Justification Bias in Self-Reported Disability:
New Evidence from Panel Data

Nicole Black, David W. Johnston, Agne Suziedelyte
Centre for Health Economics, Monash University, Australia

Abstract
The relationship between health and work is frequently investigated using self-assessments of disability from social surveys. The complication is that respondents may overstate their level of disability to justify non-employment. This study provides new evidence on the existence and magnitude of justification bias by exploiting a novel feature of an Australian longitudinal survey: each wave survey respondents are asked identical disability questions twice; near the beginning and end of the face-to-face interview. Fixed-effects regression models provide evidence of significant justification bias; especially for men and women who receive disability pensions. Additional analysis suggests mental illness is the most over-reported condition.

This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Social Services (DSS) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views reported in this paper, however, are those of the authors and should not be attributed to either DSS or the Melbourne Institute.

Corresponding author: David Johnston; david.johnston@monash.edu; +61 399058155; Centre for Health Economics, Building 75, Monash University, Clayton 3800, Victoria, Australia.

1. Introduction

Understanding the relationship between health and work is central to labor and health economics research and crucial for the design of health policies, social welfare systems, and strategies for productivity and growth. This relationship is often investigated using self-assessments of health and disability from social surveys. However, there exists a legitimate concern that thresholds for reporting a work-limiting disability may vary systematically according to individual circumstances (Kapteyn et al. 2007). In particular, individuals without a paid job may overstate their disability or health problems in order to justify their non-employment status. This so called 'justification bias' implies that the estimated importance of health and disability on labor supply decisions is most likely inflated. To more precisely measure the role of health in economic decision making, it is therefore critical that we ascertain the magnitude of justification bias and characterize the types of individuals for whom justification bias is largest. In this paper, we present new evidence on these issues. We additionally identify aspects of survey design that inadvertently exacerbate justification bias.

Justification bias has received considerable attention over the past three decades. In an early economics contribution, Parsons (1982, p.83) observed that "The self-rated poor health group will be composed of two distinct subsets: those who would rate themselves in poor health in an incentive-neutral environment, and those who are induced by the economic environment to declare themselves in poor health." As this reasoning suggests, it is possible that a proportion of survey respondents deliberately overstate their health-related work limitations because of financial incentives, such as qualifying for a disability pension. It is also possible that social context and psychological factors compel the non-employed to use illness to rationalize their inability to fulfil a socially prescribed role (Shuval et al. 1973). Such self-rationalization would lead non-workers to have a lower threshold for equating a health condition with a work limitation (Kreider and Pepper 2008). For example, non-workers may be significantly more inclined than workers to interpret back pain as a work-limiting disability.

Despite the long-running recognition and attention devoted to the issue of justification bias, there is conflicting evidence about its importance. In early investigations, Anderson and Burkhauser (1985, p.324) state "we are persuaded that self-reports of health are unsatisfactory measures", while on the other hand, Stern (1989, p.392) concludes that "standard disability measures are powerful and reasonably exogenous predictors of labour force participation". A decade later, Kerkhofs and Lindeboom (1995) and Kreider (1999) find substantial over-reporting of work limitations, whereas, Dwyer and Mitchell (1999) find no evidence in support of the justification hypothesis. More recently, Benítez-Silva et al. (2004, p.649) are "unable to reject the hypothesis that self-reported disability is an unbiased indicator", while in contrast, Baker et al. (2004, p.1090) find "evidence that the error in self-reported chronic conditions is related to labor market status", and the results in Lindeboom and Kerkhofs (2009, p.1042) "show that justification bias is substantial and that failing to account for this may change estimation results considerably". Further recent evidence on the importance of justification bias can be found in Gannon (2009), Datta Gupta and Larsen (2010), Datta Gupta and Jürges (2012), and Gosling and Saloniki (2014).

We contribute to the literature on justification bias by using an approach that differs in important ways from previous studies. First, we exploit a unique feature of an Australian longitudinal survey. In each wave, the disability status of certain respondents is self-reported twice using the same question – once at the beginning and once at the end of the face-to-face interview.[1] This question identifies disabilities or health conditions that have lasted six months or more, restrict everyday activity and cannot be corrected by medication. The second disability question is, however, preceded by a series of detailed questions about employment, occupation, job search, income and pension receipt. Therefore, it is likely that respondents are inadvertently 'primed' to consider these issues when reporting disability the second time. How survey design can induce or exacerbate misreporting of health and disability has received little acknowledgement in the justification bias literature, but it has been shown that responses to life evaluation questions are extremely sensitive to question-order effects (Deaton, 2012). Priming has also been used in economic experiments to increase the salience of certain concepts and issues (Benjamin et al., 2010; Callen et al., 2014; Cohn et al., 2015).[2]

The second novel feature of our approach is that we fully utilize the panel dimension of our data by estimating fixed-effects (FE) regression models. Essentially, we investigate how changes over time in the variation between the two self-reported disability measures correlate with changes over time in employment status. This modelling approach allows us to control for all time-invariant factors that influence reporting behavior, such as cognitive ability. Failing to control for these factors is likely to upward bias the estimated magnitude of justification bias,

[1] Previous studies have exploited repeated health questions in surveys to investigate reporting bias and heterogeneity; see Crossley and Kennedy (2002), Clarke and Ryan (2006) and Lumsdaine and Exterkate (2013). In these studies the questions regard general health (rather than disability status) and either the survey mode (i.e. face-to-face versus self-completion), question wording, or available response options differ between the two survey questions. In addition, none of the studies use longitudinal data.

[2] For example, Cohn et al. (2014) increases the salience of bank employees' professional identity by asking several questions about their professional background, such as "At which bank are you presently employed?" and "What is your function at this bank?".

because individual characteristics associated with greater reporting heterogeneity are positively associated with non-employment. Results show that ordinary least squares (OLS) estimates of justification bias are indeed substantially larger than FE estimates.

The FE results demonstrate that non-employed respondents and disability pension recipients are significantly more likely to misreport or exaggerate their level of disability. For example, we find that conditional on the responses to the disability question at the beginning of the interview, unemployed and out of labor force (OLF) males are 2.7 and 5.4 percentage points respectively more likely to report a disability at the end of the interview than are employed males. The corresponding effects are smaller for females (2.0 and 2.1 percentage points respectively). This tendency to justify one's non-employment status with poor health is substantially greater for those receiving a disability support pension. We also find that the degree of justification bias appears to be largest for those respondents who then subsequently report a mental health condition.

2. Data and Methods

2.1. The Household, Income and Labour Dynamics in Australia (HILDA) Survey

We use data from the HILDA Survey, a nationally-representative longitudinal study of Australian households that began in 2001. Wave 1 contained a sample of 19,914 panel members from 7,682 households, and in each year since members of these households have been followed-up, along with new household members resulting from changes in the composition of the original household and new households from the wave 11 top-up sample. Annual data is currently available from 2001 to 2012, and each year includes detailed information on income, employment, health and other demographic and socio-economic information.[3] The survey comprises three face-to-face survey instruments – Household Form, Household Questionnaire, and Person Questionnaire – and a confidential self-completion questionnaire.

Disability status is first elicited in the Household Form, which we label throughout as Questionnaire 1 (Q1). Q1 is an initial face-to-face questionnaire designed to record basic information about each member of the household before commencing the detailed questionnaires. It is administered to one member of the household, which can vary from wave-

[3] The household response rates range from 87.0 per cent in wave 2 to 70.8 per cent in wave 11, while the household response rates for those households responding in the previous wave ranges from 87.0 per cent in wave 2 to 96.4 per cent in wave 11 (Summerfield et al., 2012).

to-wave, and takes on- average 6 minutes to complete. Most significantly, the Q1 respondent is asked “does anyone here have any long-term health condition, disability or impairment such as these?”, and is shown a card with the description “Disabilities/health conditions which have lasted, or are likely to last, 6 months or more; restrict everyday activity; and cannot be corrected by medication or medical aids” followed by a list of 17 types of disability.[4] The Q1 respondent answers “yes” or “no” for all household members.

Disability status is subsequently elicited in the Person Questionnaire, which we label throughout as Questionnaire 2 (Q2). Q2 is the main survey instrument and is administered face-to-face to every member of the household aged 15 years and over. It takes on-average 33 minutes to complete and contains sections on family background, education, employment, income, family formation, and health, with the health section occurring near the end of the questionnaire. In the very beginning of the health section, respondents are presented with a disability card identical to that used in Q1 (same description and list of disability types), and are asked whether they have any long-term health condition, impairment, or disability. Again, they can only answer “yes” or “no”. If the respondent answers “yes”, they are then asked which of the 17 disabilities they have (multiple types can be provided) and whether the condition limits the “type of work or the amount of work you can do?” These follow-up questions were not asked in Q1.

Within each wave of HILDA, each respondents’ disability status is therefore measured twice; through either two self-reports (Q1 and Q2), or through a partner-report (Q1) and a self-report (Q2).[5] We are unaware of any comparable data sets that repeatedly ask identical disability or health questions, especially not consistently across waves. HILDA data therefore provide a unique opportunity to investigate reporting heterogeneity in self-reported disability.

Table 1 presents the proportions of individuals aged 18-60 who have a disability according to Q1 and Q2, separately for the sample of respondents for whom Q1 is self-reported (Sample A) and for the sample of respondents for whom Q1 is partner-reported (Sample B).

---

[4] The 17 disability types are: Sight problems not corrected by glasses or contact lenses; Hearing problems; Speech problems; Blackouts, fits or loss of consciousness; Difficulty learning or understanding things; Limited use of arms or fingers; Difficulty gripping things; Limited use of feet or legs; A nervous or emotional condition which requires treatment; Any condition that restricts physical activity or physical work (e.g. back problems, migraines); Any disfigurement or deformity; Any mental illness which requires help or supervision; Shortness of breath or difficulty breathing; Chronic or recurring pain; Long-term effects as a result of a head injury, stroke or other brain damage; A long-term condition or ailment which is still restrictive even though it is being treated or medication is being taken for it; Any other long-term condition such as arthritis, asthma, heart disease, Alzheimer’s disease, dementia etc.

[5] Other combinations also occur. For instance, it is common for the parent of a young person to be the Q1 respondent. We do not use these combinations in our analyses.

Note that many respondents (42% of women and 60% of men) appear in both samples, as the household member who completes Q1 can vary across waves. The disability rates are also presented for three employment states: employed (E), unemployed (U) and out-of-the-labor-force (OLF). A respondent is defined as unemployed if they want to work, and are actively looking for work or available to start work within four weeks. The out-of-the-labor-force category includes respondents who do not want to work, and respondents who want to work but are not actively looking and are not available to start work. This category includes persons who are retired, homemakers, carers, disabled, travelling / on holiday, and volunteers.

The summary statistics reveal several interesting features of the data. First, self-reported disability rates are around 20 percentage points for both males and females, which is similar to the United States (Kreider and Pepper, 2008). The self-reported rates are slightly larger for Sample A (who self-reported Q1) than for Sample B (whose partner reported Q1), reflecting the fact that the samples are non-randomly selected. Second, Q1 disability rates are lower than Q2 disability rates, especially when Q1 disability is partner-reported. Third, the differences between Q1 and Q2 disability rates are larger for unemployed and out-of-the-labor-force respondents than for employed respondents. For example, the percentage point differences for men in Sample A equal 1.5, 3.9 and 4.5 for employed, unemployed and OLF, respectively. The corresponding figures for women in Sample A equal 1.6, 3.3 and 3.3 percentage points. Overall, the data indicate that respondents are inconsistent in their answers to survey disability questions, and that the inconsistency is higher for the non-employed.

Figure 1 additionally highlights the inconsistency between the Q1 and Q2 disability measures by presenting nonparametric kernel regression estimates of the proportions of OLF men and women with a reported disability by age (using Sample A). The graphs show that a significantly higher proportion of OLF respondents report a disability in Q2 than in Q1; though, this difference is much larger for males than females. For males the difference peaks at around 40 years of age and equals almost 10 percentage points. This remarkably large gap generated by differential responses to identical questions asked on the same day, demonstrates that reporting bias and heterogeneity in self-reported disability is an important issue; especially considering that population disability rates are fundamental to the design and budgeting of health and social welfare policies.

2.2. Identifying Justification Bias

Our main aim is to evaluate the extent of justification bias in responses to the disability question appearing in Questionnaire 2 (Q2), which measures serious, long-term, untreatable conditions.

This question is administered individually to each adult member of the household and constitutes the HILDA survey's most important measure of ill-health and disability. The disability question is relatively standard, and is comparable to the measures contained in the Health and Retirement Study (HRS), Panel Study of Income Dynamics (PSID) and Survey of Income and Program Participation (SIPP), and as in each of these surveys, the Q2 disability question appears after questions on employment outcomes and welfare receipt.

Justification bias is identified by comparing the variation between Q1 self-reported disability status and Q2 self-reported disability status over time, with the variation in employment status over time. Variation between self-reports in Q1 (recorded at the beginning of the survey) and Q2 (recorded at the end of the survey) is most likely generated by survey priming effects – the process in which respondents are (inadvertently) primed to think about an issue, representation or association while answering subsequent questions. In HILDA, the survey modules between the two disability measures include multiple questions on employment history, job search, reasons for not working, reasons for not searching for work, and pension receipt. Therefore, it is likely that respondents are considering these issues when answering the disability question for the second time. This consideration may prompt some respondents to deliberately misreport their disability status in Q2; for example, those respondents who are fraudulently collecting disability-related welfare payments.

Another potentially more frequent mechanism is that answering detailed questions about employment and welfare receipt changes respondents' thresholds for equating poor health with a disability. For example, a non-employed respondent suffering migraines may not usually consider themselves as having a disability that "restricts everyday activity". Their assessment, however, changes after they have been asked to rationalize their non-employment spell (i.e. their threshold decreases). Conversely, an employed respondent may initially assess their back pain as restricting their everyday activity, but after describing their employment history, work hours, training activities, and tasks, duties and responsibilities, they change their assessment (i.e. their threshold increases). This proffered "self-rationalization" mechanism is based partially on the concept that "disability" is not an objective binary health state, but more so a categorization that is based on self, doctor, or government evaluations and definitions. As Autor and Duggan (2006; p.85) write: "While certain medical conditions are clearly disabling, "disability" is not a medical condition. Disability is a dividing line (or zone) chosen by policymakers on a continuum of ailments affecting claimants' capability to engage in paying work and their pain and discomfort in doing so", and "Beyond the subset of clearly incapacitating medical and mental disorders, the extent of "disability" is ultimately a variable determined by policy." In this paper, following the convention in the literature, we label both the variation generated by deliberate misreporting and the variation generated by self-rationalization (changing thresholds) as justification bias.[6]

Another possibility is that the variation between self-reports in Q1 and Q2 is caused by an initial reluctance by respondents in Q1 to reveal ill-health and disability (e.g. due to stigma effects), which is partially overcome in Q2 after the respondent has become more relaxed and trusting of the interviewer. Robustness models using more 'experienced' respondents and models including additional health controls from a confidential self-completion questionnaire (reported in Section 3), suggest that this alternative explanation does not explain our results.

The main advantage of our identification strategy is that the two self-reports are based on identical face-to-face survey questions measured on the same day, and therefore differences between them cannot be generated by differences in an individual's true multidimensional health status. Such differences may arise, for example, if we compared self-reported disability status with self-reported general health, or if we compared self-reports measured in different weeks or from different survey modes (face-to-face versus self-completion), or if we compared self-reports with administrative records of health utilization data. The second advantage is that by exploiting within-individual variation over time, we can control for unobservables that are associated with employment status and the variation between Q1 and Q2 disability measurements. For example, cognitive ability, English language skills, personality and culture likely influence the comprehension and interpretation of the question clauses (i.e. "restrict everyday activity" and "cannot be corrected by medication or medical aids") and also influence employment outcomes.

A disadvantage of this identification strategy is that it will generate an under-estimate of justification bias if a non-zero proportion of 'justifying' respondents do so consistently in both Q1 and Q2. Another disadvantage is that it relies upon a non-random sub-sample of survey respondents who self-reported Q1. This sub-sample is more likely to be non-employed and to be disabled (see Table 1), and the non-random selection may therefore generate a sample selection bias. To investigate these possible biases, we report estimates from samples of single men and women, for whom Q1 was necessarily self-reported. We also conduct an analysis that exploits the variation between Q1 partner-reported disability status and Q2 self-reported disability status. We contend that partners (cohabitating partners and married spouses) are

[6] See Bound and Burkhauser (1999) and Kreider and Pepper (2008) for further discussion of how self-rationalization may lead to the misreporting of disability status.

ordinarily aware of serious, long-term, untreatable conditions suffered by their partners, and are less likely to justify their partners' non-employment than their partners themselves.[7] In addition, a significant proportion of respondents with partner-reported Q1 measurements also have self-reported Q1 measurements (in different waves). This means we are able to compare estimates for the same respondents derived using self-reported Q1 and partner-reported Q1 to gauge the magnitude of any sample selection bias.

Figure 2 illustrates the extent of variation between the self-reported Q1 and Q2 measures. Specifically, Figure 2A presents the probability of reporting a disability in Q2, conditional on not reporting a disability in Q1: $Pr\left(D_{it}^{Q2} = 1 \middle| D_{it}^{Q1} = 0\right)$. If reporting was consistent across Q1 and Q2, we would expect the rates to be near zero, but instead we observe a remarkable pattern. Figure 2A shows that the propensity for male respondents to change their assessment and report a disability in Q2 increases sharply with non-employment. Conditional on a previous self-report of no disability in Q1, the Q2 disability rates equal 5% for employed, 11% for unemployed, and 21% for OLF. In contrast, the gradient is nearly flat for women. The male pattern suggests that having to rationalize non-employment significantly decreases non-employed men's thresholds for equating poor health with a disability.

Figure 2B presents the opposite variation; the probability of not reporting a disability in Q2, conditional on reporting a disability in Q1: $Pr\left(D_{it}^{Q2} = 0 \middle| D_{it}^{Q1} = 1\right)$. The figure again shows a steep gradient, but in the reverse direction, and for both men and women. Conditional on a previous self-report of an existing disability in Q1, the probability of *not* reporting a disability in Q2 equals 20% for employed men, 7% for unemployed men and 3% for OLF men. The corresponding figures for women equal 17%, 9% and 6%. These patterns suggest that having to detail employment conditions, significantly increases employed respondents' thresholds for equating poor health with a disability.[8]

2.3. Regression Modelling Approach

As discussed previously, to more rigorously examine the extent of justification bias, we wish to exploit the variation between Q1 self-reported disability status and Q2 self-reported

[7] While individuals may have an incentive to present themselves in a more socially acceptable manner in front of an interviewer, a proxy respondent is less likely to have the same incentive. A proxy respondent may certainly report the health status of another individual with error, but it is less likely that the error would be systematically related to employment status. We therefore expect that spouse-reports are less likely to suffer from justification bias than self-reports.

[8] These descriptive results cast doubt on the assumption made in previous studies (e.g. Kreider, 1999; McGarry, 2004) that employed respondents consistently provide accurate assessments of their disability status.

disability status within-individuals over time. We do this by estimating linear fixed-effects regression models:

$$D_{it}^{Q2} = \beta D_{it}^{Q1} + NE'_{it}\gamma + X'_{it}\delta + \alpha_i + \mu_t + \varepsilon_{it} \quad (1)$$

where D_{it}^{Q2} is disability status from Q2, D_{it}^{Q1} is disability status from Q1, NE_{it} is a vector of non-employment states, X_{it} is a vector of additional control variables, α_i is an individual-level fixed-effect , μ_t is a time fixed-effect, and ε_{it} is a random error term. Statistically significant positive effects of non-employment on self-reported disability ($\hat{\gamma} > 0$) are then interpreted as evidence of justification bias.

We estimate equation (1) using fixed-effects model to account for the potential correlation between employment status and unobserved heterogeneity. In this model, the parameter of interest γ is identified from the variation in an individual's employment status over time. For comparison, we also present ordinary least squares (OLS) estimates of equation (1). These OLS estimates are expected to be biased upwards, given that time-invariant characteristics that increase reporting heterogeneity, such as low cognitive ability and poor English language skills, are positively correlated with non-employment.

We also present variants of equation (1) that condition directly on the value of D_{it}^{Q1}, rather than including D_{it}^{Q1} as a covariate, broadly replicating the modelling approach used by Baker et al. (2004). Specifically, we separately estimate linear regression models of D_{it}^{Q2} for the sample with $D_{it}^{Q1} = 0$ and the sample with $D_{it}^{Q1} = 1$. We additionally present estimates from multinomial logit models of the four mutually exclusive outcomes: (1) $D_{it}^{Q1} = 0, D_{it}^{Q2} = 0$ (no-no); (2) $D_{it}^{Q1} = 1, D_{it}^{Q2} = 1$ (yes-yes); (3) $D_{it}^{Q1} = 0, D_{it}^{Q2} = 1$ (no-yes); and (4) $D_{it}^{Q1} = 1, D_{it}^{Q2} = 0$ (yes-no). The results from these two alternative modelling approaches demonstrate the robustness of our main results. [9]

For our methodology to be valid, we must assume that the relationship between non-employment and unobserved true disability (D_{it}^{T}) is captured by Q1 disability and other controls. That is, conditional on D_{it}^{Q1} (and X_{it} and α_i), an individual's non-employment provides no

[9] We do not present estimates of a modified equation (1) that restricts $\beta = 1$, or in other words, models the difference between D_{it}^{Q2} and D_{it}^{Q1}. This restriction holds under the null hypothesis of 'rational unbiased reporting' of disability status (Benítez-Silva et al., 2004), but does not hold when the conditional probability to report a disability in Q1 differs from the conditional probability to report a disability in Q2.

additional information regarding their true disability status (Kerkhofs and Lindeboom, 1995). More specifically, we rely on a conditional independence assumption, which implies:

$$pdf\left(D_{it}^{T} \middle| D_{it}^{Q1}, X_{it}, \alpha_i, NE_{it}\right) \equiv pdf\left(D_{it}^{T} \middle| D_{it}^{Q1}, X_{it}, \alpha_i\right) \quad (2)$$

where for notational simplicity we conflate the binary observed measurements of disability with the continuous latent constructs that underlie them. Crucially, the identifying assumption can hold even if D_{it}^{Q1} is observed with error. What is vital is that the error is not related to the employment status of the respondent (Lindeboom and Kerkhofs, 2009). As mentioned above, an example of 'problematic' error is if the respondent consistently justifies non-employment in both Q1 and Q2. In this instance, the employment status coefficients ($\hat{\gamma}$) would be downward biased (i.e. provide lower bound estimates), and can be interpreted as the extent to which justification bias in Q2 disability is larger than in Q1 disability.

To help demonstrate the validity of our methodological approach, we present in Section 3 a number of sensitivity tests. Specifically, we examine the sensitivity of our key coefficients to: (i) using additional self-reported health information from HILDA's self-completion questionnaire; (ii) the inclusion of lagged Q2 disability measures (estimated using GMM dynamic panel data models); and (iii) using modified outcome variables representing work and non-work limiting disabilities, and types of illnesses/conditions. As aforementioned, we also present results from models that compare self- and partner-reports rather than two self-reports.

3. Results

3.1. Estimated Effects of Non-Employment on Self-Reported Disability Status

Estimates of equation (1) are reported in Table 2, separately by gender. Columns (1) and (3) report estimates for the sample of respondents from linear regression models without individual fixed-effects, and columns (2) and (4) report estimates from linear regression models with fixed-effects. Included in each regression model is non-employment status (unemployed, OLF), self-reported disability status from Questionnaire 1 (Q1) and a set of unreported control variables: a quadratic function in age, marital status, having dependent children, interview circumstances[10], and year effects.

Column (1) shows that conditional on self-reported disability from Q1, unemployed men are 8.4 percentage points and OLF men are 14.0 percentage points more likely to report a disability in Questionnaire 2 (Q2) than are employed men. The corresponding estimates for women in Column (3) equal 5.7 percentage points and 6.2 percentage points. These results are comparable to the raw male and female differences seen in Figure 2, which are also largely driven by between-individual variation. The within-individual estimates presented in Columns (2) and (4) are considerably smaller. The estimates for unemployed men and OLF men drop to 2.7 percentage points and 5.4 percentage points, respectively. While the estimates for unemployed and OLF women equal 2.0 percentage points and 2.1 percentage points.

The set of estimates in Table 2 provide several interesting results. First, individual fixed-effects are important controls; time-invariant characteristics that are associated with employment status (e.g. cognitive ability), also seem to be important determinants of reporting heterogeneity in disability. This finding has implications for previous research that has relied upon between-individual variation to identify justification bias and other determinants of misreporting. Second, the fixed-effect estimates suggest that for men there is considerable justification bias: a 5.4 percentage point increase in the propensity to report a disability equates to a 24.7% effect relative to the disability rate for all men (22%) and a 7.4% effect relative to the disability rate for OLF men (73%). Notably, this effect may be downward biased, given the potential for employed men to under-report and non-employed men to over-report in both Q1 and Q2. Third, statistically significant effects for women, though small, are relatively unique to the literature, with most previous studies using data that is pooled across genders or that includes only men.

3.2. Alternative specifications and robustness checks

Appendix Tables A1 and A2 present results from two alternative modelling approaches. First, we condition directly on Q1 disability status rather than including Q1 disability as a covariate, and estimate linear regression models of Q2 disability (see Table A1). Estimates from these models are consistent with our main results in Table 2, although the size of the effects are

[10] In particular, we include indicator variables for the respondent being suspicious about the study; the respondent being uncooperative; the respondent having poor eyesight, hearing problems, reading difficulties, English being second language, or other language problems; being interviewed in follow-up fieldwork period; and needing more than four calls to complete all the interviews.

slightly smaller under the fixed-effects estimator. Second, we present results from multinomial logit models of the four mutually exclusive responses to the two disability questions (yes-yes, no-no, no-yes, yes-no) (see Table A2). These results also show that non-employed men and women are significantly more likely to answer "no" on Q1 and "yes" in Q2 than their employed counterparts, which corresponds with our main findings. The estimated effects on the converse yes-no outcome (columns 4 and 8) are small and statistically insignificant. Likewise, in Table A1 (columns 4 and 8) among those who reported a disability in Q1, there is no difference in reporting a disability in Q2 by employment status. These findings suggest the main results (Table 2) are driven from differences by employment status in reporting a Q2 disability after not reporting a Q1 disability, rather than vice-versa.

To further test the robustness of the estimates in Table 2, we estimate fixed-effect models using subsamples of single respondents. For singles, Q1 is necessarily self-reported and therefore the estimated non-employment effects cannot be driven by bias from non-random selection of Q1 respondents. The estimates for singles are presented for males and females in columns (1) and (4) of Table 3, respectively, and are close in magnitude to those from Table 2. Conditional on self-reported disability from Q1, single OLF men are 4.9 percentage points and single OLF women are 3.8 percentage points more likely to report a disability in Q2 than are employed men and women.

Our identification strategy may still under-estimate justification bias if respondents tend to justify their employment status consistently in both Q1 and Q2. To test for this, we compare self-reported disability measures for partnered respondents (married or cohabitating) with estimates generated using partner-reported disability from Q1 and self-reported disability from Q2. Because partner-reports of disability are much less likely to suffer from justification bias, this allows us to gauge the extent to which our main estimates may be under-estimated. Columns (2) and (5) of Table 3 report male and female fixed-effects estimates for partnered respondents for whom Q1 is self-reported; and columns (3) and (6) of Table 3 report fixed-effects estimates for partnered respondents for whom Q1 is partner-reported. The estimated effects for unemployed and OLF are remarkably similar across the two specifications, for both genders, which suggests the downward bias of the estimates in Table 2 may be small. Or in other words, employed men are unlikely to under-report and non-employed men are unlikely to over-report their disability status unless they have been primed to think about their employment status.[11] Another interesting feature of the results in Table 3 is that the estimated effects for single men are smaller than the estimated effects for partnered men, and the reverse is true for women. This suggests there may be heterogeneity by marital status.

While priming is the most likely explanation for the variation in responses to Q1 and Q2 disability questions, it is also possible that the variation may be driven by an initial reluctance by some respondents to reveal any disability early in the interview (in Q1), which is then overcome by Q2. To test for this, we re-estimate the fixed-effect models excluding observations from the first three surveys for each respondent. This allows the effect to be based on surveys where respondents are more likely to trust the HILDA survey and therefore more comfortable throughout the interview process. Table A3 shows that the effects are incredibly similar if we constrain the sample to the more 'experienced' respondents, suggesting it is unlikely our results are driven by initial distrust.

We also test the robustness of the estimates in Table 2 by adding additional health measures as covariates to the fixed-effects regression models. As discussed in Section 2.3, for our empirical approach to be valid, we must assume that conditional on the included covariates an individual's employment status provides no information regarding true disability status. The validity of this assumption is therefore strengthened by the inclusion of additional health information. First, we include cubic functions of respondents' SF-36 physical functioning and mental health scores obtained from HILDA's confidential self-completion questionnaire. Second, we exploit the panel nature of our data set and include the Q2 self-reported disability measure from the previous wave. This model is estimated using a system GMM specification that contains the disability equation in levels and differences (Blundell and Bond, 1998). In this method, lagged first differences are used as instruments for the equation in levels and lagged levels are used as instruments for the equation in first differences. Importantly, these additional health variables are themselves likely to contain justification bias as they are recorded after questions on employment status and welfare receipt, and it is therefore expected that estimates from these robustness models will be smaller than those previously presented. Nevertheless, if the estimated non-employment effects are generally robust to these additional

[11] In some waves Questionnaire 1 includes a single question on employment prior to the disability question ('What is your current employment status?'). Using the variation across waves in the appearance of this question we are able to test whether the results are sensitive to its inclusion. When we estimate equation (1) separately for waves 1-2, in which the employment question was not included, and for waves 3-4, in which the employment question was included, the estimated coefficients on unemployed and OLF are not statistically different from one another using these two samples. This suggests that the inclusion of this single employment question did not sufficiently 'prime' the respondents when answering Q1.

controls, it is a strong indication that our estimates are not driven by some unobserved element of health that is associated with employment status.

For comparative purposes, Column (1) of Table 4 reports the regression models from Table 2 for the sample that has non-missing SF-36 physical functioning and mental health scores. Column (2) additionally includes the cubic SF-36 score functions. Overall the estimated non-employment effects change very little for men. For women, the estimated effects are moderately affected and the OLF estimate loses statistical significance. If we omit Q1 disability from the model and only control for true disability status using the SF-36 cubic functions – an approach similar to previous studies in the literature – the male unemployed and OLF estimates rise to 4.4 percentage points and 9.2 percentage points, respectively. In column (4), the inclusion of lagged Q2 disability reduces the size of the estimates, but the male OLF estimate remains statistically significant at the 10%-level, and indicates that conditional on Q1 self-reported disability and Q2 self-reported disability from last year, respondents who are out-of-the-labor-force are 3.6 percentage points more likely to report a disability in Q2 than are employed respondents. The comparability of estimates across models – particularly those reported in Columns (1), (2) and (4) – are supportive of our assumption that the relationship between non-employment and unobserved true disability is captured by Q1 disability and other controls.

3.3. Results Disaggregated by Type of Disabling Condition

After the main disability question in Q2, respondents are asked which of the 17 conditions they have and whether the condition limits the "type of work or the amount of work you can do?". We use the answers to these two follow-up questions to disaggregate the disability outcome variable in to: (i) work limiting and non-work limiting conditions (Table 5); and (ii) seven groups of condition types (Table 6). If respondents are primed by the detailed questions regarding employment and welfare receipt then the estimated effects of non-employment on work limiting conditions should be large and the effects of non-employment on non-work limiting conditions should be small.

The results in Table 5 indicate this is the case. The estimated effects of non-employment on the probability of reporting a non-work limiting condition are small and statistically insignificant; whereas the estimated effects of non-employment on the probability of reporting a work-limiting condition are large and statistically significant at the 1% level. OLF men and women are estimated to be 8.1 percentage points and 3.5 percentage points more likely to report a work limiting condition, respectively, than are employed men and women.[12] Note also that the estimated effects of self-reported Q1 disability are significantly larger in the non-work limiting regression models than in the work-limiting regression models; reflecting the fact that respondents are more likely to be consistent between Q1 and Q2 when their reported disability is unrelated to work.

Table 6 reports the estimated effects for seven outcomes representing disability caused by: (1) mental health condition; (2) limited use of limbs; (3) chronic or recurring pain; (4) hearing and sight problems; (5) other specified conditions (all with low frequency); (6) unspecified restrictive condition; and (7) any other unspecified condition (see Table 6 note for further details). Column (1) shows that conditional on self-reported disability from Q1, unemployed men are 2.0 percentage points and OLF men are 7.0 percentage points more likely to report a mental health condition than are employed men. Relative to sample mean levels, these effects are far larger than for any other condition type. The non-employment effects are also statistically significant for 'limited use of limbs', 'chronic or recurring pain', 'other specified conditions', and 'unspecified restrictive condition'.[13] The conditions for which there are no non-employment effects are hearing and sight problems (Column 4) and the catch-all category 'any other unspecified condition' (Column 7).

For females the estimated effects are also largest for the 'mental health condition' category: unemployed women are 2.5 percentage points and OLF women are 2.9 percentage points more likely to report a mental health condition than employed women (conditional on Q1 disability). The estimated effects for the categories 'limited use of limbs' and 'unspecified restrictive condition' are also large for women, largely mirroring the results for men.

The comparatively large effect sizes on the 'mental health condition' category, which is comprised of the two conditions "a nervous or emotional condition which requires treatment" and "any mental illness which requires help or supervision", are particularly interesting given the very large increases over time in the proportions of individuals receiving disability pensions for mental ill-health. The explicit role of disability pension receipt in generating justification bias is investigated in the next sub-section; but briefly, the proportion of Australian Disability Support Pension recipients listing a psychological or psychiatric condition as their primary

[12] The non-work limiting outcome variable equals one if the respondent reports a non-work limiting disability and zero if the respondent does not have a disability. Similarly, the work-limiting outcome variable equals one if the respondent reports a work limiting disability and zero if the respondent does not have a disability.

[13] Results for the 17 conditions estimated separately (shown in Appendix Table A4) indicate that the estimated effects for 'shortness of breath or difficulty breathing' and 'blackouts, fits or loss of consciousness' are driving the 'other specified conditions' effect.

medical condition has risen from 23% in 2001 to 31% in 2013.[14] The growth in disability pension receipt for mental ill-health has also been documented in other developed countries. For example, Autor and Duggan (2006) show that in the US the proportion of Disability Insurance (DI) Awards for the diagnosis group 'mental disorders' has risen from 16% in 1983 to 25% in 2003 (see their Table 1).

3.3. Results Disaggregated by Disability Welfare Receipt

In this subsection we investigate the role of disability pension receipt on the propensity to report a disability in Q2, conditional on Q1 disability status. Between the Q1 and Q2 disability measures, respondents are asked whether they currently receive the Disability Support Pension (DSP), how much they received in their most recent payment, whether they received the DSP at any time during the last financial year, how many weeks in the last financial year they received the DSP, and how much they received in total during the last financial year. As background information, in Australia 16-64 year-olds are eligible for the DSP if they have a physical, intellectual, or psychiatric condition that prevents them from working 15 hours or more per week within the next 2 years. In our sample time-frame, eligibility was based on a report from the claimant's doctor; though, a recently proposed policy change will entail new DSP claimants to be assessed by a government-appointed doctor. The maximum fortnightly DSP payment for singles without children equals $782, which is substantially higher than the unemployment benefit ($519).[15] Finally, DSP recipients may work up to 30 hours per week and continue to receive a part pension.

To examine the impact of receiving DSP (or more accurately reporting the receipt of DSP) on the propensity to report a disability in Q2, we disaggregate the employed, unemployed and OLF variables by DSP receipt, and re-estimate linear fixed-effects regression models, with 'Employed without DSP' used as the omitted category. Results are reported in Table 7 separately by gender and for two outcome variables: any disability (as per Tables 2-4) and work-limiting disability (as per Table 5). The latter outcome is added given the earlier results that demonstrate that only work-limiting disabilities are 'over-reported'.

[14] The top 5 most commonly claimed for conditions in 2013 are: (1) psychological / psychiatric (31%); (2) musculo-skeletal and connective tissue (26%); (3) intellectual / learning (12%); (4) nervous system (5%); and (5) circulatory system (4%). The distribution of primary medical conditions is similar for both sexes. For more information see the 2013 report by the Australian Government Department of Social Services on "Characteristics of Disability Support Pension Recipients", available at www.dss.gov.au.

[15] For more information see www.humanservices.gov.au/customer/services/centrelink/disability-support-pension and http://www.humanservices.gov.au/customer/services/centrelink/newstart-allowance.

The estimates in Table 7 indicate that male employed, unemployed and OLF respondents receiving the DSP are 9.1 percentage points, 13.3 percentage points and 12.7 percentage points more likely than employed respondents not receiving the DSP to report a Q2 disability, conditional on Q1 disability status. Equivalent figures for a work-limiting disability equal 14.3 percentage points, 17.9 percentage points and 18.1 percentage points. In both instances, the estimates for these three groups are not significantly different from one another (F-test p-values equal 0.153 and 0.226). Similar results exist for females. Unemployed and OLF DSP recipients are 10.3 percentage points and 9.7 percentage points more likely to report any disability, and are 15.1 percentage points and 14.0 percentage points more likely to report a work-limiting disability. The estimated effect for employed women receiving the DSP is much smaller; though the sample size for this category is the smallest of all those presented (only 0.7% of the female sample are employed and receiving the DSP). These results suggest that reporting DSP receipt between the Q1 and Q2 disability measures is an important driver of 'over-reporting' disability.

Importantly, the estimates are also statistically significant for unemployed and OLF men and women who do not receive the DSP. Given that DSP payments are substantially higher than unemployment benefit (UB) payments, and given the onerous 'mutual obligation' requirements required of UB recipients (Saunders, 2007), individuals who are unemployed or OLF and truly disabled have a strong incentive to apply for the DSP. Therefore, it is highly likely that a non-trivial proportion of the non-working non-DSP recipients who are reporting a disability in Q2 but not Q1 are not disabled; indicating that having to discuss (and justify) one's non-employment status sufficiently primes respondents to 'over-report' their disability status.

4. Conclusion

The last two decades has seen the number of disability pension recipients more than double in Australia (Broadway et al. 2014), with similarly worrying trends in the United States (Liebman 2015). Notably, the increase has been especially large for hard-to-verify impairments such as back pain and mental health problems (Liebman 2015). Consequently, the Australian government has recognized that "many new applications for the disability pension are not triggered by the acquisition of an impairment or disability, but by changes in an individual's employment circumstances" (Macklin 2009). Our study provides new evidence on the directly related issue of justification bias; the tendency for survey respondents to exaggerate their level of disability in order to justify non-employment. Of particular interest is how increasing the salience of respondents' employment status and pension receipt may lower a non-employed respondents' threshold for equating poor health with disability.

This paper contributes to the literature on justification bias in two novel and important ways. First, it provides evidence from unique longitudinal data in which respondents answer identical disability questions at the beginning and end of the interview. Given the two self-reports are based on identical face-to-face survey questions measured on the same day, differences between them cannot be generated by differences in an individual's true multidimensional health status. Second, it presents estimates based on within-individual variation over time. This modelling approach controls for all time-invariant factors that are associated with non-employment and that influence response behavior, such as cognitive ability, English language skills, personality and culture. Results show that estimates based on within-individual variation are significantly different from those based on between-individual variation.

We find that non-employed respondents and disability pension recipients are significantly more likely to misreport or exaggerate their level of disability. For example, we find that conditional on the responses to the disability question at the beginning of the interview, unemployed and OLF males are 2.7 and 5.4 percentage points respectively more likely to report a disability at the end of the interview than are employed males. The effects of non-employment on misreporting of disability are generally smaller for females. We also find that individuals receiving a disability support pension, including those who are employed, are particularly likely to exaggerate their disability. For example, employed males receiving a disability pension are 9 percentage points more likely to report a disability than are employed males not receiving a disability pension. Additional analysis shows that the most misreported health condition is mental illness.

We argue that the variation in the two disability measures in our study is caused by survey priming effects. Between the two disability measures are multiple questions on employment history, job search, reasons for not working, reasons for not searching for work, and pension receipt. Therefore, it is likely that respondents are considering these issues when answering the second disability question. Self-reported disability will continue to be a practical and informative measure of disability in large surveys; however, we demonstrate that the ordering of questions can have a considerable impact on the reporting of health limitations. Our results therefore have potentially important implications for the many major surveys that include health and disability questions after questions on income, employment, and retirement.

References

Anderson, K. H. and R. V. Burkhauser (1985). "The Retirement-Health Nexus: A New Measure of an Old Puzzle." Journal of Human Resources 20(3).

Autor, DH. and MG Duggan (2006). “The growth in the social security disability rolls: a fiscal crisis unfolding.” Journal of Economic Perspectives 20(3): 71-96.

Baker, M., M. Stabile and C. Deri (2004). "What do self-reported, objective, measures of health measure?" Journal of Human Resources 39(4): 1067-1093.

Benítez-Silva, H., M. Buchinsky, H. Man Chan, S. Cheidvasser and J. Rust (2004). "How large is the bias in self-reported disability?" Journal of Applied Econometrics 19(6): 649-670.

Benjamin, D.J., J.J. Choi and A.J. Strickland (2010). “Social Identity and Preferences.” American Economic Review 100(4): 1913–28.

Bound, J. (1991). "Self-Reported versus Objective Measures of Health in Retirement Models." Journal of Human Resources 26(1): 106-138.

Broadway, B., A. Chigavazira and S. Kassenboehmer (2014). “Labour Force Potential of Disability Support Pension Recipients.” Melbourne Institute of Applied Economic and Social Research Report.

Callen, M, M. Isaqzadeh, J.D. Long and C. Sprenger (2014). “Violence and Risk Preference: Experimental Evidence from Afghanistan.” American Economic Review 104(1): 123–48.

Clarke, P.M. and C. Ryan (2006). “Self-reported health: reliability and consequences for health inequality measurement.” Health Economics 15: 645-652.

Cohn, A., J. Engelmann, E. Fehr, M.A. Maréchal (2015). “Evidence for Countercyclical Risk Aversion: An Experiment with Financial professionals”. American Economic Review, 105 (2): 860-885.

Cohn, A., E. Fehr and M.A. Maréchal (2014). “Business Culture and Dishonesty in the Banking Industry.” Nature 516(7529): 86–89.

Crossley, T.F. and S. Kennedy (2002). “The reliability of self-assessed health status.” Journal of Health Economics 21: 643-658.

Datta Gupta, N. and M. Larsen (2010). "The impact of health on individual retirement plans: self-reported versus diagnostic measures." Health Economics 19(7): 792-813.

Datta Gupta, N. and H. Jürges (2012). "Do workers underreport morbidity? The accuracy of self-reports of chronic conditions." Social Science and Medicine 75: 1589-1594.

Deaton, A. (2012). “The financial crisis and the well-being of Americans.” Oxford Economic Papers 64(1): 1-26.

Dwyer, D. S. and O. S. Mitchell (1999). "Health problems as determinants of retirement: Are self-rated measures endogenous?" Journal of Health Economics 18(2): 173-193.

Gannon, B. (2009). "The influence of economic incentives on reported disability status." Health Economics 18(7): 743-759.

Gosling, A. and E. Saloniki (2014). "Correction of misclassification error in disability rates." Health Economics 23(9): 1084–1097.

Kapteyn, A., JP 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.

Kerkhofs, M. and M. Lindeboom (1995). "Subjective health measures and state dependent reporting errors." Health Economics 4(3): 221-235.

Kreider, B. (1999). "Latent work disability and reporting bias." Journal of Human Resources: 734-769.

Liebman, J.B. (2015). “Understanding the Increase in Disability Insurance Benefit Receipt in the United States.” Journal of Economic Perspectives 29(2): 123-150.

Lindeboom, M. and M. Kerkhofs (2009). "Health and work of the elderly: subjective health measures, reporting errors and endogeneity in the relationship between health and work." Journal of Applied Econometrics 24(6): 1024-1046.

Lumsdaine, R.L. and A. Exterkate (2013). “How survey design affects self-assessed health responses in the Survey of Health, Ageing, and Retirement in Europe (SHARE).” European Economic Review 63: 299-307.

Macklin, J (2009) “Valuing the future: policymaking for the long term.” Speech for the Per Capita Policy Exchange Conference, Canberra, 21 October 2009.

McGarry, K. (2004). “Health and Retirement: Do Changes in Health Affect Retirement Expectations?” Journal of Human Resources 39: 624-648.

Myers, R. J. (1982). "Why Do People Retire from Work Early." Social Security Bulletin 45(9): 10-14.

Parsons, D. O. (1982). "The Male Labour Force Participation Decision: Health, Reported Health, and Economic Incentives." Economica 49(193): 81-91.

Saunders, P. (2007). “Mutual Obligation, Unemployment and Well-being”. Australian Journal of Labour Economics 10: 167-184.

Shuval, J. T., A. Antonovsky and A. M. Davies (1973). "Illness: A mechanism for coping with failure." Social Science & Medicine (1967) 7(4): 259-265.

Stern, S. (1989). "Measuring the Effect of Disability on Labor Force Participation." Journal of Human Resources 24(3): 361-395.

Summerfield M, Freidin S, Hahn M, Ittak P, Li N, Macalalad N, et al. (2012). HILDA User Manual –Release 11. Melbourne: Melbourne Institute of Applied Economic and Social Research, University of Melbourne

Figures & Tables

Figure 1: Estimated proportions of OLF respondents self-reporting a disability across ages

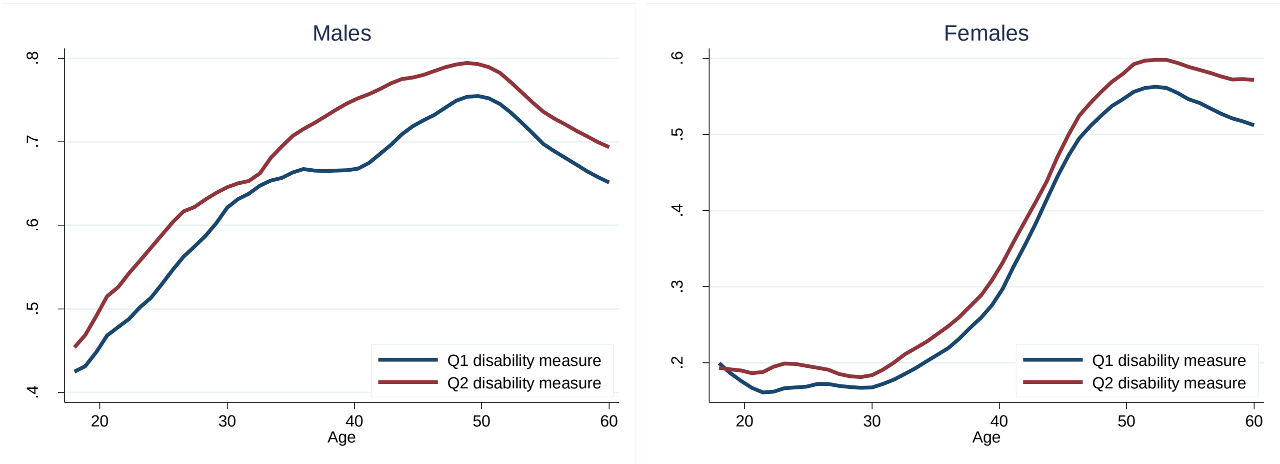

Figure 2: Variation between the self-reported Q1 and Q2 disability measures

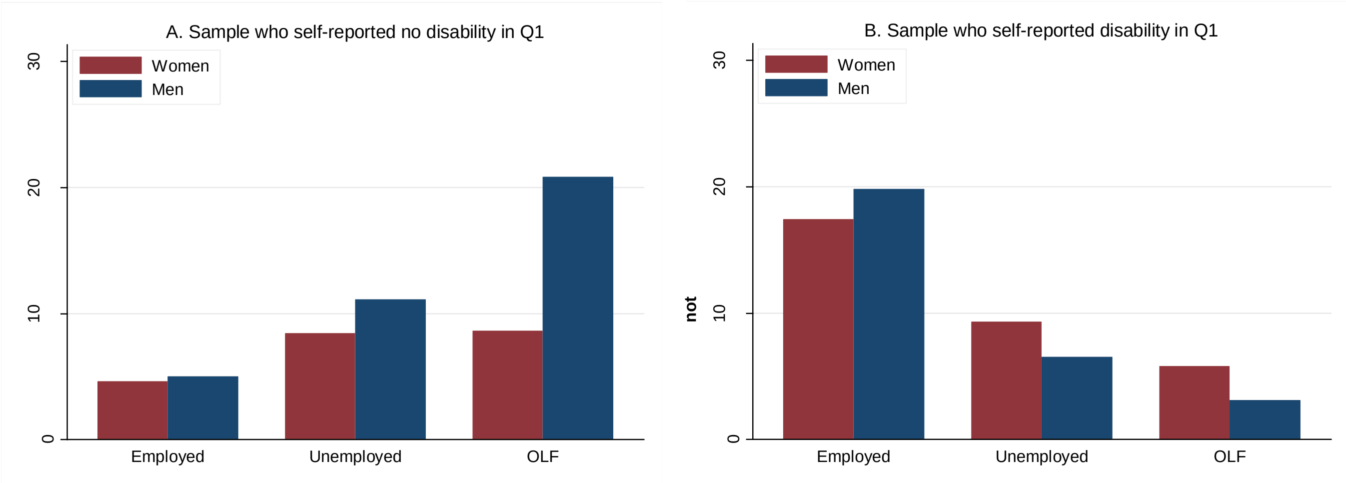

Table 1: Variation in reported disability rates across questionnaires, genders and employment states

	Male				Female			
	All	E	U	OLF	All	E	U	OLF
A) Sample who self-reported Q1								
Self-report in Q1	0.200	0.141	0.408	0.684	0.195	0.137	0.285	0.368
Self-report in Q2	0.219	0.156	0.447	0.729	0.216	0.153	0.318	0.401
Individuals	4,545	4,178	889	363	6,078	5,135	1,910	2,327
Observations	22,971	19,605	1,661	1,705	35,690	25,249	4,044	6,397
B) Sample whose partner reported Q1								
Partner-report in Q1	0.157	0.121	0.357	0.723	0.153	0.117	0.21	0.265
Self-report in Q2	0.197	0.16	0.427	0.777	0.2	0.159	0.28	0.319
Individuals	3,889	3,705	457	348	3,005	2,516	605	1,028
Observations	18,808	17,241	736	831	12,041	8,743	956	2,342

Note: Figures are sample means calculated using our estimation sample. The employment status categories are: employed (E), unemployed (U) and out of the labor force (OLF).

Table 2: Regression models of disability status recorded in Questionnaire 2

	Males		Females	
	(1)	(2)	(3)	(4)
Unemployed	0.084***	0.027***	0.057***	0.020***
	(0.009)	(0.010)	(0.006)	(0.006)
OLF	0.140***	0.054***	0.062***	0.021***
	(0.010)	(0.013)	(0.005)	(0.006)
Self-reported Q1 disability	0.751***	0.542***	0.790***	0.591***
	(0.008)	(0.011)	(0.005)	(0.009)
Fixed-effects	×	✓	×	✓
Observations	22971	22971	35690	35690

Notes: Standard errors (clustered at individual level) are presented in parentheses. Samples consist of individuals for whom disability in Q1 is self-reported. All regressions control for quadratic function in age, marital status, having dependent children, interview conditions, and year effects. Models (1) and (3) control for education and country of birth. Models (2) and (4) include an individual-specific fixed-effect. *** denote statistical significance at the 0.01 level.

Table 3: Fixed-effect regression models of disability status recorded in Questionnaire 2 using subsamples of single and partnered respondents

	Males			Females		
	(1)	(2)	(3)	(4)	(5)	(6)
Unemployed	0.025**	0.028*	0.028*	0.024**	0.017**	0.013
	(0.012)	(0.015)	(0.015)	(0.011)	(0.007)	(0.014)
OLF	0.049***	0.063***	0.069***	0.038***	0.014**	0.016
	(0.018)	(0.020)	(0.018)	(0.013)	(0.007)	(0.011)
Self-reported Q1 disability	0.542***	0.526***	-	0.593***	0.578***	-
	(0.017)	(0.016)		(0.014)	(0.011)	
Partner-reported Q1 disability	-	-	0.421***	-	-	0.441***
			(0.013)			(0.017)
Sample	Singles	Partnered	Partnered	Singles	Partnered	Partnered
Q1 disability is reported by	Self	Self	Partner	Self	Self	Partner
Observations	10051	12920	18808	12483	23207	12041

Notes: Standard errors (clustered at individual level) are presented in parentheses. All regressions control for quadratic function in age, marital status, having dependent children, interview conditions, and year effects. Models (1), (2), (4) and (5) include an individual-specific fixed-effect, whereas Models (3) and (6) include an individual-couple-specific fixed-effect. *, ** and *** denote statistical significance at the 0.10, 0.05 and 0.01 level, respectively.

Table 4: Sensitivity to controlling for additional self-reported health information

	(1)	(2)	(3)	(4)
A. Males				
Unemployed	0.024**	0.020*	0.044***	0.014
	(0.011)	(0.010)	(0.013)	(0.014)
OLF	0.049***	0.043***	0.092***	0.036*
	(0.014)	(0.014)	(0.018)	(0.021)
Self-reported Q1 disability	0.542***	0.529***	-	0.503***
	(0.012)	(0.012)		(0.019)
Observations	19507	19507	19507	14321
B. Females				
Unemployed	0.017***	0.013**	0.034***	0.006
	(0.006)	(0.006)	(0.008)	(0.011)
OLF	0.017***	0.010	0.036***	0.010
	(0.006)	(0.006)	(0.008)	(0.010)
Self-reported Q1 disability	0.590***	0.571***	-	0.546***
	(0.009)	(0.009)		(0.014)
Observations	31717	31717	31717	24076
Q1 disability	✓	✓	×	✓
SF-36 physical health	×	✓	✓	×
SF-36 mental health	×	✓	✓	×
Lagged Q2 disability	×	×	×	✓

Notes: Presented figures are coefficient estimates from fixed-effects models. Standard errors (clustered at individual level in models 1 to 3) are presented in parentheses. The sample consists of individuals for whom disability in Q1 is self-reported. SF-36 physical and mental health covariates are cubic functions of the SF-36 physical functioning and mental health subscales. All regressions additionally control for quadratic function in age, marital status, having dependent children, interview conditions, year effects and individual-level fixed-effects. *, ** and *** denote statistical significance at the 0.10, 0.05 and 0.01 level, respectively.

Table 5: Fixed-effect regression models of work-limiting disability status reported in Questionnaire 2

	Males		Females	
	Non Work Limiting	Work Limiting	Non Work Limiting	Work Limiting
Unemployed	-0.004	0.041***	-0.003	0.031***
	(0.009)	(0.009)	(0.005)	(0.005)
OLF	-0.019	0.081***	-0.005	0.035***
	(0.016)	(0.013)	(0.005)	(0.005)
Self-reported Q1 disability	0.482***	0.455***	0.531***	0.504***
	(0.015)	(0.014)	(0.011)	(0.012)
Work limiting disability	×	✓	×	✓
Sample mean	0.071	0.148	0.062	0.154
Observations	19566	21338	30198	33472

Notes: Presented figures are coefficient estimates from fixed-effects models. Standard errors (clustered at individual level) are presented in parentheses. The sample consists of individuals for whom disability in Q1 is self-reported. All regressions additionally control for quadratic function in age, marital status, having dependent children, interview conditions, year effects and individual-level fixed-effects. *** denote statistical significance at the 0.01 level.

Table 6: Regression models of specific health conditions reported in Questionnaire 2

	Mental Health Condition	Limited Use of Limbs	Chronic or Recurring Pain	Hearing & Sight Problems	Other Specified Conditions	Unspecified Restrictive Condition	Any Other Unspecified Condition
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Males							
Unemployed	0.020**	0.019**	0.018**	0.000	0.007	0.033***	0.007
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.010)
Out of LF	0.070***	0.044***	0.040***	0.007	0.041***	0.065***	0.014
	(0.015)	(0.014)	(0.015)	(0.015)	(0.015)	(0.016)	(0.016)
Sample mean	0.041	0.051	0.054	0.046	0.042	0.116	0.069
Observations	15815	16022	16072	15918	15836	17260	16360
B. Females							
Unemployed	0.025***	0.017***	0.008*	0.002	0.014***	0.019***	0.013**
	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)	(0.006)	(0.006)
Out of LF	0.029***	0.031***	0.014***	0.004	0.012***	0.027***	0.013**
	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)	(0.006)	(0.005)
Sample mean	0.051	0.050	0.057	0.027	0.042	0.116	0.083
Observations	24787	24753	24965	24047	24509	26724	25733

Notes: Presented figures are coefficient estimates from fixed-effects models. Standard errors (clustered at individual level) are presented in parentheses. The sample consists of individuals for whom disability in Q1 is self-reported. All regressions additionally control for Q1 disability status, quadratic function in age, marital status, having dependent children, interview conditions, year effects and individual-level fixed-effects. *, ** and *** denote statistical significance at the 0.10, 0.05 and 0.01 level, respectively.

Table 7: Effects of disability benefit receipt on disability status reported in Questionnaire 2

	Males		Females	
	Any	Work Limiting	Any	Work Limiting
	(1)	(2)	(3)	(4)
No DSP - Unemployed	0.020**	0.034***	0.018***	0.027***
	(0.010)	(0.010)	(0.006)	(0.005)
No DSP - OLF	0.052***	0.079***	0.016***	0.030***
	(0.015)	(0.014)	(0.006)	(0.006)
Receive DSP - Employed	0.091***	0.143***	0.048	0.086***
	(0.023)	(0.026)	(0.030)	(0.030)
Receive DSP - Unemployed	0.133***	0.179***	0.103***	0.151***
	(0.026)	(0.027)	(0.019)	(0.020)
Receive DSP - OLF	0.127***	0.181***	0.097***	0.140***
	(0.023)	(0.024)	(0.016)	(0.016)
Observations	22971	21338	35690	33472
F-statistic	1.877	1.486	1.875	2.585
p-value	0.153	0.226	0.153	0.075

Notes: Presented figures are coefficient estimates from fixed-effects models. Standard errors (clustered at individual level) are presented in parentheses. Omitted category is ‘No DSP - Employed’. Bottom panel presents F-statistics and p-values for the F-test of the equality of ‘Receive DSP - Employed, ‘Receive DSP - Unemployed’, and ‘Receive DSP - OLF’. The sample consists of individuals for whom disability in Q1 is self-reported. All regressions additionally control for Q1 disability status, quadratic function in age, marital status, having dependent children, interview conditions, year effects and individual-level fixed-effects. ** and *** denote statistical significance at the 0.05 and 0.01 level, respectively.

APPENDIX TABLES

Table A1: Regression models of Q2 disability status for subsamples defined by Q1 disability status

	Males				Females			
	$D_{it}^{Q1}=0$	$D_{it}^{Q1}=1$	$D_{it}^{Q1}=0$	$D_{it}^{Q1}=1$	$D_{it}^{Q1}=0$	$D_{it}^{Q1}=1$	$D_{it}^{Q1}=0$	$D_{it}^{Q1}=1$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Unemployed	0.049***	0.080***	0.017	0.014	0.034***	0.055***	0.015**	0.012
	(0.010)	(0.013)	(0.010)	(0.018)	(0.006)	(0.011)	(0.006)	(0.015)
OLF	0.093***	0.100***	0.031*	0.011	0.031***	0.079***	0.007	0.006
	(0.020)	(0.010)	(0.019)	(0.020)	(0.005)	(0.009)	(0.006)	(0.014)
Fixed-effects	×	×	✓	✓	×	×	✓	✓
Observations	18037	3868	18039	3868	28327	6021	28333	6021

Notes: Standard errors (clustered at individual level) are presented in parentheses. Samples consist of individuals for whom disability in Q1 is self-reported. All regressions control for quadratic function in age, marital status, having dependent children, interview conditions, and year effects. Models (1), (2), (5), and (6) control for education and country of birth. Models (3), (4) (7), and (8) include an individual-specific fixed-effect. *, ** and *** denote statistical significance at the 0.10, 0.05 and 0.01 level, respectively.

Table A2: Multinomial logit models of Q1 and Q2 disability answers

	Males				Females			
	No-No	Yes-Yes	No-Yes	Yes-No	No-No	Yes-Yes	No-Yes	Yes-No
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Unemployed	-0.193***	0.172***	0.020***	0.001	-0.172***	0.150***	0.021***	0.001
	(0.008)	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)	(0.003)	(0.002)
OLF	-0.310***	0.278***	0.029***	0.002	-0.192***	0.182***	0.013***	-0.003
	(0.009)	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)	(0.003)	(0.002)
Observations	22969	22969	22969	22969	35684	35684	35684	35684

Notes: No-No = No disability in Q1 or Q2. Yes-Yes = Disability in Q1 and Q2. No-Yes = No disability in Q1 but disability in Q2. Yes-No = Disability in Q1 but no disability in Q2. Presented figures are average partial effects. Standard errors (clustered at individual level) are presented in parentheses. Samples consist of individuals for whom disability in Q1 is self-reported. All regressions control for quadratic function in age, marital status, having dependent children, education, country of birth, and interview conditions, and year effects. *** denote statistical significance at the 0.01 level, respectively.

Table A3: Effects of non-employment on reporting disability in Q2 for experienced respondents

	Males		Females	
	Base	Experienced respondents	Base	Experienced respondents
	(1)	(2)	(3)	(4)
Unemployed	0.027***	0.024**	0.020***	0.025***
	(0.010)	(0.012)	(0.006)	(0.008)
OLF	0.054***	0.042**	0.021***	0.014*
	(0.013)	(0.017)	(0.006)	(0.007)
Self-reported Q1 disability	0.542***	0.515***	0.591***	0.574***
	(0.011)	(0.014)	(0.009)	(0.011)
Observations	22971	16300	35690	24838

Notes: Presented figures are coefficient estimates from fixed-effects models. Standard errors (clustered at individual level) are presented in parentheses. The samples consist of individuals for whom disability in Q1 is self-reported. In columns (2) and (4), observations in the first 3 years that respondents are surveyed are excluded. All regressions control for quadratic function in age, marital status, having dependent children, interview conditions and year effects, and individual-level fixed-effects. *, ** and *** denote statistical significance at the 0.10, 0.05 and 0.01 level, respectively

Table A4: Heterogeneity by specific health condition, estimates of coefficient on OLF.

	Males			Females		
	Mean	Coef.	S.E.	Mean	Coef.	S.E.
Sight problems	0.019	0.003	(0.014)	0.012	-0.003	(0.003)
Hearing problems	0.030	0.013	(0.013)	0.016	0.000	(0.003)
Speech problems	0.004	-0.004	(0.005)	0.002	0.002	(0.001)
Difficulty breathing	0.018	0.022*	(0.012)	0.023	0.008*	(0.004)
Disfiguration/deformity	0.007	-0.007	(0.009)	0.004	0.002	(0.002)
Slow at learning	0.010	0.001	(0.010)	0.008	0.009**	(0.004)
Head injury/stroke	0.009	0.016	(0.010)	0.006	0.003	(0.003)
Limited use of arms/fingers	0.022	0.029**	(0.013)	0.024	0.014***	(0.004)
Limited use of feet/legs	0.036	0.047***	(0.014)	0.030	0.023***	(0.005)
Difficulty gripping things	0.018	0.023*	(0.012)	0.023	0.011***	(0.004)
Blackouts, fits	0.006	0.016**	(0.007)	0.008	0.005*	(0.003)
Nervous/emotional condition	0.033	0.058***	(0.016)	0.043	0.027***	(0.006)
Mental illness	0.016	0.034**	(0.015)	0.016	0.011***	(0.004)
Chronic pain	0.054	0.028*	(0.015)	0.057	0.008	(0.005)
Restriction of physical activity	0.085	0.043**	(0.018)	0.081	0.017***	(0.006)
Other treated condition	0.056	0.069***	(0.019)	0.061	0.034***	(0.006)
Other LT condition	0.069	-0.000	(0.017)	0.083	0.005	(0.007)
Observations	19305			29947		

Notes: Presented figures are coefficient estimates from fixed-effects models. Standard errors (clustered at individual level) are presented in parentheses. The sample consists of individuals for whom Q1 disability is self-reported. All regressions additionally control for Q1 disability status, quadratic function in age, marital status, having dependent children, interview conditions, year effects and individual-level fixed-effects. *, ** and *** denote statistical significance at the 0.10, 0.05 and 0.01 level, respectively.