Work Limitation Reporting and Disability Programs in Europe and the U.S.

Na Yin, Ph.D.
Baruch College, City University of New York (CUNY), New York City
CUNY Institute for Demographic Research, New York City
E-mail: Na.Yin@baruch.cuny.edu

Frank Heiland, Ph.D.
Baruch College, City University of New York (CUNY), New York City
CUNY Institute for Demographic Research, New York City
E-mail: Frank.heiland@baruch.cuny.edu

April 6th, 2015

Prepared for the 2015 Annual Meetings of the Population Association of America

Abstract

Recent studies have documented substantial variations across countries in the way people characterize an identical work disability (presented in a vignette). However, to date systematic analysis of the potential determinants and mechanisms underlying the observed cross-country differences in disability reporting has been lacking. We explore the role that cross-country disability policy differences, measured using a broad range of program characteristics, play in shaping individuals' disability reporting styles. We use anchoring vignettes available in comparable U.S. and European survey data to test and adjust for reporting differences in self-reported work disability measures. We find evidence that people under more generous disability regimes apply a more inclusive (i.e., lenient) scale in their assessments of work limitations. This research contributes to a better understanding of the role of disability policy and reporting heterogeneity in comparative disability research. Our findings have important implications for disability policy design and delivery.

Keywords: Work Limitations, Disability Policy, Anchoring Vignettes, HRS, SHARE

I. Introduction

Work disability is a complex social construct related to work capacity not just physical health. In practice, researchers often rely on self-reported disability measures that combine a variety of factors determining work limitations into a single measure. Many household surveys ask respondents to rate the severity of their work limitations on a five-point scale. An important concern with these instruments is reporting heterogeneity: Individuals may characterize the same objective level of work limitation differently by adopting a personal classification rule that is unobserved.

Reporting style may be influenced by individual-level as well as aggregate factors such as country-level contextual variables. This provides unique challenges, especially when conducting comparative research on work disability and health. Looking at chronic health and functional limitations among those 50 years and older in the U.S. and European surveys, Americans stand out as the population with the most health problems and limitations (see Figure 1). However, the distribution of the self-reported work limitations in the U.S. is very similar to that in some European countries (see Figure 2). Although work limitations and health/functional limitations are not identical, the changes in country rankings are striking as we move from the relatively more objective measures of health conditionals and functional limitations to the subjective measures of work limitations. One possible explanation for this discrepancy could be the presence of cross-country differences in reporting styles.

Recent studies (e.g. Angelini et al., 2012; Bago d'Uva et al., 2008; Grol-Prokopczyk et al., 2011; Kapteyn et al., 2007 and 2009; Murray et al., 2003; Sadana et al., 2002.) employ vignette data to identify and adjust for reporting heterogeneity in self-reported measures of health and disability. These analyses have documented substantial variation across and within countries in

the way individuals characterize the severity of health conditions or work limitations presented in vignettes. However, to date systematic analysis of the potential determinants and mechanisms underlying the observed cross-country differences in disability and health reporting has been lacking. Angelini et al. (2012) explore the impact of country-level factors, including public disability spending, the employment rate among the elderly, and consumer prices, on disability reporting in selected European countries. They show that more generous spending on disability programs in a country is associated with higher prevalence of work disability but do not find an effect of public disability spending on reporting styles.

In this paper, we explore the role that cross-country differences in disability policy play in how individuals characterize the severity level of work limitations, accounting for individual characteristics. We focus on the effect on reporting styles of the generosity of disability policies measured using a broad range of program characteristics that include disability policy coverage, eligibility, benefit level and duration, medical and vocational assessment, and interactions between disability policy and other labor market institutions such as unemployment benefits. The extensive set of program characteristics deepens the analysis relative to Angelini et al. (2012), while including the U.S. in the study broadens its scope and may enhance the validity of the findings.

We take advantage of work disability vignette data available in high-quality nationally representative household surveys from the U.S. and the Europe. Vignette data are becoming increasingly popular in social science research. For health and disability vignettes, respondents are asked to rate, using an ordinal categorical scale, the severity of the health or work limitations of identical hypothetical individuals. Vignette ratings allow researchers to evaluate whether respondents exhibit different response styles. The vignettes represent a fixed level of work

disability, so the differentials in the disability severity classification are attributable to different scales used by the respondents.

The cross-national comparative approach is particularly appealing for examining the effects of institutional differences on disability reporting styles. An understanding of how disability policies in a society affect the disability rating styles among the society members could provide an opportunity to refine policy design and delivery so as to change beliefs about work disability. The changing beliefs could affect individuals' behavior and decision-makings such as labor force participation and disability applications.

The paper is structured as follows: In Section II, we compare the disability policies in the United States and seven European countries, and briefly discuss the mechanisms of how individuals perceive work limitations under different policy environments. In Section III, we describe the anchoring vignette approach, the data and preliminary evidence. In Section IV, we present our estimation results and conduct policy simulations. Section V concludes.

II. Disability Policy across Countries

OECD Classification for the Disability Compensation Policy

The main variables of our interest are measures of disability policy generosity, which is prepared by the OECD (2003). OECD provides classification for disability compensation policy for all its member countries. For every country, 10 disability policy dimensions are evaluated: coverage, minimum disability level, disability levels for full benefits, maximum benefit level, permanence of benefits, medical assessment, vocational assessment, sickness benefit level, sickness benefit duration, and unemployment benefit level and duration. For each dimension, OECD scores on a scale from 0 to 5 for each country, with 5 being the most generous in that

policy dimension. The OECD measure captures not only the formal disability program rules but also the implementation and administration of the rules. Table 1 provides the detailed classification for disability compensation policy dimension. In the table we also list the eight countries in our analytical sample according to their scoring in each policy dimension.

Overall, according to a generosity index that we constructed by summing the scores in all 10 dimensions, the U.S. ranks as the least generous disability system (index = 21) and Sweden ranks the most generous (index = 34). However, countries are ranked differently in each specific policy dimension.

The U.S. federal disability system intends to award benefits only to the individuals who are fully and permanently disabled. It does not allow partial or temporary disabilities. The U.S. scores the lowest by the OECD classification for its strictest "minimum disability level". To receive any disability benefits, an American must have at least 71-85% disability level, compared to only 0-25% for a German, a Dutch, or a Swede, 26-40% for a Spaniard, and 56-70% for a Belgian, French or an Italian.

Sickness benefits in many countries usually target at individuals with less severe work limitations and serve as a precursor to long-term disability programs. The U.S. has the shortest duration and lowest benefit level for sickness absence compared to the European countries.

A worker who suffers health impairment and job separation but still has residual work capacity may choose to apply for unemployment benefits or partial disability benefits. This is mostly likely the case in European countries whose disability systems offer partial benefits. In the U.S., the disability system awards benefits only to the fully disabled individuals who are not supposed to qualify for any unemployment benefits. Compared to Spain where disability benefits are more generous than unemployment benefits, and the U.S. where disability benefit levels are

generally lower than unemployment benefit levels, Sweden offers similar levels of disability benefits and unemployment benefits for a resident but the duration is longer for disability benefits.

In terms of vocational assessment, Sweden, like the U.S., takes into account all the jobs available in the national economy, whereas some other countries, such as Belgium and France, use one's previous earnings/occupations as reference in evaluating one's residual earning/work capacity.

Despite the relatively lenient standard used in Europe, compared to the U.S., in awarding disability benefits to people with less severe work limitation (reflected in the lower minimum disability level, higher and more durable sickness benefits, and more generous disability benefits than unemployment benefits), European countries require a higher or similar disability severity level to qualify for full disability benefits. For example, Sweden requires as high as 86-99% disability level for full disability benefits while its minimum disability level is only 0-25%. The U.S. system does not offer benefits for mildly or moderately work limited individuals. It allows disability benefits only to those with a disability level of 71-85% or higher.

With respect to medical assessment, the U.S. disability determination relies on the applicants' treating doctors' opinion predominantly whereas the European systems mainly depend upon insurance doctors' evaluations. This is the only measure, among the ten, that the U.S. takes the lead in leniency.

Sweden scores 5 for having the most extensive policy coverage for its population. The U.S. disability system covers labor force plus a means-tested non-contribution scheme. According to the proportion of the work earnings that can be replaced by the maximum benefit level, Sweden

and Netherlands rank the highest with the most generous replacement rate (>=75%), and the U.S. sets a moderate replacement rate of 50-75%, the same as the rate in France and Italy.

In Sweden, the disability compensation is neither permanent like in Spain (strictly permanent), Belgium and U.S. (*de facto* permanent), nor temporary (unless for fully disabled) like in France, Germany and Italy. The Swedish system requires periodic reviews on the beneficiaries and decides on a case-by-case basis.[1]

Some of the ten policy dimensions are highly correlated, as evidenced by the magnitudes of the correlation coefficients between them (Appendix Table 1). For example, policy 2, 8, 9 and 10 are strongly and positively correlated, as the correlation coefficients between any two of those dimensions are about 0.7 or larger. Moreover, these four policy dimensions affect the vignettes' ratings in the similar fashion, as indicated by the correlation coefficients between each policy dimension and the vignettes classifying (Appendix Table 2). Policy 1 and 4 are also highly correlated. So in our regression estimation shown later, we group the correlated policy dimensions to address the potential collinearity (In an alternative model specification, we include an index that summarizes the ten policy dimensions.)

Policy Perception and Disability Reporting

Tremendous research effort has been devoted to estimating the behavioral responses, such as labor force participation and disability applications, to disability policy changes. We argue that there may be alternative pathway through which disability policy changes get to individuals, for example, by affecting people's perception about work limitations. We take advantage of the variations in disability policy generosity scores across countries and evaluate how policy differences affects what an individual think should constitute a work disability.

Before conducting estimation, we discuss the possible mechanisms of policy effects on individuals' perception. The public disability policy, in a large part, reflects the opinions towards work disability from the majority of electorates in a country. (Sanderson and Scherbov in 2007 had a similar discussion about aging policies and the electorates.) The policy may then shape an individual's opinion about work disability by signaling social norms. In other words, the disability policy provides knowledge of a threshold or reference that an individual uses to classify a work disability. For example, many European countries have long used partial disability benefits as a way to encourage people to remain at work, or to return to employment (See Note 2 and Note 3). In these countries, being disabled and having work capacity are not necessarily mutually exclusive. In countries where partial disability is an option and a larger percentage of the population is a part of the disability policy, people may consider a mild disability as a possibility and adopt a less strict definition of disability in general. In a disability welfare state that recognizes various extents of work limitations, the society members may rate a health problem on a wider severity spectrum.

In a country that offers little support to disabled workers and has more limited and stringent disability policy and programs, people may have less exposure to and experience with the disability policy and programs. The disability programs are usually reserved for the most severe disability cases that have little residual work capacity. Many people may get the impression that disability support is not easily available and mild or moderately severe health problems are not defined as "disabling" or qualify for disability subsidies. As a result, members of such societies may have a higher threshold for classifying someone as disabled.

The main instruments we use to carry out the analysis below are a series of vignette characters with different levels of health limitations. The characters are generally described as

someone currently engaging in employment, with explicit expressions such as "…pain in her back…sometimes prevents her from doing her work", "…pain gets worse while he is working", "…gets depressed once a week at work…", and "…heart problem…cholesterol level. Sometimes if he feels stressed at work he feels pain in his chest…" For such vignettes, a respondent from a stricter disability welfare state would less likely rate the health problem as work limiting, since the vignette person is working. His rating likely lies towards the milder end of the disability severity spectrum. An individual from a more generous disability welfare state would likely rate the same health problem on a wider spectrum of disability severity, for example, ranging from mildly, moderately to severely and extremely work limited. The different categorizations for disability severity, everything else being constant, could be related to the disability grids used in the disability programs in the country.

III. Anchoring Vignette Approach and Data

Many general household surveys contain self-reported measures of health and work limitations. These instruments are often based on ordinal categorical response scales. A commonly used question is: "To what extent are you limited in the kind or amount of work you can do because of an impairment or health problem?" Respondents are asked to answer on a 5-point scale: Not at all limited; mildly limited; moderately limited; severely limited; extremely limited (or cannot do any work).

An important concern with categorical self-reports of health and work limitations are so-called "scaling" or "anchoring" effects (King et al. 2004). Individuals may characterize the same objective condition or degree of limitation differently. Such measurement error will especially affect comparative research using data on self-reported health and disability. Different

populations may use systematically different thresholds (or cut-points) when rating health, which can, for example, result in different self-reports for individuals who have the same level of "objective" health or work limitation.

Vignette data can be used to study reporting heterogeneity and. A vignette describes the work limitation of a hypothetical person and then asks the respondent to evaluate the severity of the vignette work limitation on the same five-point scale used for their own health assessment. Since the vignettes are identical for all the respondents, the differences in respondents' evaluations must be due to different reporting styles. We hypothesize that the scale that the respondents use to classify the severity of a given vignette character's work limitation is a function of the country's disability policy, particularly a severity classification scale used by their country's disability system.

We use the 2004 wave of the Health and Retirement Study (HRS), a bi-annual panel with a representative sample of the US population age over 50 and their spouses. It has been conducted by the University of Michigan since 1992. Information collected includes health, socio-economic status, and social program participation. We use a subsample of respondents who first completed a face-to-face interview and later completed a leave-behind questionnaire that consists of a series of work disability vignettes.

The Survey of Health, Ageing and Retirement in Europe (SHARE) is a longitudinal dataset on European citizens of age 50 and older and their spouses. We use the 2004 wave of the survey. SHARE was purposely modeled after the HRS and follows a common set-up across all countries with the goal of facilitating cross-country research. For a subset of countries that agreed to participate, SHARE included a set of self-assessments and vignette questions on work limitations as part of a drop-off questionnaire. The eight countries that participated in this vignette

experiment are Belgium, France, Germany, Greece, Italy, the Netherlands, Spain, and Sweden. The work disability vignettes were identical to the work disability vignettes in the HRS leave-behind questionnaire. In our analysis, we exclude Greece because the comparable index for the disability policy generosity is not available for Greece in the OECD report.

The work disability vignettes describe work limitation of a hypothetical character in three domains: pain, depression, and cardiovascular health. In each domain, several vignette questions are asked. We use the common nine vignettes in HRS and SHARE survey. The text for all the vignettes is provided in a data appendix. For each vignette, the respondent is asked: "How much is s/he limited in the kind or amount of work s/he could do?" The answer follows a five-point scale: 1) None; 2) Mild; 3) Moderate; 4) Severe; 5) Extreme/Cannot do any work. Preceding the vignette questions, respondents are asked about their own work limitations: "Do you have any impairment or health problem that limits the kind or amount of work you can do?" with the same answer categories.

As shown in Figures 3-1 and 3-2, we observe considerable differences in how residents across countries characterize the work limitation severity characterized in the vignettes. The differences between the U.S. and European countries are even more striking. European respondents, compared to Americans, tend to classify a vignette as more work limiting. American respondents are the strictest overall in rating work limitations, followed by the Italians, the Belgians, the Dutch, and the French. The Swedes and the Spaniards are at the other end of the spectrum and inclined to rate a work limitation as more severe. The German respondents are in the middle.

These patterns are notable in light of our discussion of disability program generosity above. More inclusive rating scales are associated with more generous disability regimes, as indicated

by the fact that the country ranking according to the inclusiveness of disability rating is in line with the ranking in terms of the generosity of the country's disability system.

Figure 3-1 illustrates a fairly strong correlation between the generosity of disability system and disability vignette rating. The x-axis represents a country's disability policy generosity index. The y-axis refers to the percentage of respondents in a country who classify a vignette as not at all limited. Each diamond in the graph represents the rating for a specific vignette. There are apparent variations between countries with different disability policy generosity in classifying the severity of the same disability vignettes. The difference in rating styles is striking between the U.S. (generosity index=21) and other European countries. The graph indicates a strong and negative correlation between disability policy and disability ratings. That is, more generous disability policy seems to be associated with more inclusive disability classifying styles. Together with Figure 3-2, it shows that respondents under more generous disability regimes are more likely to report the same vignettes as more work limiting. A detailed characterization of the empirical model is provided in the Appendix.

IV. Main Results

In this section, we present results of our analysis of the determinants of reporting styles (Table 3.1 and 3.2). In the model we include detailed sets of individual-level factors as well as country-level factors, as previous literature has shown that disability results from both individual factors and environmental factors (e.g. Mitra, 2006). We are particularly interested in the role of cross-country differences in disability policy generosity in affecting reporting scales. Then to illustrate the effects of accounting for different reporting styles, we also estimate a model that determines the self-reported work limitation allowing for response scale heterogeneity, and

compare it with a model assuming a homogeneous scale (Table 4). Lastly, we conduct some policy simulation to illustrate the impact of disability policy generosity on the country's self-reported disability distribution.

It is of our main interest to relate reporting styles for work limitation to disability institutional arrangements. We look at variables characterizing the generosity of a country's disability policy. In the estimation (Table 3.1), we include an index that summarizes in the ten disability policy dimensions how a country scores in terms of policy generosity according to OECD. To capture the remaining country-level effect on disability rating styles, we include a series of country dummies (U.S. is the omitted/reference country). Due to the collinearity between the country dummies and the country disability policy index, we group some countries together that share similar values in the major areas of life (religion, politics, and economic and social life). The value and culture proximity was measured in the Inglehard-Welzel Cultural Map based on the World Values Survey data.

While it is interesting to estimate the effects of the overall generosity of disability policy on work limitation ratings, in an alternative model (Table 3.2), we also examine the predictive power of individual policy dimensions.

When modeling the reporting thresholds using the vignette data (Tables 3.1 and 3.2), the model also includes standard demographic covariates—gender (dummies for being female), age (in groups), education (in years), and a series of health indicators (dummies for high blood pressure, diabetes, cancer, lung problems, heart problems, arthritis, obesity, and number of ADL limitations, and number of IADL limitations). Work disability, not just a medical measure, also depends on the work context, as explained in previous literature (E.g. Bernell, 2003). So we also include 7 dummies for (last) occupation (technical/sales/administrative support, service,

farming/fishing/forestry, precision production/craft/repair, operators/fabricators/laborers, elementary occupation, and managerial/professional specialty as omitted/reference occupation). Finally, we note that since we are pooling the vignette data, the (latent) severity level is specified as a function of 9 vignette dummies as shown in equation (3) in the model appendix.

In the model of self-reported severity of work limitations (Table 4), we include a similar set of variables to those in the thresholds equation except that we exclude the index for disability policy generosity and include a full set of country dummies to account for unobserved country-level differences in resulting work limitations. The severity of work limitations is a function of health problems, occupations, and demographic characteristics such as age, gender, and education. We have also tried to include a full set of interactions between the occupations and the health conditions in order to evaluate the risk of having a work limitation for individuals who have certain health problems and engaged in different types of work (results not shown).

Table 2 lists the means of the explanatory variables and self-reported work disability by country. The table shows large differences in years of education, with low means in the southern European countries. There are also obvious differences in the age composition, with, for example, relatively few 66-70 year olds in Sweden. All the chronic conditions and functional limitations are much more prevalent in the U.S. than in European countries. Still, the distribution of self-reported severity of work disability in the U.S. is quite similar to what is observed in European countries overall (columns “US” vs. “Europe”).

However, there are notable differences in the raw distributions across the 7 European populations. For example, while all distributions are fairly right-skewed (severity ranging from “none” to “extreme”), Sweden is particularly concentrated at “none” and the Netherlands is very concentrated at “none” and “mild”, while the distribution is relatively more equal across the

severity categories in Belgium. There are apparent differences in the occupational distribution. For example, the most common occupations are managerial and professional specialty in the U.S. and Sweden, while Germany and France are dominated by occupations associated with technical support, sales and administrative support.

Evidence of Response-Scale Heterogeneity

Table 3.1 and 3.2 shows the estimates from the threshold and the vignette equation corresponding to the HOPIT model (explained in the model appendix) that adjusts for scaling heterogeneity. The estimated effects from the four cut-point equations are listed from left to right for the four thresholds μ_1 ("not at all limited" to "mildly limited") to μ_4 ("severely limited" to "extremely limited"). The results are consistent with systematic reporting heterogeneity at the individual and country level as some of the covariates are found to be predictive of the location of the thresholds.

At the country level, the disability policy generosity scores show statistically significant predictive power for the respondents' disability classification scales, with the effects stronger and more statistically significant at the lower end and the middle of the scale (μ_1, μ_2, and μ_3).

Respondents under more generous disability regimes tend to apply a more inclusive scale in classifying a mild, or moderate or severe work limitation. But at the cut-point μ_4, the policy effect flips the sign and becomes positive. This means that the classifying scales become less inclusive as the generosity scores increase when it comes to classify an extreme work limitation. The reporting patterns may relate to the fact that the relatively more generous European disability systems set a more lenient standard for admitting people with less serious work limitations but a stricter requirement for allowing the most severe work limitations.

After accounting for differences in disability policy generosity (and individual-level factors), the country dummies estimate the residual differences in reporting styles between countries. We observe that most of the European countries, compared to the U.S., tend to place a lower threshold (i.e. more inclusive scale) at the lower end and the higher end of the scale. For the middle of the scale, the differences in reporting styles between European countries and the U.S. are less consistent. For example, compared to the U.S., the Netherlands place a higher threshold for classifying moderate and severe work limitations, and Germany uses a less inclusive scale for defining severe work limitations.

Having observed that more lenient disability institutions generally predict more inclusive reporting styles for work limitations except for the extremely severe cases, we are also interested to understand what specific policy dimensions drive the results and whether different policy aspects link to the rating styles differently. To that end, we re-estimate the same specifications as before replacing the generosity index and country dummies with separate policy dimensions (with a few grouped together as explained earlier). The estimation results are presented in Table 3.2. More extensive coverage and higher maximum benefit level predict more inclusive rating styles at all four cut-points, that is, over the whole work limitation severity distribution. Permanence of the benefits is also associated with applying lower thresholds, with the strongest effects at the middle of the work limitation distribution. Easier entry into disability programs with a mild work limitation, reflected in lower minimum disability level, more generous sickness benefits, and better disability benefits compared to unemployment benefits, predicts more inclusive rating styles for work limitation severity.

More lenient medical assessment in a country, that is, the disability system places greater weight on the opinion of the applicant's treating doctor, seem to predict lower cut-points in

rating work limitation severity among its residents. This inclusive style applies to the middle and the right end of the work limitation distribution (μ_2, μ_3, and μ_4), but not to the first cut-point when it comes to classifying a vignette as mildly limited vs. not at all limited.

Vocational assessment shows much larger effects compared to other policy dimensions with the effects concentrated in the first three thresholds. Under a disability regime with relatively lenient vocational assessment, that is, eligibility for disability benefits is based on inability to do one's own usual occupation (rather than inability to take on any job available), a respondent seems to be more likely to classify a given health problem as work limiting except when it comes to rate extremely severe health problems.

Disability levels for full disability benefits show remarkably different effects on reporting styles. More lenient standards to determine disability levels for full disability benefits used in a country are associated with stricter rating styles for work limitations among its residents. The effects are concentrated in the middle of the work limitation distribution (μ_2 and μ_3). But the sign of the effects looks counterintuitive and puzzling.

At the individual level, there is evidence that the thresholds are shifted to the right for women, which indicates that they apply stricter thresholds when describing the severity level. In other words, women tend to classify the same degree of work limitations (vignette case) lower on the five-point scale than men, all else equal. Individuals with more education are more lenient at the lower end of the scale (none to mild and mild to moderate work disabilities) but place a higher threshold after that, especially at the higher end of the disability severity distribution (severe to extreme).

We also observe non-monotonic effects of age (the reference age group is 50-55): older individuals tend to place their cut-points lower when classifying a given vignette as mild or

moderate work disability but do not do so for classifying more severe work disability. Having certain health conditions, such as high blood pressure, diabetes, heart conditions, and obesity, is associated with a lower threshold at certain parts of the disability severity distribution.

However, respondents with cancer, lung problems or arthritis tend to apply a higher cut-point when classifying severe or extreme work limitations. More ADL limitations are associated with more inclusive scales at both ends of the disability severity spectrum. There are also obvious differences in reporting styles between occupations. Respondents with managerial or professional specialty seem to be much stricter in rating disability severity compared to other occupations.

Determinants of Self-reported Work Limitation

Table 4 presents estimation results for the model of work disability severity with and without adjustment for reporting scale heterogeneity. The results in the table are estimated for the pooled sample including the U.S. and European countries. The model explains the self-reported work disability on a five-point scale. The model "without adjustment" is a HOPIT model that does not allow the cut-points to vary with respondents' characteristics. It is similar to a standard ordered probit model. The model "with adjustment" is a HOPIT model allowing for reporting scale variation. The former model is rejected by the data (as the log-likelihoods at the bottom of the table indicate).

We find that work disability increases with age, and decreases with schooling. Having any of the health conditions that are included in the model increases the likelihood that a person reports being more severely disabled. Individuals (formerly or currently) in technical / sales / administrative support, precision production / craft / repair, and operators / fabricators / laborers

occupations are more likely to report work limitations compared to those in managerial/professional specialty occupations.

Comparing the estimates in the models with and without scaling adjustment, we observe that most of the effects, except the country fixed effects, do not change significantly after adjustment. The health protective effect of education increases meaningfully when scaling heterogeneity is accounted for. This means that the educational effect on work disability found in the homogeneous reporting model was a mixture of a negative effect of education on work disability and an overall tendency for the more educated to report more severe work limitations. There are also obvious changes in the effects of health conditions and the effects of occupations after the adjustment. Prevalence of (self-reported) work limitation also appears to be related to country-level effects. Controlling for demographics, health conditions and occupations, and accounting for scaling heterogeneity, we find that Europeans suffer less from work limitations than Americans. Before the adjustment, individuals in Germany, Sweden, France and Belgium are more likely to report a disability than Americans, all else constant. It is likely a result of better health and overall more inclusive disability reporting styles among European respondents, the latter effect being linked to the more generous disability institutions in Europe.

Counterfactual Simulations of the Effects of Disability Policy Generosity

To illustrate the effect of different disability policy environments on response scales and thus on disability severity distributions, we conduct some policy simulations and show the results in Figure 4. In each graph, the first column represents the disability severity distribution adjusted by demographics, health conditions and occupations in that country and applying the response scale from that country. The second and third columns represent the simulated disability severity

distributions when we assign the U.S. and Sweden disability policy scores, respectively, to each country (the graphs for U.S. and Sweden have only two columns).

The U.S. is the strictest while Sweden is the most inclusive according to the overall policy generosity scores. When we apply U.S. disability policy scores to the European countries, we observe a dramatic fall in reported work disability for all the seven European countries. For instance, the percentage of respondents without a work disability goes up from 33.3% to 64.7% for Spain. The smallest rise in the non-disabled population occurs in Italy (from 59.6% to 62.2%) due to the very similar disability policy generosity scores between Italy and the U.S. Again this is a result of the fact that in a country with disability policies as strict as the U.S., respondents are less likely to classify a given health problem as work limiting.

In contrast, when we apply Sweden's disability policy score to the U.S. and the other European countries, we observe a marked rise in reported work disability for all the countries. The effects for Italy and the U.S. are among the most pronounced - the percentage of respondents with any work disability increases by 42 and 36 percentage points, respectively. Substantial increases appear in every category of disability severity. For example, in the U.S., the percentage of respondents reporting mild work disability goes up from 35% to 45%, the percentage reporting moderate disability from 11% to 29% while the rate of severe or extreme disability rises from 7% to 15%.

V. Conclusions

Recent studies have documented substantial variations across countries in the way people characterize an identical work disability (presented in a vignette). However, to date systematic analysis of the potential determinants and mechanisms underlying the observed cross-country

differences in disability reporting has been lacking. We explore the role that cross-country disability policy differences, measured using a broad range of program characteristics, play in shaping individuals' disability reporting styles.

We use anchoring vignettes available in comparable U.S. and European survey data to test and adjust for reporting differences in self-reported work disability measures. Using disability policy generosity scores provided by the OECD, we link the variation in disability institutional context across countries to the systematic differences in disability reporting. We find evidence that people under more generous disability regimes apply a more inclusive (i.e., lenient) scale in their assessments of work limitations.

This research contributes to a better understanding of the role of disability policy and reporting heterogeneity in comparative disability and health research. Our comparative analysis of disability policy generosity extends earlier work on country-level factors such as Angelini et al. (2012), who focus on the impact of public disability spending in selected European countries. Using an anchoring vignette strategy similar to our approach, they find no statistically significant effects of public disability spending on reporting style. This is interesting in light of our results and suggests that a measure of spending alone may not fully capture the complex nature of cross-country differences in disability generosity.

One concern with the evidence presented here is that disability policy could be endogenous to disability severity ratings. Disability policy and disability perception have reciprocal effects on each other. Citizens' attitudes towards disability shape the country's disability policy and the policy would in turn shape people's opinion about disability and change the disability culture in the country. Alesina and Giuliano (2013) review the two-way effects between culture and institutions in general. In a society with a tough culture, members may not find generous

disability support important, which is then reflected in the stricter disability policy. In the paper, we focus on how the existing disability policies in a country affect people's perception about work disability. How citizens' attitudes towards work disability affect the development and passage of disability policies is beyond the scope of the present study. We note, however, that in most cases the disability policies in the countries investigated here have been in place for a long time, so disability institutions and disability culture are likely interwoven and hence hard to distinguish.

Disability programs are a substantial and rising component of public social expenditures and an important dimension of the social safety net in many developed countries. An understanding of how disability policy affects the opinions about work disability among society members would have important implications for the policy design and delivery. As we show, particular aspects, such as partial benefits design, medical and vocational assessments, rules about sickness benefits and unemployment benefits, in the disability institutional arrangements may change the way people think about work disability. The policies that aim at promoting employment among people with disabilities, such as the Ticket to Work Program and the Benefit Offset Program in the U.S., will also likely modify the perceptions about work disabilities as time passes by and thus change the culture about work in the wake of health limitations (See Note 4). The changing beliefs, in turn, could affect individuals' behavior and decision-making related to health and work. This aspect of disability policy has received little attention in prior literature and, in light of the result of the present study, we propose it as a fruitful avenue of future research.

References

Alesina, A. and Giuliano, P. (2013). *Culture and Institutions*. (NBER Working Paper No. 19750). Cambridge, MA*:* National Bureau of Economic Research Working Paper.

Angelini, V., Cavapozzi, D., Paccagnella, O. (2012). Cross-Country Differentials in Work Disability Reporting Among Older Europeans. *Social Indicator Research,* 105: 211-226.

Bago d'Uva, T., O'Donnell, O., van Doorslaer, E. (2008). Differential Health Reporting by Education Level and Its Impact on the Measurement of Health Inequalities among Older Europeans. *International Journal of Epidemiology*, 37: 1375-83.

Bernell, S.L. (2003). Theoretical and Applied Issues in Defining Disability in Labor Market Research. *Journal of Disability Policy Studies*, 14(1): 36-45.

Brandt, D. E., Houtenville, A. J., Huynh, M. T., Chan, L. & Rasch E. K. (2011) Connecting Contemporary Paradigms to the Social Security Administration's Disability Evaluation Process. *Journal of Disability Policy Studies*, 22(2), 116-128.

Garcia-Gomez, P., Jimenez-Martin, S. & Vall Castello, J. (2011). *Health, Disability and Pathways to Retirement in Spain*. (NBER Working Paper No. 17048). Cambridge, MA*:* National Bureau of Economic Research Working Paper.

Grol-Prokopczyk, H., Freese, J., Hauser, R. (2011). Using Anchoring Vignettes to Assess Group Differences in General Self-Rated Health. *Journal of Health and Social Behavior,* 52(2): 246–61.

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

Kapteyn, A., Smith, J.P., van Soest, A. (2009). *Work Disability, Work, and Justification Bias in Europe and the U.S.* (RAND Working Paper WR-696). Santa Monica, CA: RAND, Labor and Population Working Paper.

King, G., Murray, C.J.L., Salomon, J., Tandon, A. (2004). Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research. *American Political Science Review*, 98: 184-91.

Livermore, G., Mamun, A., Schimmel, J., Prenovitz, S. (2013). *Executive Summary of the Seventh Ticket to Work Evaluation Report*. Washington, DC: Center for Studying Disability Policy.

Mitra, S. (2006). The Capability Approach and Disability. *Journal of Disability Policy Studies*, 16(4), 236-47.

Mitra, S. (2009). Temporary and Partial Disability Programs in Nine Countries: What Can the United States Learn from Other Countries? *Journal of Disability Policy Studies*, 20(1), 14-27.

Murray, C.J.L., Ozaltin, E., Tandon, A., Salomon, J., Sadana, R., Chatterji, S. (2003). Empirical Evaluation of the Anchoring Vignettes Approach in Health Surveys. In: Murray, C.J.L., Evans, D.B. ed. *Health Systems Performance Assessment: Debates, Methods and Empiricism*. Geneva: World Health Organization.

Organisation for Economic Co-operation and Development. (2003). *Transforming Disability into Ability: Policies to Promote Work and Income Security for Disabled People*. Paris: OECD Publishing.

Organisation for Economic Co-operation and Development. (2010). *Sickness, Disability and Work: Breaking the Barriers*. Paris: OECD Publishing.

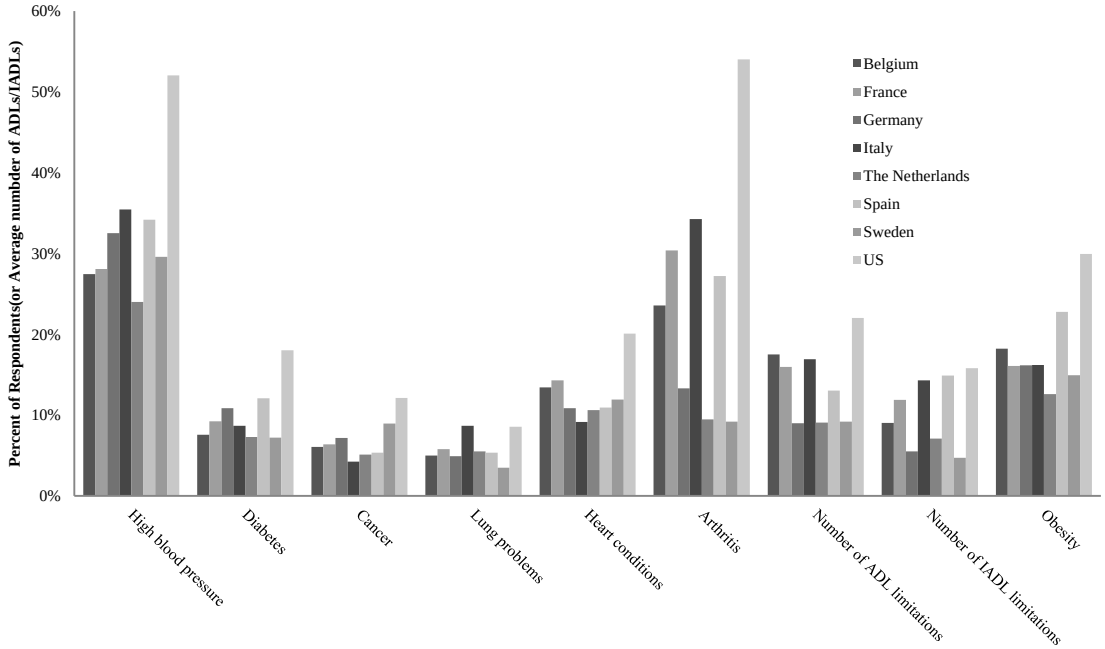
Sadana, R., Tandon, A., Murray, C.J.L., Serdobova, I., Cao, Y., Xie, W.J., Chatterji, S., Ustun, B.L. (2002). *Describing Population Health in Six Domains: Comparable Results from 66 Household Surveys*. (Discussion Paper No. 43). Geneva: World Health Organization Global Programme on Evidence for Health Policy Discussion Paper.

Sanderson W.C., Scherbov, S. (2007). A Near Electoral Majority of Pensioners: Prospects and Policies. *Population and Development Review*, 33(3): 543-554.

Weathers II, R. R. and Hemmeter, J. (2011). The Impact of Changing Financial Work Incentives on the Earnings of Social Security Disability Insurance (SSDI) Beneficiaries. *Journal of Policy Analysis and Management*, 30 (4), 708-728.

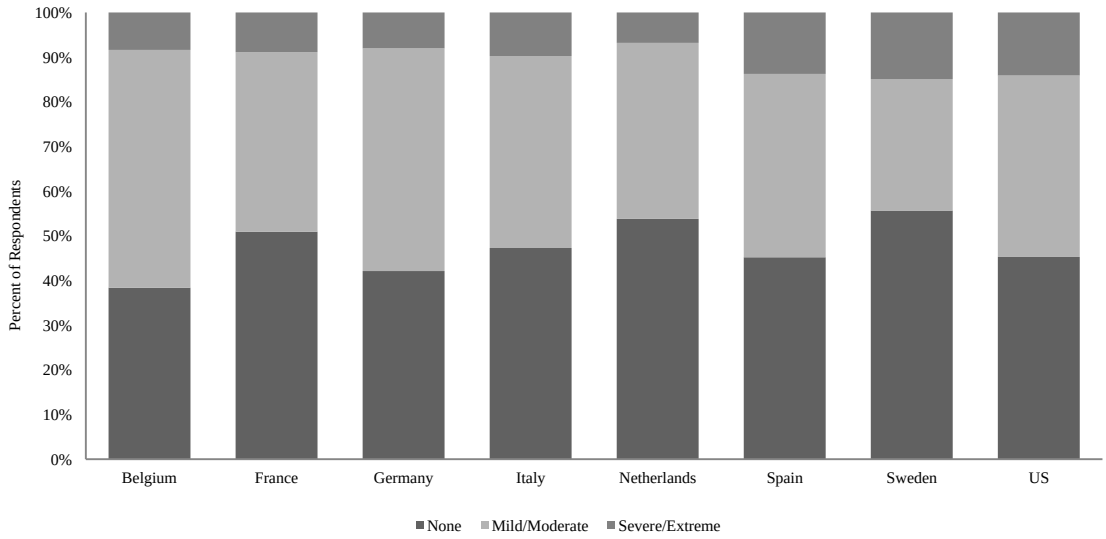
Yin, N. (2015). Partial Benefits in the Social Security Disability Insurance Program. *Journal of Risk and Insurance*. Forthcoming. DOI: 10.1111/jori.12028

Figure 1: Health Conditions, Functional Limitations and Obesity Status by Country



Source: Authors' calculations using data from the HRS and the SHARE.

Figure 2: Self-Reported Work Limitation by Country



Source: Authors' calculations using data from the HRS and the SHARE.

Figure 3-1: Disability Policy Generosity and Percentage of Respondents Classifying 9 Vignettes as Not At All Limited

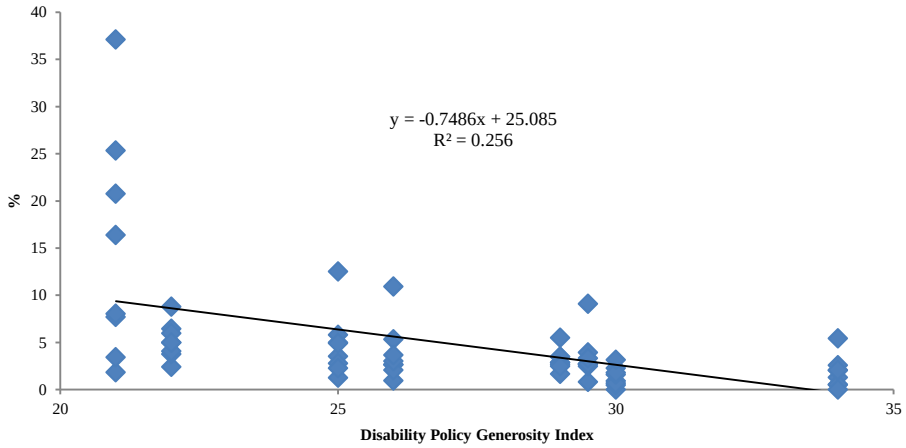


Figure 3-2: Disability Policy Generosity and Percentage of Respondents Classifying 9 Vignettes as Severely/Extremely Limited

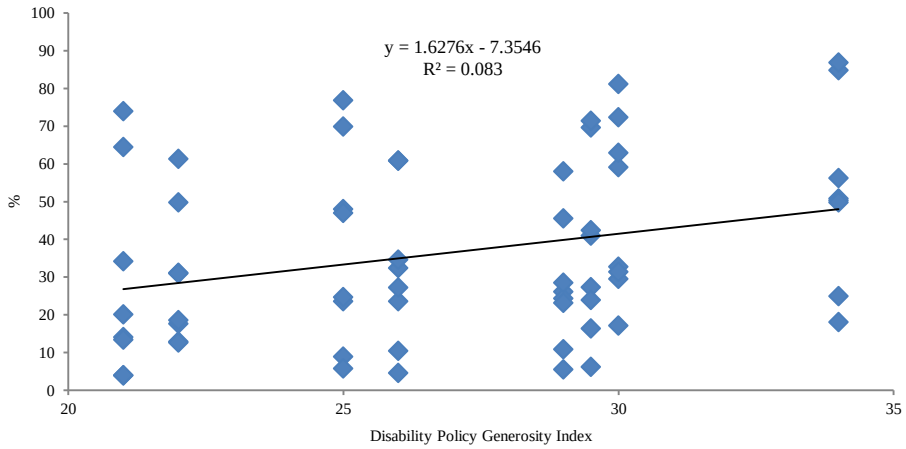
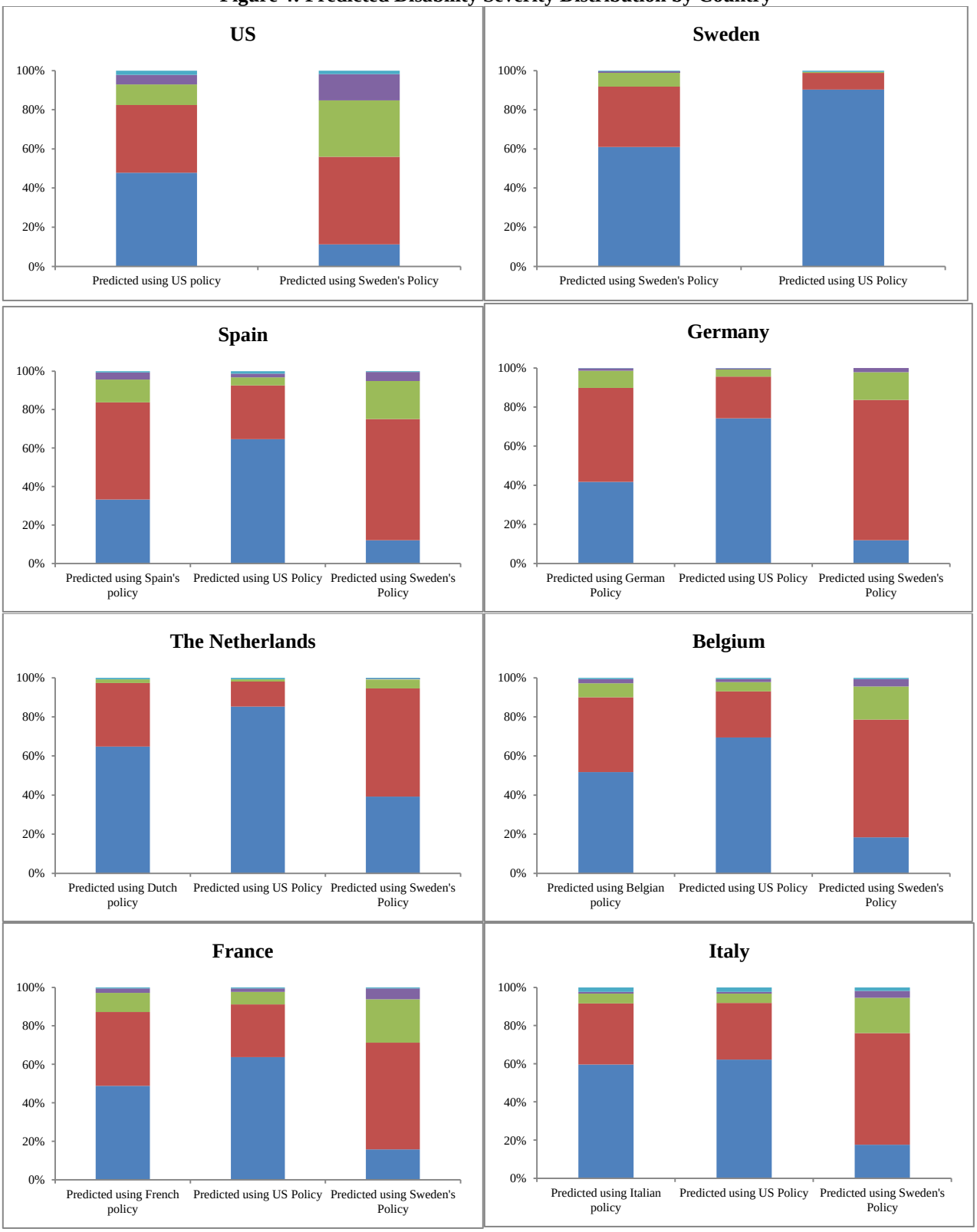


Figure 4: Predicted Disability Severity Distribution by Country



Note: Each column in the graphs represents a disability severity distribution. From bottom up in the column, disability severity ranges from none, mild, moderate, severe up to extreme.

Table 1: OECD Classification for Disability Compensation Policy Dimension (Based on OECD 2003 Table A2.1.)

	5 points	4 points	3 points	2 points	1 point	0 point
Policy 1: Coverage	total population (residents)	some of those out of the labor force	labor force plus means-tested non-contribution scheme	labor force with voluntary self-insurance	labor force	employees
	Sweden	*Netherlands*	*Belgium, France, Italy, Spain, US*	*Germany*		
Policy 2: Minimum Disability Level	0-25%	26-40%	41-55%	56-70%	71-85%	86-100%
	Germany, Netherlands, Sweden	*Spain*		*Belgium, France, Italy*	*US*	
Policy 3: Disability Levels for Full Disability	<50%	50-61%	62-73%	74-85%	86-99%	100%
			Belgium, Germany	*Netherlands, US*	*France, Spain, Sweden*	*Italy*
Policy 4: Maximum Benefit Level	rr>=75%, reasonable minimum	rr>=75%, minimum not specified	75>rr>=50%, reasonable minimum	75>rr>=50%, minimum not specified	rr<50%, reasonable minimum	rr<50%, minimum not specified
	Netherlands, Sweden	*Spain*	*France, Italy, US*	*Germany*	*Belgium*	
Policy 5: Permanence of Benefits	strictly permanent	de facto permanent	self-reported review only	regulated review procedure	strictly temporary, unless fully (=100%) disabled	strictly temporary in all cases
	Spain	*Belgium, US*	*Netherlands, Sweden*		*France, Germany, Italy*	
Policy 6: Medical Assessment	treating doctor exclusively	treating doctor predominantly	insurance doctor predominantly	insurance doctor exclusively	team of experts in the insurance	insurance team and two-step procedure
		US	*Germany, Sweden*	*Belgium, France*	*Italy, Netherlands*	*Spain*
Policy 7: Vocational Assessment	strict own or usual occupation assessment	reference is made to one's previous earnings	own-occupation assessment for partial benefits	current labor market conditions are taken into account	all jobs available taken into account, leniently applied	all jobs available taken into account, strictly applied
		Belgium, France, Germany (3.5)	*Italy, Spain*		*Netherlands, Sweden, US*	
Policy 8: Sickness Benefit Level	rr=100% also for long-term sickness absence	rr=100% (short-term); rr >=75% (long-term) sickness absence	rr=75% (short-term); rr >=50% (long-term) sickness absence	75>rr>=50% for any type of sickness absence	rr>=50% (short-term); rr <50% (long-term) sickness absence	rr<50% also for short-term sickness absence
		Germany, Sweden	*Belgium, Italy, Netherlands*	*France, Spain, US*		
Policy 9: Sickness Benefit Duration	one year or more, short or no wage payment period	one year or more, significant wage payment period	6-12 months, short or no wage payment period	6-12 months, significant wage payment period	less than 6 months, short or no wage payment period	less than 6 months, significant wage payment period
	France	*Germany, Spain, Sweden*	*Italy, Netherlands*	*Belgium*		*US*
Policy 10: Unemployment Benefit (UE) Level and Duration	DI>UE level, short duration of unemploy-ment	DI>UE level, long duration of unemployment	similar levels, short duration of unemploy-ment	similar levels, long duration of unemployment	DI<UE level, short duration of unemploy-ment	DI<UE level, long duration of unemployment
		Spain	*Italy, Sweden*	*Belgium, France, Germany, Netherlands*	*US*	

Note: rr = replacement rate; DI = Disability benefit.

Table 2: Sample Means by Country

	Belgium	France	Germany	Italy	Netherlands	Spain	Sweden	US	Europe
Demographics and Education									
Female	0.55	0.57	0.56	0.56	0.51	0.56	0.52	0.57	0.55
Age	63.9	64.9	63.8	63.7	62.8	64.8	64.1	64.6	64.1
Age 50-55	0.23	0.22	0.22	0.20	0.24	0.22	0.21	0.20	0.22
Age 56-60	0.21	0.20	0.17	0.23	0.25	0.17	0.20	0.12	0.20
Age 61-65	0.15	0.15	0.21	0.18	0.16	0.15	0.21	0.19	0.17
Age 66-70	0.15	0.13	0.20	0.16	0.14	0.16	0.09	0.23	0.15
Age 70+	0.27	0.30	0.21	0.23	0.21	0.30	0.28	0.25	0.26
Years of education	10.2	8.3	13.1	7.2	11.5	7.1	10.4	12.7	9.6
Health									
High blood pressure	0.27	0.28	0.33	0.35	0.24	0.34	0.30	0.52	0.30
Diabetes	0.08	0.09	0.11	0.09	0.07	0.12	0.07	0.18	0.09
Cancer	0.06	0.06	0.07	0.04	0.05	0.05	0.09	0.12	0.06
Lung problems	0.05	0.06	0.05	0.09	0.06	0.05	0.03	0.09	0.06
Heart conditions	0.13	0.14	0.11	0.09	0.11	0.11	0.12	0.20	0.12
Arthritis	0.24	0.30	0.13	0.34	0.09	0.27	0.09	0.54	0.22
Number of ADL limitations	0.17	0.16	0.09	0.17	0.09	0.13	0.09	0.22	0.13
Number of IADL limitations	0.09	0.12	0.06	0.14	0.07	0.15	0.05	0.16	0.10
Obesity	0.18	0.16	0.16	0.16	0.13	0.23	0.15	0.30	0.17
Self-reported work limitation									
None	0.38	0.51	0.42	0.47	0.54	0.45	0.56	0.45	0.48
Mild	0.36	0.22	0.29	0.29	0.31	0.22	0.16	0.23	0.26
Moderate	0.18	0.18	0.21	0.14	0.09	0.19	0.13	0.18	0.16
Severe	0.07	0.06	0.07	0.06	0.04	0.11	0.11	0.08	0.07
Extreme	0.02	0.02	0.01	0.04	0.02	0.02	0.04	0.06	0.02
Occupation									
Managerial/Professional specialty	0.23	0.18	0.18	0.16	0.24	0.10	0.28	0.27	0.19
Technical/Sales/Administrative support	*0.23*	*0.32*	*0.28*	*0.15*	0.23	*0.09*	*0.27*	*0.22*	*0.23*
Service	0.08	0.11	0.12	0.05	*0.16*	0.07	0.15	0.12	0.11
Farming/Fishing/Forestry	0.02	0.06	0.03	0.06	0.01	0.07	0.03	0.02	0.04
Precision production/Craft/Repair	0.12	0.12	0.13	0.16	0.08	0.13	0.10	0.10	0.12
Operators/Fabricators/Laborers	0.05	0.05	0.08	0.10	0.06	0.10	0.08	0.13	0.07
Elementary occupation	0.13	0.10	0.11	0.11	0.12	0.16	0.06	0.02	0.11
Occupation info missing	0.14	0.06	0.08	0.21	0.10	0.29	0.02	0.12	0.12
Disability Policy									
Disability policy generosity index	26	25	29.5	22	29	30	34	21	27.6
No. of observations	*543*	*833*	*489*	*426*	*508*	*430*	*402*	*3021*	*3631*

Table 3.1: Estimation Results for Vignette Ratings, Including a Disability Policy Index

	Thresholds Equation											
	Cut-point 1			**Cut-point 2**			**Cut-point 3**			**Cut-point 4**		
	Not at all limited =>Mildly limited			Mildly limited =>Moderately limited			Moderately limited =>Severely limited			Severely limited=>Extremely limited		
	Coef.		S.E.	Coef.		S.E.	Coef.		S.E.	Coef.		S.E.
Demographics and Education												
Female	0.034	*	0.018	0.091	***	0.013	0.081	***	0.013	0.025		0.019
Age56-60	-0.011		0.027	-0.003		0.019	0.024		0.018	0.023		0.026
Age61-65	-0.022		0.026	0.030		0.019	0.074	***	0.018	0.147	***	0.027
Age66-70	-0.007		0.026	-0.002		0.019	0.037	**	0.019	0.084	***	0.027
Age70+	-0.056	**	0.025	-0.037	**	0.018	0.066	***	0.018	0.217	***	0.026
Years of education	-0.027	***	0.003	-0.003	*	0.002	0.003	*	0.002	0.010	***	0.003
Health												
High blood pressure	-0.047	***	0.017	-0.023	*	0.013	0.001		0.012	-0.011		0.019
Diabetes	-0.001		0.024	-0.033	*	0.018	-0.055	***	0.018	-0.121	***	0.025
Cancer	0.024		0.027	0.025		0.020	0.049	**	0.020	0.119	***	0.032
Lung problems	0.016		0.031	-0.022		0.023	0.013		0.023	0.051		0.034
Heart conditions	-0.063	***	0.023	-0.039	**	0.017	-0.022		0.016	0.019		0.024
Arthritis	-0.049	***	0.018	-0.011		0.013	0.005		0.013	0.045	**	0.020
Number of ADL limitation	-0.057	***	0.017	-0.030	***	0.012	-0.025	**	0.011	-0.106	***	0.014
Number of IADL limitation	0.011		0.019	0.005		0.014	-0.028	**	0.013	-0.034	*	0.017
Obese	-0.003		0.020	-0.033	**	0.015	-0.061	***	0.014	-0.110	***	0.020
Occupation												
Ref: Managerial/Professional specialty												
Technical/Sales/Administrative support	-0.031		0.025	-0.057	***	0.018	-0.057	***	0.017	-0.043		0.027
Service	0.018		0.030	-0.080	***	0.022	-0.101	***	0.022	-0.233	***	0.032
Farming/Fishing/Forestry	0.040		0.051	0.085	**	0.037	-0.034		0.036	-0.111	**	0.053
Precision production/Craft/Repair	-0.073	**	0.033	-0.021		0.023	-0.045	**	0.023	-0.110	***	0.034
Operators/Fabricators/Laborers	-0.036		0.032	-0.021		0.024	-0.098	***	0.023	-0.213	***	0.034
Elementary occupation	-0.083	*	0.042	-0.051	*	0.028	-0.057	**	0.027	-0.179	***	0.039
Disability Policy and Country												
Disability policy generosity index	**-0.069**	*******	**0.007**	**-0.064**	*******	**0.004**	**-0.063**	*******	**0.004**	**0.012**	******	**0.006**
Germany	-0.218	***	0.072	-0.117	***	0.043	0.235	***	0.039	-0.039		0.060
Sweden	-0.023		0.104	-0.032		0.062	0.030		0.055	-0.755	***	0.079
The Netherlands	-0.306	***	0.069	0.235	***	0.041	0.435	***	0.038	-0.343	***	0.055
Spain and Italy	-0.635	***	0.043	-0.331	***	0.029	-0.024		0.029	-0.191	***	0.040
France and Belgium	-0.516	***	0.041	-0.278	***	0.026	0.045	*	0.024	-0.035		0.036
Constant	-0.050		0.156	0.478	***	0.095	1.298	***	0.088	0.892	***	0.128

Note: We include in the estimation an indicator for missing occupation information. * p<0.10, ** p<0.05, *** p<0.01.

Table 3.2: Estimation Results for Vignette Ratings, Including Multiple Disability Policy Dimensions

	Thresholds Equation											
	Cut-point 1 Not at all limited =>Mildly limited			**Cut-point 2** Mildly limited =>Moderately limited			**Cut-point 3** Moderately limited =>Severely limited			**Cut-point 4** Severely limited=>Extremely limited		
	Coef.		S.E.	Coef.		S.E.	Coef.		S.E.	Coef.		S.E.
Demographics and Education												
Female	0.035	**	0.018	0.091	***	0.013	0.081	***	0.013	0.029		0.019
Age56-60	-0.013		0.027	-0.001		0.019	0.025		0.018	0.020		0.026
Age61-65	-0.021		0.026	0.029		0.019	0.074	***	0.018	0.153	***	0.027
Age66-70	-0.006		0.026	-0.006		0.019	0.034	*	0.019	0.096	***	0.027
Age70+	-0.056	**	0.025	-0.039	**	0.018	0.066	***	0.018	0.224	***	0.026
Years of education	-0.026	***	0.003	-0.006	***	0.002	0.001		0.002	0.013	***	0.003
Health												
High blood pressure	-0.047	***	0.017	-0.024	*	0.013	-0.001		0.012	-0.009		0.019
Diabetes	0.001		0.024	-0.034	*	0.018	-0.055	***	0.018	-0.111	***	0.025
Cancer	0.025		0.027	0.024		0.020	0.049	**	0.020	0.122	***	0.032
Lung problems	0.017		0.031	-0.023		0.023	0.013		0.023	0.042		0.034
Heart conditions	-0.063	***	0.023	-0.038	**	0.017	-0.022		0.016	0.020		0.024
Arthritis	-0.049	***	0.018	-0.010		0.013	0.006		0.013	0.050	**	0.020
Number of ADL limitation	-0.057	***	0.017	-0.032	***	0.012	-0.026	**	0.011	-0.107	***	0.014
Number of IADL limitation	0.011		0.019	0.005		0.014	-0.028	**	0.013	-0.032	*	0.017
Obese	-0.002		0.020	-0.037	**	0.015	-0.064	***	0.014	-0.102	***	0.020
Occupation *Ref: Managerial/Professional specialty*												
Technical/Sales/Administrative support	-0.029		0.025	-0.058	***	0.018	-0.057	***	0.017	-0.031		0.027
Service	0.021		0.030	-0.086	***	0.022	-0.105	***	0.022	-0.213	***	0.031
Farming/Fishing/Forestry	0.048		0.051	0.084	**	0.037	-0.033		0.036	-0.083		0.053
Precision production/Craft/Repair	-0.068	**	0.033	-0.030		0.023	-0.051	**	0.023	-0.088	***	0.034
Operators/Fabricators/Laborers	-0.033		0.032	-0.033		0.024	-0.106	***	0.023	-0.189	***	0.034
Elementary occupation	-0.075	*	0.042	-0.070	**	0.028	-0.069	***	0.027	-0.149	***	0.039
Disability Policy												
Coverage & Max. benefit level	**-0.053**	*******	**0.017**	**-0.033**	*******	**0.011**	**-0.042**	*******	**0.010**	**-0.037**	******	**0.015**
Disability levels for full disability	**-0.040**	*****	**0.022**	**0.124**	*******	**0.013**	**0.145**	*******	**0.012**	**0.030**	*****	**0.018**
Permanence of benefits	**-0.072**	*******	**0.013**	**-0.115**	*******	**0.007**	**-0.141**	*******	**0.007**	**-0.030**	*******	**0.010**
Medical assessment	**0.084**	*******	**0.016**	**-0.127**	*******	**0.010**	**-0.179**	*******	**0.009**	**-0.040**	*******	**0.012**
Vocational assessment	**-0.154**	*******	**0.021**	**-0.192**	*******	**0.014**	**-0.143**	*******	**0.013**	**0.034**	*****	**0.020**
Min. disability level & Sickness benefit & Unemployment benefit	**-0.058**	*******	**0.006**	**-0.075**	*******	**0.003**	**-0.070**	*******	**0.003**	**-0.031**	*******	**0.005**
Constant	-0.780	***	0.172	0.598	***	0.113	1.664	***	0.107	1.576	***	0.159

Note: We include in the estimation an indicator for missing occupation information. * p<0.10, ** p<0.05, *** p<0.01.

Table 4: Estimation Results for Work Limitation, With and Without Adjustment for Reporting Heterogeneity

Self-reported Work Limitation						
	Without Adjustment			With Adjustment		
	Coef.		S.E.	Coef.		S.E.
Demographics and Education						
Female	-0.004		0.044	0.060		0.047
Age56-60	0.104		0.066	0.102		0.072
Age61-65	0.097		0.064	0.114		0.070
Age66-70	0.106		0.065	0.122	*	0.071
Age70+	0.340	***	0.061	0.353	***	0.066
Years of education	-0.030	***	0.006	-0.048	***	0.007
Health						
High blood pressure	0.134	***	0.043	0.102	**	0.046
Diabetes	0.318	***	0.059	0.282	***	0.063
Cancer	0.386	***	0.067	0.450	***	0.072
Lung problems	0.587	***	0.075	0.620	***	0.081
Heart conditions	0.644	***	0.054	0.626	***	0.058
Arthritis	0.704	***	0.045	0.732	***	0.049
Number of ADL limitation	0.616	***	0.037	0.571	***	0.040
Number of IADL limitation	0.380	***	0.045	0.366	***	0.049
Obese	0.256	***	0.048	0.229	***	0.052
Occupation						
Ref: Managerial/Professional specialty						
Technical/Sales/Administrative support	0.038		0.061	-0.011		0.066
Service	0.249	***	0.075	0.190	**	0.081
Farming/Fishing/Forestry	0.010		0.122	0.045		0.132
Precision production/Craft/Repair	0.267	***	0.077	0.218	***	0.084
Operators/Fabricators/Laborers	0.326	***	0.080	0.268	***	0.087
Elementary occupation	0.158	*	0.092	0.067		0.100
Country						
Germany	0.527	***	0.078	-0.081		0.086
Sweden	0.361	***	0.089	-0.479	***	0.099
The Netherlands	0.144	*	0.081	-0.432	***	0.090
Spain and Italy	0.046		0.071	-0.702	***	0.078
France and Belgium	0.121	**	0.060	-0.485	***	0.065
Constant	-2.841	***	0.128	-2.392	***	0.138
Thresholds						
Cut-point 1	-2.220	***	0.016			
Cut-point 2	-1.200	***	0.014	See estimation results for thresholds equation (Table 3.1)		
Cut-point 3	-0.162	***	0.013			
Cut-point 4	1.113	***	0.014			
Vignettes						
Vignette 2	-1.250	***	0.019	-1.287	***	0.019
Vignette 3	0.240	***	0.019	0.240	***	0.019
Vignette 4	-0.998	***	0.019	-1.019	***	0.019
Vignette 5	-0.905	***	0.019	-0.920	***	0.019
Vignette 6	-1.486	***	0.019	-1.549	***	0.019
Vignette 7	-0.586	***	0.019	-0.587	***	0.019
Vignette 8	-0.504	***	0.019	-0.513	***	0.019
Vignette 9	0.204	***	0.019	0.219	***	0.019
Sigma	1.358	***	0.020	1.422	***	0.022
No. observations/Log likelihood	*6652*		*-85446*	*6652*		*-82658*

Note: The "Without Adjustment" results are estimated from HOPIT procedures without correcting for scale differences. The "With Adjustment" results are estimated from HOPIT procedures that allow for heterogeneous reporting scales. * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table 1: Correlation of Disability Policy Dimensions

	Policy 1	Policy 2	Policy 3	Policy 4	Policy 5	Policy 6	Policy 7	Policy 8	Policy 9	Policy 10
Policy 1	1.000									
Policy 2	0.344	1.000								
Policy 3	-0.346	-0.022	1.000							
Policy 4	0.741	0.443	-0.511	1.000						
Policy 5	0.154	-0.314	0.323	0.075	1.000					
Policy 6	-0.153	-0.591	0.431	-0.240	0.287	1.000				
Policy 7	-0.391	0.238	-0.098	-0.494	-0.576	-0.593	1.000			
Policy 8	0.253	0.771	0.158	0.048	-0.421	-0.262	0.234	1.000		
Policy 9	0.142	0.705	-0.397	0.193	-0.681	-0.751	0.731	0.473	1.000	
Policy 10	0.264	0.677	-0.521	0.305	-0.238	-0.862	0.516	0.435	0.770	1.000

Appendix Table 2: Correlation between Generosity of Each Disability Policy Dimension and Vignettes Classifying

	Percentage of Respondents Classifying the Vignettes as *Not At All Limited*	Percentage of Respondents Classifying the Vignettes as *Mildly or Moderately Limited*	Percentage of Respondents Classifying the Vignettes as *Severely or Extremely Limited*
Policy 1	-0.145	-0.145	0.164
Policy 2	-0.439	-0.094	0.192
Policy 3	0.060	0.091	-0.096
Policy 4	-0.139	-0.139	0.157
Policy 5	0.054	-0.122	0.095
Policy 6	0.381	-0.067	-0.035
Policy 7	-0.199	0.086	-0.027
Policy 8	-0.272	-0.021	0.086
Policy 9	-0.503	-0.091	0.205
Policy 10	-0.445	-0.156	0.249

Appendix: Empirical Model

Standard ordered regression models (e.g. ordered probit) are often used to analyze self-reported work limitation. Let h_i^s be a self-reported categorical work limitation measure, for individual i. It is assumed that h_i^s is generalized by a latent continuous limitation variable, h_i^{s*}. The ordered probit model assumes the following specification:

$$h_i^{s*} = x_i\beta + \varepsilon_i^s, \quad \varepsilon_i^s \sim N(0,1) \tag{1}$$

x_i is a vector of observed respondent characteristics, and ε_i^s is a random error term that is independent of x_i. The observed categorical response h_i^s relates to h_i^{s*} as follows:

$$h_i^s = k \qquad if\ \mu^{k-1} \le h_i^{s*} < \mu^k, \quad k = 1, \dots, 5 \tag{2}$$

$\mu^0 < \mu^1 \dots < \mu^5$, and $\mu^0 = -\infty\ and\ \mu^5 = +\infty$. The ordered probit model assumes that cut-points μ^k are constant. If this does not hold, in particular, if the cut-points vary according to the covariates x_i, then imposing this assumption will lead to biased estimates of the coefficients β in the latent health index. This is because they will reflect both health and reporting effects, the latter being effects of x_i on the cut-points. Vignette data can be used to model the cut-points as functions of respondent characteristics. These cut-points can then be imposed on the model for self-reported work limitation, making it possible to identify health effects rather than a mixture of health effects and reporting effects. This can be done using the Hierarchical Ordered Probit procedure or HOPIT (King et al., 2004).

The HOPIT model has two components: the vignette component reflects reporting behavior (i.e. it models the cut-points as functions of respondents' characteristics, thus allowing for reporting heterogeneity) and the health component representing the relationship between the respondent's own work limitation and covariates (with cut-points determined by the vignette component).

For the vignette component analysis, we use a set of vignettes in the HRS and the SHARE. Apart from measurement error, all respondents perceive each particular vignette *j* (*j* = 1, 2,…,9), to be consistent with the same latent level of work limitation, ruling out any association between h_{ij}^{v*} and an respondent's characteristics. Consequently, the latent work limitation of each vignette *j* as perceived by respondent *i* can be specified as an intercept plus random measurement error:

$$h_{ij}^{v*} = \alpha_j + \varepsilon_{ij}^v, \quad \varepsilon_{ij}^v \sim N(0,1) \tag{3}$$

We normalize $\alpha_1$to be zero, and ε_{ij}^v independent of each other and of x_i. The respective observed categorical rating h_{ij}^v is related to h_{ij}^{v*} through the following mechanism:

$$h_{ij}^v = k\ if\ \mu_i^{k-1} \le h_{ij}^{v*} < \mu_i^k, \quad k = 1, \dots, 5 \tag{4}$$

with $\mu_i^1 < \mu_i^2 \dots < \mu_i^5$, and $\mu_i^0 = -\infty\ and\ \mu_i^5 = +\infty$. The exclusion restriction in (3) allows us to identify the cut-points as functions of the respondents' characteristics:

$$\mu_i^k = \gamma_0^k + x_i\gamma^k + z_c\delta^k + c\eta^k, \quad k = 1, \dots, 4 \tag{5}$$

where x_i is a vector of a respondent's characteristics; z_c refers to country-level characteristics, including a country's disability policy generosity score; *c* is the country fixed effect. The effect δ^k reflects the differentials in response scales as the disability policy generosity score in a country increases by one point. A positive estimate of δ^k suggests that respondents under a more generous disability regime are more likely to evaluate a given vignette person as more severely work limited.

As in the standard ordered probit model, the second component of the HOPIT defines the latent level of the respondent's own work limitation, h_i^{s*} and the process that links this latent variable to the observed categorical variable, h_i^s. The difference is that the cut-points are no longer constants but can vary across respondents, and are determined by the vignette component of the model. Similar to the ordered probit, the second component of the HOPIT defines the latent level of individual own work limitation, h_i^{s*}, as:

$$h_i^{s*} = \beta_0 + x_i\beta + \varepsilon_i^s, \quad \varepsilon_i^s \sim N(0, \sigma^2) \tag{6}$$

We assume that ε_i^s is independent of x_i and ε_{ij}^v; and through similar mechanism as above the process this latent variable is linked to the observed work limitation severity categories as:

$$h_i^s = k \; if \; \mu_i^{k-1} \leq h_i^{s*} < \mu_i^k, \;\; k = 1, \ldots, 5 \qquad (7)$$

where $\mu_i^1 < \mu_i^2 \ldots < \mu_i^5$, and $\mu_i^0 = -\infty \; and \; \mu_i^5 = +\infty$. We assume that respondents use the same response scales to classify the vignettes and their own work limitation, so μ_i^k are equal to the cut-points in equation (5) of the vignette component.

Data Appendix: Vignette Questionnaires

Pain Vignettes:

1. [Yvonne] has almost constant pain in her back and this sometimes prevents her from doing her work.

2. [Catherine] suffers from back pain that causes stiffness in her back especially at work but is relieved with low doses of medication. She does not have any pains other than this generalized discomfort.

3. [Mark] has pain in his back and legs, and the pain is present almost all the time. It gets worse while he is working. Although medication helps, he feels uncomfortable when moving around, holding and lifting things at work.

Cardiovascular Vignettes:

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

2. [Tom] has been diagnosed with high blood pressure. His blood pressure goes up quickly if he feels under stress. Tom does not exercise much and is overweight.

3. [Dan] has undergone triple bypass heart surgery. He is a heavy smoker and still experiences severe chest pain sometimes.

Depression Vignettes:

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

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

3. [Henriette] generally enjoys her work. She gets depressed every 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.

Note:

[1] See Brandt et al., 2011 for a review and analysis of the U.S. federal disability evaluation process; See Garcia-Gomez et al., 2011 for the institutional arrangements of disability insurance in Spain and its recent reforms and interaction with other social programs.

[2] Most countries with such systems, including some Nordic (e.g., Sweden), western (e.g., Germany and the Netherlands), and central and eastern European countries, offer a full benefit to those assessed to be incapable of work as well as various degrees of partial benefits consistent with reduced work capacity. They offer one (e.g., Germany) to four different levels (e.g., Sweden) of partial benefits and in some cases offer finer gradations (e.g., the Netherlands). Other countries, including France and Spain, have a quasi-partial benefit for people who are unable to work in their usual occupation and a full benefit only for those unable to work in any occupation. Hence, the capacity threshold is the same for both benefits but the reference is different. The partial benefit is allowed to be supplemented to some extent by earnings from a job in another occupation. Even European countries that do not offer partial benefits as part of their main disability program, such as Belgium and Italy, have universal sickness programs as a precursor to their long-term disability program and also more alternative public programs to complement their disability program. For example, while the Italian disability pension scheme awards full benefits only to people totally unable to work, a means-tested disability allowance compensates for partial work capacity loss. Refer to OECD (2010) for a detailed discussion about OECD countries' recent experience and policy lessons in reforming their disability programs.

[3] Mitra (2009) provides a detailed analysis of partial disability program in 9 countries and great discussion about the advantages and disadvantages of such programs. Yin (2015) provides a detailed analysis of the incentives provided by the U.S. all-or-nothing disability system in contrast with the European partial disability system, simulating the effects of introducing partial benefits into the U.S. system on application behavior and employment.

[4] Livermore et al. (2013) provides a detailed summary of the main findings from evaluating the Ticket to Work Program. Weathers and Hemmeter (2011) describe the Benefit Offset National Demonstration pilot projects and present preliminary results.