Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors*

Brent Kreider
Iowa State University
bkreider@iastate.edu

John V. Pepper
University of Virginia
jvpepper@virginia.edu

January 2003

Abstract

Long-standing debates about relationships between labor supply behavior and health status among persons nearing retirement age have centered largely on disagreements about the reliability of self-reported health indicators. In light of reporting errors in work capacity, this paper considers the problem of predicting how employment rates vary with disability status when "true" disability is unobserved. Rather than imposing the strong assumptions required to obtain point identification, we take a step back to evaluate what can be inferred under a variety of assumptions that are weaker but arguably more credible than those imposed in the existing literature. Although these assumptions do not identify the conditional employment rates except in special cases, nonparametric bounds for these parameters can be obtained. Using data from the Health and Retirement Study and the Survey of Income and Program Participation, we estimate a set of bounds that formalize the identifying power of a number of different assumptions that appear to have broad consensus in the literature. Our results suggest that models estimated under the assumption of fully accurate reporting lead to biased inferences. In particular, it appears that nonworkers in both datasets tend to overreport disabilities.

1 Introduction

This paper considers the problem of predicting how employment rates of older persons vary with their disability status when "true" disability is unobserved. Health status is widely recognized as an important predictor of labor supply behavior, especially among persons nearing retirement age. During the past twenty years, however, researchers have come to recognize that inferences about the effects of health and government policies on labor market outcomes can be highly sensitive to the type of health measure used in the analyses. Anderson and Burkhauser (1984), for example, found

*We gratefully acknowledge financial support from the 2001-2002 Joint Center for Poverty Research / Census Bureau Research Development Grant, the 2001-2002 Sandell Grant Program, and from the W.E. Upjohn Institute for Employment Research.

that their estimated wage elasticity of work participation varied five-fold depending on whether work capacity was measured using a self-reported measure of disability or an indicator of subsequent mortality. Ongoing debates regarding the influence of Social Security Disability Insurance (SSDI) policy on declining labor force participation rates have also emphasized issues regarding the reliability of self-reported disability information [e.g., Haveman and Wolfe (1984) vs. Parsons (1984); Bound (1991b) vs. Parsons (1991)].

Many researchers have controlled for work capacity based on responses to questions of the general type: "Do you have a health condition that limits the kind or amount of work you can perform?" Some researchers favor the use of self-reported measures in labor supply studies because they provide direct information about work ability. In contrast, specific health conditions or functional limitations do not necessarily imply work disability. Other researchers, however, emphasize that self-reports of work capacity are especially prone to reporting bias. Little has been resolved about the appropriate measurement of work capacity since Anderson and Burkhauser (1984) characterized this problem as "the major unsettled issue in the empirical literature on the labor supply of older workers."

The debate has grown stronger over time, with some maintaining that self-reported measures of disability status can be treated as perfectly reliable (e.g., Benitez-Silva et al., 1997; 1999) and others arguing that self reported measures are completely unreliable (e.g., Meyers, 1982; Bowe, 1993). Still others suggest that self-reports can be considered reliable for certain subpopulations but not others (e.g., McGarry, 2002). Recent Institute of Medicine workshop reports (Wunderlich, 1999; Mathiowetz and Wunderlich, 2000) highlight the lack of information on reporting errors and call for more research on the degree and nature of these errors.

In the absence of this information, studies that have modeled and assessed the reliability of self-reported work limitations have not been able to resolve these issues. Stern (1989), Dwyer and Mitchell (1999), and Benitez-Silva et al. (1999), for example, find little evidence that labor market outcomes affect reporting behavior. In contrast, Kerkhofs and Lindeboom (1995, 2002), O'Donnell

(1998), and Kreider (1999, 2000) estimate large reporting errors that are systematically related to labor force status.[1] Under a set of weak nonparametric assumptions, Kreider and Pepper (2001) also find evidence that self-reports are invalid. Bound (1991a) highlights the econometric issues surrounding reporting errors in the framework of a parametric simultaneous equations model.

Given the uncertainty about reporting behavior, we do not focus on providing point estimates of the true conditional employment rates. Instead, following our recent work in Kreider and Pepper (2001) which provides bounds on the true disability rate, we take a step back to evaluate what can be inferred under a variety of nonparametric assumptions that are weaker but arguably more credible than those imposed in the existing literature. We examine conditional employment rates among respondents nearing retirement age in the Health and Retirement Study (HRS) who report on the existence of impairments that limit the kind or amount of work that can be performed. We then estimate these same probabilities for the population more generally using data from the Survey of Income and Program Participation (SIPP). After describing these data in the next section, Section 3 formalizes the identification problem created by misreporting. Without assumptions about the prevalence of inaccurate reports relative to social norms or other criteria specific to the research question,[2] the sampling scheme is uninformative: the true conditional employment rates could lie anywhere between 0 and 100 percent.

In Section 4, we evaluate what can be inferred when it is assumed that members of certain observed subgroups (e.g., workers) – or at least some lower bound fraction of members of these groups – provide valid self-reports. In Section 5, we apply and extend results from Manski and Pepper (2000) and Kreider and Pepper (2001) to consider the additional identifying power of arguably innocuous monotonicity assumptions that link disability to certain observed covariates. In particular, we assume that the disabled are less likely to work, that the disability rate does not

[1] Earlier studies that found evidence of systematic reporting errors include Gordon and Blinder (1980), Myers (1982), Chirikos and Nestel (1984), Bazzoli (1985), and Anderson and Burkhauser (1985).

[2] The SSA requires that beneficiaries are unable "to engage in any substantial gainful activity by reason of any medically determinable physical or mental impairment which can be expected to result in death or can be expected to last for a continuous period of not less than one year." Substantial gainful activity is currently defined as employment resulting in earnings in excess of $780/month.

decrease in age, and/or that employment rates conditional on disability do not increase with age. Under these nonparametric models, we find evidence that the assumption of valid self-reporting is not supported by our data. We draw conclusions in Section 6.

2 Data

Our analysis uses data from two sources. We use the Health and Retirement Study (HRS) to estimate conditional employment probabilities among those nearing retirement age and the Survey of Income and Program Participation (SIPP) to estimate conditional employment probabilities more generally across all adults younger than 70.

The HRS is a nationally representative panel survey with 7608 households whose heads were nearing retirement age (aged 51-61) in 1992-93. The HRS has become an especially popular data source for studying the effects of health status and public policy on work behavior of older persons because of its detailed information about health and disability, work history, and participation in public transfer programs. We use self-reported health and labor force participation information from all 12,652 respondents. We also record each respondent's age, years of schooling, occupation, race, gender, and whether the respondent received government assistance for a disability.[3] As part of our identification strategy, some of our analysis also incorporates reported health and employment information from the second wave, which was conducted two years after the first wave.

The Survey of Income and Program Participation (SIPP) is a nationally representative longitudinal survey with new panels introduced at regular intervals. Because of its detailed information about health and disability, work history, and participation in government transfer programs for all age groups, the SIPP is perhaps the most important data source for studying the effects of health status and public policy on work behavior and disability. By covering the U.S. civilian non-institutionalized population, these data can be used to draw inferences about the entire population rather than just those nearing retirement. We utilize data from the first wave of the 1996 panel,

[3] We imputed age for one member of the sample.

a nationally representative sample of 36,800 households. These households were interviewed every four months for four years. Because respondents older than 69 were not asked about work limitations, we restrict the SIPP sample to the 60,633 individuals between the ages of 18 and 69. We use information from the first interview. As above for the HRS data, we also incorporate reported health and employment information from the second wave.

Table 1 displays descriptive statistics. In the HRS, 21.5 percent of the sample responded in the affirmative to the question, "Do you have any impairment or health problem that limits the kind or amount of paid work you can do?" Respondents were also asked about current job status: "Are you working now, temporarily laid off, unemployed and looking for work, disabled and unable to work, retired, a homemaker, or what?" About 9 percent indicated that they were disabled and unable to work. We classify respondents as work-limited by their own self-assessments (denoted $X = 1$) if they answered yes to either of the disability questions. Using this standard, 21.8% of the respondents in the HRS sample claimed to be disabled. The corresponding fraction of respondents in the SIPP data claiming to be disabled is 13.1%.[4] In addition to self-reported disability measures, we also observe whether the respondent was employed and/or receiving disability benefits. In total, 66.2 percent of respondents in the HRS and 75.0% of the respondents in the SIPP identified themselves as currently working for pay. In the HRS sample of generally older respondents, 10.0 percent reported that they were receiving disability benefits (or were scheduled to begin receiving benefits) from Social Security Disability Insurance, Supplemental Security Income, Veterans' Disability, Workers' Compensation, or a state disability program.[5] In the nationally representative SIPP, 3.0% of adults reported receipt of disability benefits from a public or private source.

Table 2 presents labor force participation rates conditioning on self-assessed work limitation, gender, and age. In the HRS, the employment rate among those reporting to be disabled is 0.295

[4] Based on the wave five interviews from the 1996 SIPP, McNeil (2000) reports that about 16 million individuals claimed to have a work disability on the basis of being limited in the kind or amount of work that could be performed.

[5] There were also some additional cases in which the head of household reported disability income, but we could not determine which family member was receiving the benefit. Similarly, some heads of households reported disability income from a private insurance plan, but it was generally not possible for us to determine which family member was receiving the benefit. In these cases, we did not classify the respondent as receiving disability assistance.

compared with 0.765 for those reporting to be nondisabled. The difference in reported employment rates by reported disability status is thus -0.470. In the SIPP data, the employment rates among those claiming to be disabled and nondisabled are 0.358 and 0.808, respectively, resulting in a difference of -0.450.

3 Statement of the Identification Problem

Predicting how the employment rate varies by disability status requires assumptions about reporting errors. The data do not reveal the fraction of respondents who give invalid responses to disability status questions. It might be that all positive reports of disability are inaccurate, in which case everyone is nondisabled. Alternatively, it might be that all negative reports are inaccurate, in which case the entire population may be disabled. In this section, we formalize the basic identification problem that arises in corrupt data and then consider the sensitivity of inferences to prior information on the degree of possible inaccurate reporting.

To evaluate the implications of invalid response, we introduce notation that distinguishes between self-reports and accurate reports. Let L be an employment indicator, with $L = 1$ if the respondent is employed and 0 otherwise. Let X be a self-reported disability measure, where $X = 1$ if the respondent reports being limited in the ability to work and 0 otherwise. Let $W = 1$ indicate that the individual is truly limited in the ability to work relative to social norms (or other specified criterion), with $W = 0$ otherwise. Finally, let Z indicate whether a respondent provides accurate information, with $Z = 1$ if $W = X$ and $Z = 0$ otherwise. The data reveal the employment rate given the reported disability status, $P(L = 1|X)$.[6] We are interested, however, in learning (a) the employment probability given the true disability status,

$$P(L = 1|W), \tag{1}$$

[6] Observed characteristics of individual respondents (e.g., age, race, gender, education, etc.) are left implicit to simplify the exposition.

and (b) how the employment rate varies by disability status:

$$\beta = P(L = 1|W = 1) - P(L = 1|W = 0). \tag{2}$$

While the data reveal how the employment rate varies by reported disability status, the conditional employment probabilities (and therefore β) are not identified by the sampling process. To see this identification problem, consider evaluating the employment probability for the disabled, $P(L = 1|W = 1)$. Using Bayes Theorem, we decompose this conditional probability into the respective joint and marginal distributions:

$$\begin{aligned} P(L = 1|W = 1) &= \frac{P(L = 1, W = 1)}{P(W = 1)} \qquad (3) \\ &= \frac{P(L = 1, X = 1) + P(L = 1, X = 0, Z = 0) - P(L = 1, X = 1, Z = 0)}{P(X = 1) + P(X = 0, Z = 0) - P(X = 1, Z = 0)}. \end{aligned}$$

The data identify the fraction who self-report disability, $P(X = 1)$, and the joint probability of being employed and claiming to be disabled, $P(L = 1, X = 1)$, but they do not reveal the distribution of accurate reporters. Some unknown fraction of respondents, $P(X = 1, Z = 0)$, inaccurately report being disabled (i.e., false positives) while others, $P(X = 0, Z = 0)$, inaccurately report being nondisabled (i.e., false negatives). In the absence of restrictions on misreporting, the data are uninformative; we only know that the conditional employment rate lies between 0 and 1.

Assumptions about the nature and degree of reporting errors might be informative on the conditional employment rates. In the most limited informational setting, we assume only a lower bound on the fraction of respondents that accurately report disability status. We then investigate what can be learned by assuming that members of certain observed subgroups provide accurate reports. Finally, we consider the identifying power of monotonicity assumptions that link the employment rate to other covariates.

3.1 Upper Bound Error Probability

At one extreme, the sampling process is perfectly informative if it is known that all respondents provide accurate self-reports of disability status. At the other, if there is no prior information on the extent of misreporting, the data are uninformative. By varying the degree of possible misreporting, we can examine middle ground positions and explore the sensitivity of the identification result. In particular, suppose

$$P(Z = 1) \geq v \tag{4}$$

where v is an assumed lower bound on the accurate reporting rate.

The lower bound restriction in Equation (4) implies restrictions on the unknown joint distributions in Equation (3). With the degree of misreporting being no greater than some known fraction, $1 - v$, we can provide the following "degree bounds":

Proposition 1. Let $P(Z = 1) \geq v$. Then:

$$\frac{P(L = 1, X = 1) - \delta}{P(X = 1) - 2\delta + (1 - v)} \leq P(L = 1|W = 1) \leq \frac{P(L = 1, X = 1) + \gamma}{P(X = 1) + 2\gamma - (1 - v)} \tag{5}$$

where

$$\delta = \begin{cases} \min\{(1 - v), P(L = 1, X = 1)\} & \text{if } P(L = 1, X = 1) - P(L = 0, X = 1) - (1 - v) \leq 0 \\ \max\{0, (1 - v) - P(L = 0, X = 0)\} & \text{otherwise} \end{cases}$$

and

$$\gamma = \begin{cases} \min\{(1 - v), P(L = 1, X = 0)\} & \text{if } P(L = 1, X = 1) - P(L = 0, X = 1) + (1 - v) \leq 0 \\ \max\{0, (1 - v) - P(L = 0, X = 1)\} & \text{otherwise.} \end{cases}$$

The proof is provided in the Appendix. Note that when the lower bound fraction of accurate reporters is relatively small, the bounds on the conditional employment rates are uninformative. In particular, when $(1 - v) \geq P(L = 1, X = 1)$, the lower bound is zero as expected. Similarly, when $(1 - v) \geq P(L = 0, X = 1)$, the upper bound is one. Analogous bounds for $P(L = 1|W = 0)$ are obtained by replacing $X = 1$ with $X = 0$ and vice versa in the Proposition.

3.2 Results

By varying the value of v, we can effectively consider the wide range of views characterizing the debate on inaccurate reporting. Those willing to assume that all reports are accurate can set $v = 1$ (e.g., Benitez-Silva et al., 1997), in which case the sampling process identifies the conditional employment rates. Those believing that all reports are potentially inaccurate (e.g., Myers, 1982; Bowe, 1993) can set $v = 0$, in which case the sampling process is uninformative. Middle ground positions can be evaluated by setting v between 0 and 1. Abstracting from concerns over sampling variability, we consider what can be learned under these middle ground restrictions on the minimum accurate reporting rate.[7]

Given any conjectured value of the lower bound accurate reporting rate, v, Figures 1 and 2 display the Proposition 1 degree bounds on the conditional employment rate for the disabled and nondisabled, respectively, using the HRS sample. The obvious striking feature of these figures is that the bounds are uninformative across a large range of values for v. When $v = 0$, the employment rate can of course lie anywhere between 0 and 1. The data remain uninformative about the employment rate for the disabled, $P(L = 1|W = 1)$, unless it can be assumed that the accurate reporting rate exceeds $1 - P(L = 0, X = 1) = 0.84$; the lower bound is zero unless the accurate reporting rate exceeds $1 - P(L = 1, X = 1) = 0.94$. The bounds on the employment rate for the nondisabled begin to narrow when $v = 0.41$, while the upper bound remains at one until the accurate reporting rate exceeds 0.82. As the known accurate reporting rate increases beyond these thresholds, the Proposition 1 degree bounds displayed in Figures 1 and 2 reduce to the self-reported conditional employment rates of 0.295 and 0.765, respectively. In the absence of additional assumptions, the data cannot reject the possibility of fully accurate reporting.

We are also interested in learning how the employment probability varies with health status, β.

[7] In this section, we focus exclusively on the identification problem that arises because of misreporting. In the remaining sections, we also consider statistical variability via confidence intervals. With over 12,000 observations in the HRS sample and over 60,000 in the SIPP sample, however, the uncertainty created by the identification problem is much more extensive than uncertainty associated with sampling variability.

An upper (lower) bound on β can be found by subtracting the Proposition 1 lower (upper) bound on $P(L = 1|W = 0)$ from the Proposition 1 upper (lower) bound on $P(L = 1|W = 1)$. Figure 3 displays these naive bounds on β. Although these bounds on the difference between the two conditional probabilities are intuitive and simple to compute, they are not sharp: the constraint on the fraction of misreporting places further restrictions on β. In the appendix, Proposition 1A, we formalize sharp bounds on β, depicted as dotted lines in the figure.

The sharp bounds are identical to the naive bounds on β over much of the range of v in this application; where they slightly differ, the qualitative conclusions remain. In particular, under weak assumptions on the degree of accurate reporting, the employment rates given self-reported disability measures provide only modest information on the true employment rates of interest. The data are uninformative unless it is known that the accurate reporting rate exceeds 0.41 and the lower bound remains at -1 unless v exceeds 0.81. Without further assumptions, the sign of β is identified as negative (i.e., the data reveal that the disabled are less likely to work than the nondisabled) only if at least 88 percent of the respondents are known to provide accurate reports.

4 Verification of Observed Subgroups

So far, we have explored how inferences on the employment rate vary with prior information on the degree of misreporting. In practice, we find that these conservative restrictions on the extent of reporting errors are effectively uninformative except in the extreme cases in which virtually all respondents are known to provide accurate reports. We now consider several middle ground restrictions on both the degree and nature of misreporting. These models are motivated by theories of misreporting suggested in the existing literature that attempts to address inaccurate reporting using parametric latent variable models. Here, however, we avoid imposing the strong and generally unverifiable parametric assumptions that have been a primary source of controversy in the literature and instead focus on more primitive assumptions that appear to have some consensus.

4.1 Full Verification of Subgroups

Short of assuming that all respondents provide accurate self-reports of limitation, one might be willing to assume that certain types of respondents provide accurate reports, remaining agnostic about the reports from respondents in other subgroups. The existing literature provides a number of plausible restrictions (Bound and Burkhauser 1999). Concerns about misreporting focus primarily on financial and social incentives for certain types of respondents to exaggerate the effect of a health condition on lost work capacity. First, eligibility into some government assistance programs (e.g., SSDI) is contingent on being sufficiently work impaired. Given the difficulty in quantifying work capacity, even respondents intending to provide honest reports may systematically underreport work capacity given the associated financial advantages.[8] Second, it has long been asserted in the literature that many people, especially men, may feel social pressure to participate in the labor force until normal retirement age unless their ability to work is impaired (e.g., Bound, 1991a). Other factors constant, respondents who find themselves involuntarily out of work (or prefer not to work) may feel more compelled than employed workers to claim that a functional limitation (e.g., difficulty climbing stairs) interferes with the ability to work.

Thus, some respondents have clear incentives to misreport limitation while other observed subgroups seem to have few economic or psychological incentives to misreport limitation. Employed respondents, for example, are generally ineligible for government assistance and face neither strong social nor financial pressures to misreport.[9] Likewise, there appear to be few incentives to falsely claim to be nondisabled. Some might find it reasonable to assume that recent Disability Insurance beneficiaries, who faced stringent disability screening and are officially deemed incapable of substantial work, can be considered a verified work-limited subgroup (although not necessarily incapable of any work). Others might be willing to assume that respondents older than 65 offer valid reports

[8] In addition to monthly cash benefits, Supplemental Security Income (SSI) beneficiaries are immediately eligible for Medicaid benefits, and SSDI beneficiaries become eligible for Medicare benefits after a two-year waiting period.

[9] Kreider (1999) and McGarry (2002) explicitly assume that workers provide valid reports. Stern (1989) also explicitly assumes that misreporting is related to work outcomes. Each analysis, however, additionally imposes strong parametric assumptions and draws different conclusions.

since they are no longer expected to work and are no longer eligible for Disability Insurance.

Formally, let Y indicate whether a respondent's self-report of disability is verified to be accurate, where $Y=1$ if the report is verified and $Y=0$ otherwise. For simplicity, assume that the lower bound accurate reporting rate v equals the fraction of cases in verified groups, $P(Y=1)$.[10] We consider the possibility that four different observed subgroups are verified: (a) respondents who are disability beneficiaries (10 percent in the HRS) , (b) respondents who claimed no disability in the second wave of the survey despite being out of the labor force (27 percent in the HRS), (c) respondents who were gainfully employed (66 percent in the HRS), and (d) respondents who claimed no work limitation in the current wave (78 percent in the HRS). Ninety-four percent of the respondents in the HRS satisfied at least one of these criteria. Although we initially assume that all members of verified subgroups provide accurate responses, we relax this assumption below to allow for the possibility of some misreporting even within the verified groups.

As before, we can decompose the conditional distribution using Bayes Theorem:

$$\begin{aligned} P(L=1|W=1) &= \frac{P(L=1,W=1)}{P(W=1)} \\ &= \frac{P(L=1,W=1,Y=1)+P(L=1,W=1,\ Y=0)}{P(W=1,Y=1)+P(W=1,Y=0)}. \end{aligned} \tag{6}$$

The following proposition improves upon the bounds in Proposition 1:[11]

Proposition 2. If $Y=1 \Rightarrow Z=1$ (verification), then

$$\frac{P(L=1,X=1,Y=1)}{P(X=1,Y=1)+P(L=0,Y=0)} \leq P(L=1|W=1) \leq \frac{P(L=1,X=1,Y=1)+P(L=1,Y=0)}{P(X=1,Y=1)+P(L=1,Y=0)}. \tag{7}$$

In the special case where workers (and perhaps others) are verified, the joint probability of being a disabled worker, $P(L=1,W=1)$ equals the revealed probability, $P(L=1,X=1)$, and the only unknown parameter in Equation (6) is the disability rate among the unverified, $P(W=1,Y=0)$. In this case, the Proposition 2 bounds simplify to:

[10] This verification assumption can be generalized by allowing for the possibility that the lower bound probability on accurate reporting exceeds the verification probability.

[11] An alternative derivation of this result is given in the Horowitz and Manski (1998) bound for regressor censoring. Their result applies more generally to continuous outcomes.

Proposition 2, Corollary: Let workers be verified. Then

$$\frac{P(L=1,X=1)}{P(Y=1,X=1)+P(Y=0)} \leq P(L=1|W=1) \leq \frac{P(L=1,X=1)}{P(Y=1,X=1)}. \tag{8}$$

Notice that if only workers are verified so that $Y = L$, then the upper bound equals one; in the extreme case that nonworkers are not disabled, all of the disabled must be working. Again, the corresponding bounds for $P(L = 1|W = 0)$ are found by replacing $X = 1$ with $X = 0$ and vice versa.

4.2 Results

To evaluate how the bounds vary with the fraction verified, we begin by returning to Figures 1-3. The dotted lines represent the bounds for the true disability rate as a function of the verification probability $v = P(Y = 1)$. For these figures only, we assume that the joint distribution of employment and self-reported disability indicators, $\{L, X\}$, is independent of the verification indicator, Y. These figures clearly show the additional identifying power of the verification assumption. Consider, for instance, the bounds on the employment rate for the disabled. With only the Equation (4) lower bound assumption on the accurate reporting rate, the data are uninformative using the HRS sample unless $v > 0.84$, and even then are not informative on the lower bound unless $v > 0.94$. In contrast, verification in this setting is always informative for $v > 0$, with the greatest relative impact on the bounds for mid-range values of v (since both sets of bounds converge to the self-reported employment rates $P(L = 1|X)$ as v approaches 1). The verification bounds identify the sign of β for any $v > 0.73$, as opposed to $v > 0.88$ in the case where only restrictions on the degree of misreporting are imposed.

We now move from the abstract verification model depicted in Figures 1-3 to the particular results found under the four verification models described above. Estimated employment rate bounds for $P(L = 1|W = 1)$ and bootstrapped 90 percent confidence intervals (based on 1000 pseudosamples) are presented in Tables 3A and 3B using the HRS and SIPP data, respectively.[12] The second column of bounds displays the results for Proposition 2 which can be compared with the degree bounds of Proposition 1 in the preceding column. The widths of the bounds depend upon both the nature and size of the verified subpopulations. Beginning with the HRS data in Table 3A, the data remain uninformative if only those claiming to be nondisabled ($v = 0.782$) are verified. After all, under this assumption no verified cases report being disabled: $P(X = 1,\ Y = 1) = 0$. If workers alone are verified ($v = 0.662$), the data are informative on the lower bound but not on the upper bound. If the relatively small subgroup of SSDI beneficiaries are verified, the estimated bounds are informative on both sides. If all four groups in the table are verified, the Proposition 2 bounds narrow 32 points to $[0.295, 0.413]$, compared with $[0.014, 0.452]$ under Proposition 1. The results are similar for the SIPP data in Table 3B. When all four groups are verified, for example, the Proposition 2 bounds narrow by 34 points compared with Proposition 1.

Tables 4A and 4B provide corresponding results for β, the difference in the employment rate between the disabled and nondisabled. The bounds for β associated with Proposition 2 are informative (and sharp) in all cases. If those claiming to be nondisabled are verified, for example, then β is estimated to lie within $[-0.783, 0.361]$ using the HRS data and $[-0.818, 0.263]$ using the SIPP data. In this case, while the data reduce the uncertainly about how the employment rate varies with health status, the sign of the relationship remains uncertain. In contrast, the sign is identified as negative if workers and beneficiaries are verified. When all groups are verified, the employment rate difference between the disabled and nondisabled is estimated to lie within $[-0.470, -0.296]$ using the HRS data and within $[-0.450, -0.217]$ using the SIPP data.

Although there remains uncertainty about the true conditional employment rates, the verification bounds can be substantially more informative than the Proposition 1 degree bounds evaluated in the previous section. When all four subgroups are verified using the HRS, for example, the width of the bound on $P(L = 1|W = 1)$ reduces from 44 points to 12 points. Similarly, the width

[12] We report confidence intervals based on the bias-corrected percentile method (see Efron and Tibshirani, 1993).

of the bounds on β reduces from 56 points to 17 points. Still, there remains much uncertainly unless one has strong prior information on the degree and nature of misreporting. We cannot reject the possibility that all reports are accurate, nor can we generally reject the possibility that the employment rate is higher for the disabled.

4.3 Partial Verification

Thus far, we have assumed that all respondents in a verified subgroup provide valid reports. An important middle ground informational setting, especially useful for sensitivity analysis, arises if there is only partial information about an observed subgroup. There may be subgroups for which one is only willing to assume that at least some fraction of the members accurately report. Suppose, for example, that diagnostic tests used to evaluate health status and determine eligibility for assistance programs are only effective up to some known error rate (see, e.g., Parsons, 1996). Then, SSDI recipients would be partially verified. Likewise, those with few social or financial incentives to misreport (e.g., workers) may nevertheless find it difficult to accurately assess the degree to which they are disabled.

Formally, assume that at least some fraction v_y of the self-reports are known to be accurate such that $P(Z = 1|Y = 1) \geq v_y$.[13] In this case, the disability rate for the partially verified subgroup is not identified. Instead, applying the Proposition 1 bounds in Equation (5), we find:

Proposition 3. Let $P(Z = 1|Y = 1) \geq v_y$. Then:

$$\frac{P(L = 1, X = 1, Y = 1) - \delta}{P(X = 1, Y = 1) + P(L = 0, Y = 0) - 2\delta + (1 - v_y)P(Y = 1)}$$

$$\leq P(L = 1|W = 1) \leq$$

$$\frac{P(L = 1, X = 1, Y = 1) + P(L = 1, Y = 0) + \gamma}{P(X = 1, Y = 1) + P(L = 1, Y = 0) + 2\gamma - (1 - v_y)P(Y = 1)}$$

[13]There are many other hybrid versions of the partial verification model that one might consider. For instance, one might consider an alternative model where certain subgroups are fully verified, others are partially verified, and still others are not verified at all.

where

$$\delta = \begin{cases} \min\{(1-v_y)P(Y=1), P(L=1, X=1)\} & \text{if } \alpha \leq 0 \\ \max\{0, (1-v_y)P(Y=1) - P(L=0, X=0, Y=1)\} & \text{otherwise,} \end{cases}$$

$$\gamma = \begin{cases} \min\{(1-v_y)P(Y=1), P(L=1, X=0)\} & \text{if } \alpha' \leq 0 \\ \max\{0, (1-v_y)P(Y=1) - P(L=0, X=1, Y=1)\} & \text{otherwise,} \end{cases}$$

$$\alpha = P(L=1, X=1, Y=1) - P(L=0, X=1, Y=1) - P(L=0, Y=0) - (1-v_y)P(Y=1),$$

and

$$\alpha' = P(L=1, X=1, Y=1) - P(L=0, X=1, Y=1) + P(L=1, Y=0) + (1-v_y)P(Y=1).$$

Intuitively, the bounds widen if respondents in verified subgroups may misreport. For all $v_y < 1$, the Proposition 3 lower (upper) bound is always less (greater) than the analogous Proposition 2 bound. Still, for a sufficiently large lower bound accurate reporting rate v_y, partial verification improves upon the Proposition 1 bounds in Equation (5). Consider, for example, the Proposition 1 bound where $v = 0.10$, the fraction of disability beneficiaries. If we assume partial verification of beneficiaries, then with the HRS the upper bound (which is uninformative in Proposition 1) is improved if it can be assumed that even 27 percent of beneficiaries provide valid responses; improving upon the lower bound for this case requires that at least 96 percent provide valid responses.

5 Monotone Instrumental Variable Assumption

Restrictions on the relationship between disability and other observed characteristics can also be informative. In this section, we explore the identifying power of the monotone instrumental variable (MIV) assumption as discussed by Manski and Pepper (2000) and Kreider and Pepper (2001). This MIV assumption formalizes the notion that the employment rate may be known to vary monotonically with certain covariates. We first consider the restriction that the employment rate of the disabled is less than the employment rate of the nondisabled. We then consider an assumption that the conditional employment rate is nonincreasing with the age of respondents and an assumption that the disability rate is nondecreasing with age.

5.1 Employment and Disability

Disabilities, by definition, limit one's capacity to work, ceteris paribus. Here, we formally examine the implications of the arguably innocuous assumption that the disabled are less likely to work than the nondisabled. In particular, assume that

$$P(L = 1|W = 1) \leq P(L = 1|W = 0). \tag{9}$$

Thus, under Equation (9), the employment probability is higher for the nondisabled than the disabled.

With corrupt data, the conditional probabilities in Equation (9) are not identified. The restriction is informative, however, if these probabilities are bounds, say, from a verification assumption described in the previous section. To see this, note that Equation (9) implies $P(L = 1|W = 1)$ is no larger than the known upper bound on $P(L = 1|W = 0)$. Likewise, $P(L = 1|W = 0)$ is no smaller than the known lower bound on $P(L = 1|W = 1)$. Thus, we have:

Proposition 4. Let $P(L = 1|W = 1) \leq P(L = 1|W = 0)$ and let $LB(w)$ and $UB(w)$ be the known lower and upper bounds, respectively, given the available information on $P(L = 1|W = w)$. Then, it follows that

$$LB(1) \leq P(L = 1|W = 1) \leq \min\{UB(1), UB(0)\}, \tag{10}$$

$$\max\{LB(1), LB(0)\} \leq P(L = 1|W = 0) \leq UB(0), \text{ and}$$

$$LB(1) - UB(0) \leq \beta \leq \min\{0, \min[UB(1), UB(0)] - \max[LB(1), LB(0)]\}.$$

By assumption, the upper bound on the change in the employment rate by disability status, β, can be no greater than zero even though the upper bound on $P(L = 1|W = 1)$ may exceed the lower bound on $P(L = 1|W = 0)$.

The monotonicity bounds on the employment rates can be tightened further in the special case where workers are verified. To see this, recall the conditional probability of interest in Equation (3): $P(L=1|W=1)=\frac{P(L=1,W=1)}{P(W=1)}$. When workers are verified, the fraction of disabled workers is revealed so that only the denominator is not identified by the sampling process. While the corollary to Proposition 2 formalizes what is known about the employment rate when workers are verified, the restriction that the disabled are less likely to work provides additional information on the disability rate. In particular, Equation (9) implies that the unobserved probability that workers are disabled can be no greater than the probability that nonworkers are disabled: $P(W=1|L=1) \leq P(W=1|L=0)$. Thus, it follows from the law of iterated expectations that

$$\max\{P(X=1|L=1), P(X=1,Y=1)\} \leq P(W=1) \tag{11}$$

(Kreider and Pepper, 2001, Proposition 3A), which leads us to:

Proposition 4, Corollary: Let workers be verified and Equation (9) hold. Then

$$\begin{gathered}\frac{P(L=1,X=1)}{P(Y=1,X=1)+P(Y=0)} \\ \leq P(L=1|W=1) \leq \\ \min\left\{P(L=1), \frac{P(L=1,X=1)}{P(Y=1,X=1)}, \frac{P(L=1,X=0)}{P(Y=1,X=0)}\right\}\end{gathered} \tag{12}$$

and

$$\begin{gathered}\max\left\{P(L=1), \frac{P(L=1,X=0)}{P(Y=0)+P(Y=1,X=0)}, \frac{P(L=1,X=1)}{P(Y=0)+P(Y=1,X=1)}\right\} \\ \leq P(L=1|W=0) \leq \\ \frac{P(L=1,X=0)}{P(Y=1,X=0)}.\end{gathered}$$

These bounds, which only apply in the special case where workers are verified, improve on the Proposition 4 upper bound for $P(L=1|W=1)$ and lower bound for $P(L=1|W=0)$. For example, consider the upper bound estimate on the employment probability among the disabled.

Proposition 4 gives an upper bound of one. This corollary, however, reveals that the sharp bound is no larger than the employment rate, $P(L = 1)$. Note that although the bounds on the conditional employment probabilities are tighter, the bounds on β may be unaffected.

5.2 Results

The relevant columns of Tables 3 and 4 display the estimated bounds under the employment monotonicity assumption. Note that the upper bound on $P(L = 1|W = 1)$ is often improved compared with verification alone (in the preceding column). By assumption, the upper bounds for β are improved for all cases in which the upper bound is positive under verification alone.

5.3 Employment and Age

It also seems reasonable to assume that, conditional on true disability status, the employment rate declines weakly with age. For example, the fraction of employed 60 year-olds with disabilities may be assumed to be no greater than the fraction of employed 59 year-olds with disabilities, etc.

Consider the true employment rate at some specified value age, age_0. Formally, the MIV restriction implies the inequality restriction:

$$age_1 \leq age_0 \leq age_2 \Longrightarrow P(L = 1|W, age_2) \leq P(L = 1|W, age_0) \leq P(L = 1|W, age_1) \tag{13}$$

for all $age_1 \leq age_0$and all $age_0 \leq age_2$.

With corrupt data, the conditional probabilities in Equation (13) are not identified. However, we can bound these probabilities using the methods described above. Let $LB(age)$ and $UB(age)$ be the known lower and upper bounds, respectively, given the available information on $P(L = 1|W, age)$. Then, we have

$$\sup_{age_2 \geq age_0} LB(age_2) \leq P(L = 1|W, age_0) \leq \inf_{age_1 \leq age_0} UB(age_1) \tag{14}$$

(Manski and Pepper, Proposition 1, 2000). There are no other restrictions implied by the MIV assumption.

The MIV bound on the conditional employment rate, $P(L=1|W=1)$, is easily obtained using the law of total probability. Assuming the MIV age is finite set, the following bounds apply:

Proposition 5. If the employment rate is weakly decreasing with the monotone instrumental variable age, then:

$$\sum_{age_0 \epsilon U} P(age = age_0)\{\sup_{age_2 \geq age_0} LB(age_2)\} \tag{15}$$
$$\leq P(L=1|W) \leq$$
$$\sum_{age_0 \epsilon U} P(age = age_0)\{\inf_{age_1 \leq age_0} UB(age_1)\}.$$

Thus, to find the MIV bounds on the disability rate, one takes the appropriate weighted average of the upper and lower bounds across the different values of the instrument.

In the special case where workers (and perhaps others) are verified, the only unknown parameter in the conditional employment rate is the true incidence of disability. Thus, rather than imposing an MIV assumption on the conditional employment rate, one might instead consider an analogous restriction on the disability rate. In particular, as in Kreider and Pepper (2001), assume that the true disability rate weakly increases with age (as opposed to the conditional employment rate weakly decreasing with age). In this setting, the MIV bounds in Equations (14) and (15) apply, with the parameter of interest being the disability rate, $P(W=1)$, rather than the conditional employment rate, $P(L=1|W)$.

Formally, let $LB_w(age)$ and $UB_w(age)$ be the known lower and upper bounds, respectively, given the available information on $P(W=1|age)$. Then, the MIV bounds in Equation (15) imply the following corollary:

Proposition 5, Corollary.

$$\frac{P(L=1,X=1)}{\sum_{age_0 \epsilon U} P(age=age_0)\{\inf_{age_2 \geq age_0} UB_w(age_2)\}} \leq P(L=1|W=1) \leq \frac{P(L=1,X=1)}{\sum_{age_0 \epsilon U} P(age=age_0)\{\sup_{age_1 \leq age_0} LB_w(age_1)\}}. \tag{16}$$

5.4 Results

The last two columns in Tables 3 and 4 display the estimated bounds for $P(L = 1|W = 1)$ and β, respectively, under the assumption that age is a monotone instrumental variables.[14] For the HRS sample, we divide the sample into 25 age groups containing 500 respondents per group (except that the oldest group is somewhat larger). For the age-representative SIPP sample, the age groups are defined as 18-29, 30-31, 32-33, ..., 68-69. For each grouping, we separately apply the MIV restrictions in Equation (14) and (16). The MIV assumption alone has no identifying power in corrupt data; in the absence of additional assumptions, the conditional employment rate for each age grouping is unknown. Thus, we combine the MIV assumption with the verification and employment monotonicity assumptions in Propositions 2 and 4.[15]

The age monotonicity assumptions noticeably narrow the uncertainty regarding the conditional employment rates. When workers alone are verified, for example, the bounds on $P(L = 1|W = 1)$ using the HRS data narrow by 4 points (8 percent) when age is an MIV in employment compared with 11 points (22 percent) when age is an MIV in disability. Similarly, the bounds for β narrow by 6 points (7 percent) when age is an MIV in employment, and by 45 points (54 percent) when age is an MIV in disability. In this application, the assumption that disability is weakly increasing in age (last column) generally has substantially more identifying power than the assumption that

[14] While consistency of these bound estimates is easy to establish, the finite sample properties are not well understood (Manski and Pepper, 2000). Applications of the monotonicity bounds require taking infs and sups of collections of nonparametric regression estimates. We do not attempt to resolve these statistical questions in the present paper. With over 12,000 observations in our HRS sample and over 60,000 observations in our SIPP sample, however, we are not especially concerned about small sample biases.

[15] To save space, we present only the results obtained when both the verification and employment monotonicity assumptions are imposed. Results for different combinations of assumptions are available upon request.

the conditional employment is weakly decreasing in age (previous column).

As more groups are verified, the bounds under the MIV assumption nearly collapse to a point. When all four subgroups are assumed to provide valid reports, the MIV in employment narrows the bounds for $P(L = 1|W = 1)$ in Table 3A to the six-point range [0.333, 0.391], 50 percent narrower than the bounds without the MIV assumption. In Table 4A, β is constrained to lie within the 7 point range $[-0.404, -0.333]$, nearly a 60 percent reduction compared with the case in which MIV is not imposed. In the last column of Table 3A, the MIV assumption in disability effectively reduces the bounds for β to a point, with the estimated lower bound just exceeding the estimated upper bound. The fact that the estimated lower bound exceeds the upper bound might be evidence that the maintained MIV or verification assumptions are invalid, or it may reflect sampling variability in the estimated parameters. Since the upper bound exceeds the lower bound in the 90 percent confidence interval $[-0.353, -0.335]$, however, it seems reasonable to proceed as if the maintained assumptions are valid. That is, β appears to be nearly identified.

A striking result emerges when all four groups are assumed to provide accurate reports. In this case, the estimated MIV bounds in Table 3A do not contain the point estimate under the assumption that all respondents provide accurate reports, $P(L = 1|X = 1) = 0.295$, (nor do their 90 percent confidence intervals overlap). Since the unverified group consists of nonworkers who claim to be disabled, the results support suggestions in the literature that members of this group systematically over-report disability. Similarly, the MIV bounds on β in Table 4A exceed the point estimate $P(L = 1|X = 1) - P(L = 1|X = 0) = -0.470$ based on self-reported disability classifications. The patterns are identical using the SIPP data in Tables 3B and 4B. In particular, some of the MIV bounds for $P(L = 1|W = 1)$ in Table 3B do not contain the point estimate $P(L = 1|X = 1) = 0.358$ and some of the bounds for β in Table 4B do not contain the point estimate $P(L = 1|X = 1) - P(L = 1|X = 0) = -0.450$. Thus, if disability weakly increases with age, or if employment weakly decreases with age conditional on disability, these results suggest that conventional models which presume valid self-reports are likely to be misspecified.

6 Conclusion

Concerns over the validity of disability status measures have been central in the many debates about the labor market behavior of older persons, including the effects of Social Security retirement and disability policy. Very little is known, however, about measurement error properties of self reports of disability status (Mathiowetz and Wunderlich, 2000). Given this uncertainty, the usual approach has been to identify parameters of interest by imposing strong but unverifiable assumptions on the degree and nature of misreporting. Most studies explicitly or implicitly assume fully accurate reporting, while others have modeled the nature of misreporting in the context of conventional parametric latent variable models. These solutions, however, have done little to resolve the controversy. To the contrary, the wide range of inferences drawn under different assumptions has only highlighted the degree to which results are sensitive to the identifying assumptions.

We have taken a step back to evaluate what can be learned under weaker but more credible restrictions. The methodological innovations allow us to focus on primitive assumptions that are likely to achieve broader consensus than the models found in the existing literature. By imposing layers of arguably tenable assumptions, we can nearly identify the variation in employment rates by disability status among respondents nearing retirement age. In some cases, the estimated bounds do not include the self-reported employment rates, thus casting doubt on the validity of treating self-reports as accurate. In particular, we find evidence using the Health and Retirement Study and Survey of Income and Program Participation that nonworkers tend to overreport work limitations.

Despite this strong conclusion, much of our report highlights the uncertainty create by misreporting errors. Under modest restrictions, we obtain modest results. In many cases, the bounds provide very limited information, or none at all. The uncertainty associated with misreporting errors is only substantially reduced when we impose strong assumptions on the degree of misreporting. Certainly, efforts to learn more about the nature and extent of the misreporting process seem worthwhile.

References

[1] Anderson, Kathryn and Richard Burkhauser. (1984). "The Importance of the Measure of Health in Empirical Estimates of the Labor Supply of Older Men. *Economics Letters*, 16: 375-80.

[2] __________. (1985). "The Retirement-Health Nexus: A New Measure of an Old Puzzle." *Journal of Human Resources*, 21, 118-137.

[3] Bazzoli, Gloria. (1985). "The Early Retirement Decision: New Empirical Evidence on the Influence of Health." *Journal of Human Resources*, 20, 214-34.

[4] Benitez-Silva, Hugo, Hiu-Man Chan, John Rust, and Sofia Sheidvasser. (1997). "An Empirical Analysis of the Social Security Disability Application, Appeal, and Award Process." mimeo, Yale University.

[5] Benitez-Silva, Hugo, Moshe Buchinsky, Hiu-Man Chan, Sofia Sheidvasser, and John Rust (1999). "How Large is the Bias in Self-Reported Disability?" mimeo, Yale University.

[6] Berkovec, James and Steven Stern. (1991). "Job Exit Behavior of Older Men." *Econometrica*, 59(1): 189-210.

[7] Bound, John. (1991a). "Self-Reported Versus Objective Measures of Health in Retirement Models." *Journal of Human Resources*, 26, 106-138.

[8] __________. (1991b) "The Health and Earnings of Rejected Disability Insurance Applicants: Reply," *American Economic Review*, 81, 1991a, 1427-1434.

[9] Bound, J., Schoenbaum, M., and T. Waidmann. (1995). "Race and Education Differences in Disability Status and Labor Force Attachment in the Health and Retirement Survey." *Journal of Human Resources*, 30, Supplement, S227-S267.

[10] Bound, John and Richard Burkhauser. (1999). "Economic Analysis of Transfer Programs Targeted on People with Disabilities. In Orley Ashenfelter and David Card (Eds.), Handbook of Labor Economics, Vol. 3C. Amsterdam: Elsevier Science, 3417-3528.

[11] Bowe, Frank. (1993). "Statistics, Politics, and Employment of People with Disabilities." *Journal of Disability Policy Studies*, 4(2), 84-91.

[12] Boskin, Michael and Michael Hurd. (1978). "The Effect of Social Security on Early Retirement." *Journal of Public Economics*, 10(3): 361-77.

[13] Campolieti, Michele. (2001) "Disability and the Labor Force Participation of Older Men in Canada." University of Toronto, mimeo.

[14] Chirikos, Thomas and Gilbert Nestel. (1984). "Economic Determinants and Consequences of Self-Reported Work Disability." *Journal of Health Economics*, 3(2), 117-136.

[15] Dominitz, Jeff and Robert P. Sherman. (1998). "Identification and Estimation with Contaminated and Partially Verified Data." California Institute of Technology, mimeo.

[16] Dwyer, Debra and Olivia Mitchell. (1999). "Health Problems as Determinants of Retirement: Are Self-Rated Measures Endogenous?" *Journal of Health Economics*, 18, 173-193.

[17] Dwyer, Debra, Jianting Hu, Denton Vaughan, and Bernard Wixon. (2001). "Counting the Disabled: Using Survey Self-Reports to Estimate Medical Eligibility for Social Security's Disability Programs." Social Security Administration, ORES Working Paper #90.

[18] Efron, Bradley and Robert J. Tibshirani. (1993). An Introduction to the Bootstrap. Chapman and Hall. New York: NY.

[19] Gordon, Roger and Alan Blinder. (1980). "Market Wages, Reservation Wages, and Retirement Decisions." *Journal of Public Economics*, 14, 277-308.

[20] Hanoch, Giora and Marjorie Honig. (1983). "Retirement, Wages, and Labor Supply of the Elderly." *Journal of Labor Economics*, 1(2): 131-51.

[21] Haveman, Robert, and Barbara Wolfe. (1984). "The Decline in Male Labor Force Participation: Comment." *Journal of Political Economy*, 92, 532-541.

[22] Horowitz, Joel and Charles F. Manski. (1995). "Identification and Robustness With Contaminated and Corrupted Data," *Econometrica* 63(2), 281-02.

[23] Horowitz, Joel and Charles F. Manski. (1998). "Censoring of Outcomes and Regressors Due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations, " *Journal of Econometrics*, 84, 37-58.

[24] Kerkhofs and Lindeboom (1995). "Subjective Health Measures and State Dependent Reporting Errors." *Health Economics*, 4:221-35.

[25] Kreider, Brent. (1999). "Latent Work Disability and Reporting Bias." *Journal of Human Resources*, Fall, 734-769.

[26] __________. (2000) "Behavioral Responses to Changes in Federal Disability Policy: The Role of Measured Disability on Inferences," in Research in Human Capital and Development: Essays in the Economics of Disability, David Salkever and Alan Sorkin (eds.), JAI Press: 81-107.

[27] Kreider, Brent and John V. Pepper (2001). "Inferring Disability Status from Corrupt Data." Thomas Jefferson Center for Political Economy Working Paper Series, University of Virginia, Charlottesville, VA.

[28] Lambert, D. and L. Tierney (1997). "Nonparametric Maximum Likelihood Estimation from Samples with Irrelevant Data and Verification Bias," *Journal of the American Statistical Association,* 92: 937-944.

[29] Manski, Charles F. and John V. Pepper. (2000). "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68(4): 997-1010

[30] Manski, Charles F. (1995). *Identification Problems in the Social Sciences*. Harvard University Press, Cambridge, MA.

[31] Mathiowetz, N. and G. S. Wunderlich, eds. Survey Measurement of Work Disability: Summary of a Workshop. IOM, NRC. National Academy Press, 2000.

[32] McGarry, Kathleen, "Health and Retirement: Do Changes in Health Affect Retirement Expectations," Working Paper, 2002.

[33] Myers, Robert. (1982). "Why Do People Retire from Work Early?" *Social Security Bulletin*. September: 10-14.

[34] O'Donnell, Owen. (1998). "The Effect of Disability on Employment Allowing for Work Incapacity." mimeo. University of Kent.

[35] Parsons, Donald. (1984). "Disability Insurance and Male Labor Force Participation: A Response to Haveman and Wolfe," *Journal of Political Economy*, 92: 542-549.

[36] ________. (1991). "The Health and Earnings of Rejected Disability Insurance Applicants: Comment," *American Economic Review*, 81: 1419-1426.

[37] ________. (1996). "Imperfect 'Tagging' in Social Insurance Programs." *Journal of Public Economics*, 62: 183-207.

[38] Slade, Frederick. (1984). "Older Men: Disability Insurance and the Incentive to Work." *Industrial Relations*, Spring 23: 260-69.

[39] Stern, Steven. (1989). "Measuring the Effect of Disability on Labor Force Participation." *Journal of Human Resources* 24: 361-395.

[40] Wunderlich, G. S., editor. Measuring Functional Capacity and Work Requirements: Summary of a Workshop IOM, NRC. National Academy Press, 1999.

7 Appendix

7.1 Proof of Propositions 1 and 3

Proof of Proposition 1. (degree bounds) $P(Z=1) \geq v$:

Decompose the conditional probability in Equation (3) as follows:

$$P(L=1|W=1) = \frac{P(L{=}1,X{=}1)+P(L{=}1,X{=}0,Z{=}0)-P(L{=}1,X{=}1,Z{=}0)}{P(X{=}1)+P(L{=}1,X{=}0,Z{=}0)+P(L{=}0,X{=}0,Z{=}0)-P(L{=}1,X{=}1,Z{=}0)-P(L{=}0,X{=}1,Z{=}0)}.$$

Let $b = P(L=1, X=1, Z=0)$ where $0 \leq b \leq \min[(1-v), P(L=1, X=1)]$, and let $a = P(L=1, X=0, Z=0)$ where $0 \leq a \leq \min[(1-v), P(L=1, X=0)]$. Then, for conjectured values of a and b, it follows that

$$\frac{P(L=1, X=1) - b}{P(X=1) - b + \min\{(1-v) - b, P(L=0, X=0)\}} \tag{17}$$

$$\leq P(L=1|W=1) \leq \tag{18}$$

$$\frac{P(L=1, X=1) + a}{P(X=1) + a - \min\{(1-v) - a, P(L=0, X=1)\}}.$$

Since a and b are unknown parameters, the bounds in Equation (17) are not identified. Instead, bounds are identified by finding the values of $\{a, b\}$ which maximize the upper bound and minimize the lower bound. First notice that these extremum are only realized if $(1-v)-b \leq P(L=0, X=0)$ and $(1-v) - a \leq P(L=0, X=1)$, in which case Equation (17) simplifies to

$$\frac{P(L=1, X=1) - b}{P(X=1) - 2b + (1-v)} \leq P(L=1|W=1) \leq \frac{P(L=1, X=1) + a}{P(X=1) + 2a - (1-v)}. \tag{19}$$

Differentiating this bound with respect to a and b reveals that the lower bound is minimized when $b = \delta$ and the upper bound is maximized with $a = \gamma$. Proposition 1 follows. □

Proof of Proposition 3. (partial verification) $P(Z=1|Y=1) \geq v_y$:

Proposition 3 follows Proposition 1, except now there is one subgroup, $Y = 0$, for which there is no prior information. Using Bayes' Theorem, $P(L=1|W=1) = \frac{P(L=1,W=1)}{P(W=1)}$. Decompose the numerator as

$$P(L=1,W=1) = P(L=1,X=1,Y=1) + P(L=1,W=1,Y=0)$$
$$+ P(L=1,X=0,Y=1,Z=0) - P(L=1,X=1,Y=1,Z=0)$$

and decompose the denominator as

$$P(W=1) = P(X=1,Y=1) + P(L=1,W=1,Y=0) + P(L=0,W=1,Y=0)$$
$$+ P(L=1,X=0,Y=1,Z=0) + P(L=0,X=0,Y=1,Z=0)$$
$$- P(L=1,X=1,Y=1,Z=0) - P(L=0,X=1,Y=1,Z=0).$$

Let $b = P(L=1,X=1,Y=1,Z=0)$ where $0 \leq b \leq \min[(1-v_y)P(y=1), P(L=1,X=1,Y=1)]$, and let $a = P(L=1,X=0,Y=1,Z=0)$ where $0 \leq a \leq \min[(1-v_y)P(Y=1), P(L=1,X=0,Y=1)]$. Then, for conjectured values of a and b, it follows that

$$\frac{P(L=1,X=1,Y=1)-b}{P(X=1,Y=1)+P(L=0,Y=0)-b+\min\{(1-v_y)P(Y=1)-b, P(L=0,X=0,Y=1)\}} \tag{20}$$
$$\leq P(L=1|W=1) \leq$$
$$\frac{P(L=1,X=1,Y=1)+P(L=1,Y=0)+a}{P(X=1)+P(L=1,Y=0)+a-\min\{(1-v_y)P(Y=1)-a, P(L=0,X=1,Y=1)\}}.$$

Since a and b are unknown parameters, the bounds in Equation (20) are not identified. Bounds are identified by finding the values of $\{a,b\}$ which maximize the upper bound and minimize the lower bound. First notice that these extremum are only realized if $(1-v_y)P(Y=1)-b \leq P(L=0,X=0,Y=1)$ and $(1-v_y)P(Y=1)-a \leq P(L=0,X=1,Y=1)$, in which case Equation (17) simplifies to

$$\frac{P(L=1,X=1,Y=1)-b}{P(X=1,Y=1)+P(L=0,Y=0)-2b+(1-v_y)P(Y=1)} \tag{21}$$
$$\leq P(L=1|W=1) \leq$$
$$\frac{P(L=1,X=1,Y=1)+P(L=1,Y=0)+a}{P(X=1)+P(L=1,Y=0)+2a-(1-v_y)P(Y=1)}.$$

Differentiating this bound with respect to a and b reveals that the lower bound is minimized when $b = \delta$ and the upper bound is maximized with $a = \gamma$. Proposition 3 follows. $\square$

7.2 Sharp Bounds for β

In Proposition 1, we present bounds on β found by differencing respective bounds on the two conditional employment probabilities. In particular, the upper (lower) bound on β is found subtracting the lower (upper) bound on $P(L = 1|W = 0)$ from the upper (lower) bound on $P(L = 1|W = 1)$. As shown next, the bounds formed under the degree assumptions considered in Section 3 can be improved upon in some cases. In contrast, these naive bounds on β are sharp under the full verification assumption explored in Section 4.

7.2.1 Degree Assumption (Section 3)

Suppose one has prior information on the maximum degree of inaccurate, $P(Z = 1) \geq v$. Using the same logic as in Proposition 1, we can also bound $P(L = 1|W = 0)$:

$$\frac{P(L=1, X=0) - a}{P(X=0) - 2a + (1-v)} \leq P(L=1|W=0) \leq \frac{P(L=1, X=0) + b}{P(X=1) + 2b - (1-v)}. \tag{22}$$

Combining Equations (17) and (22), we have:

Proposition 1A. Let $P(Z = 1) \geq v$. Then

$$\inf_{b \in (0, \min[(1-v), P(L=1,X=1)])} \left[\frac{P(L=1, X=1) - b}{P(X=1) - 2b + (1-v)} - \frac{P(L=1, X=0) + b}{P(X=0) + 2b - (1-v)} \right]$$
$$\leq \beta \leq \tag{23}$$
$$\sup_{a \in (0, \min[(1-v), P(L=1,X=0)])} \left[\frac{P(L=1, X=1) + a}{P(X=1) + 2a - (1-v)} - \frac{P(L=1, X=0) - a}{P(X=0) - 2a + (1-v)} \right].$$

Over part of the range for v, these bounds differ from the naive bounds obtained from Proposition 1. Consider, for example, the lower bound in Proposition 1A. If the value of the unknown parameter b that minimizes the first expression (i.e., the lower bound on $P(L = 1|W = 1)$ differs from the value of b that maximizes the second expression (i.e., the upper bound on $P(L=1|W=0)$), the two bounds on β will differ and the Proposition 1A bounds will be tighter. The two bounds will be identical when the lower bound on $P(L=1|W=1)$ and the upper bound on $P(L=1|W=0)$ are realized at same value of the unknown parameter b.

7.2.2 Verification Assumption (Section 4)

Under full verification of subgroups, the unknown parameters that minimize $P(L=1|W=1)$ also maximize $P(L=1|W=0)$ so that the naive bonds on β are sharp. To see this, write β as

$$\begin{aligned}\beta &= \frac{P(L=1,W=1)}{P(W=1)} - \frac{P(L=1)-P(L=1,W=1)}{1-P(W=1)} \\ &= \frac{P(L=1,W=1,Y=1)+P(L=1,W=1,Y=0)}{P(W=1,Y=1)+P(W=1,Y=0)} \\ &\quad -\frac{P(L=1)-P(L=1,W=1,Y=1)+P(L=1,W=1,Y=0)}{1-P(W=1,Y=1)-P(W=1,Y=0)}.\end{aligned}$$

Sharp bounds on β follow:

Proposition 2A:

$$\frac{P(L=1,W=1,Y=1)}{P(W=1,Y=1)+P(L=0,Y=0)} - \frac{P(L=1)-P(L=1,W=1,Y=1)}{1-P(W=1,Y=1)-P(L=0,Y=0)}$$

$$\leq \beta \leq \tag{24}$$

$$\frac{P(L=1,W=1,Y=1)+P(L=1,Y=0)}{P(W=1,Y=1)+P(L=1,Y=0)} - \frac{P(L=1)-P(L=1,W=1,Y=1)-P(L=1,Y=0)}{1-P(W=1,Y=1)-P(L=1,Y=0)}.$$

Notice that these are the naive bounds described above. Consider, for example, the lower bound on β. The first term in this bound equals the lower bound on $P(L=1|W=1)$ while the second term equals the upper bound on $P(L=1|W=0)$.

Figure 1
Bounds for P(L=1|W=1)
under Propositions 1 and 2

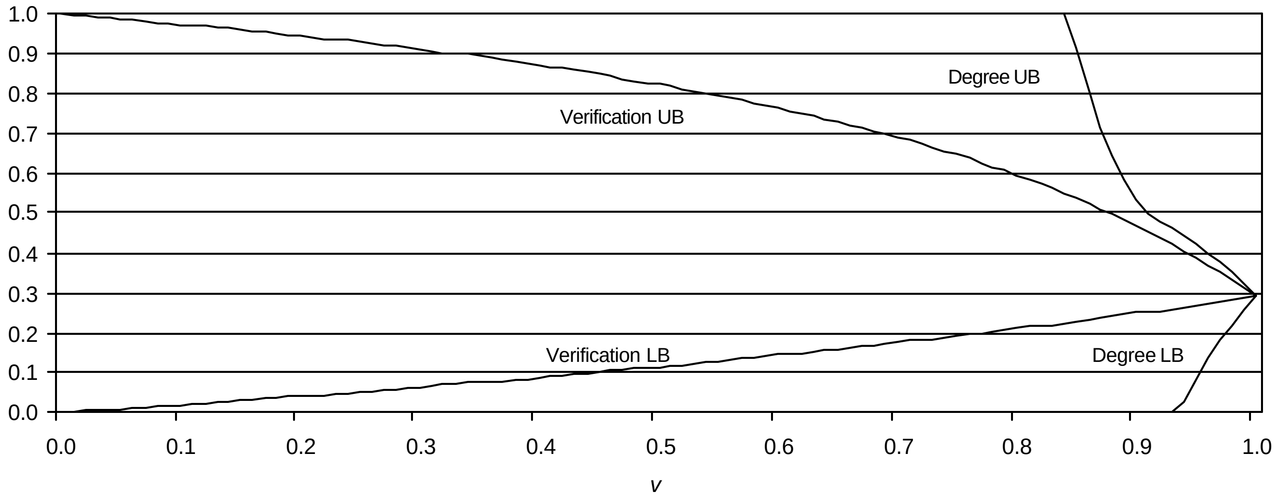

Figure 2
Bounds for P(L=1|W=0)
under Propositions 1 and 2

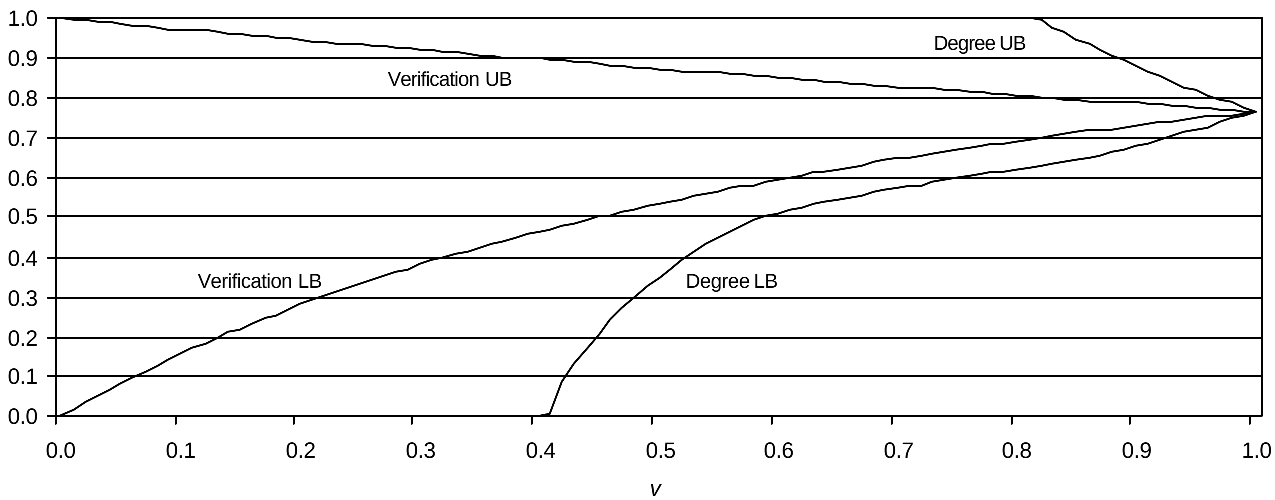

Figure 3
Bounds on Beta: P(L=1|W=1)-P(L=1|W=0)

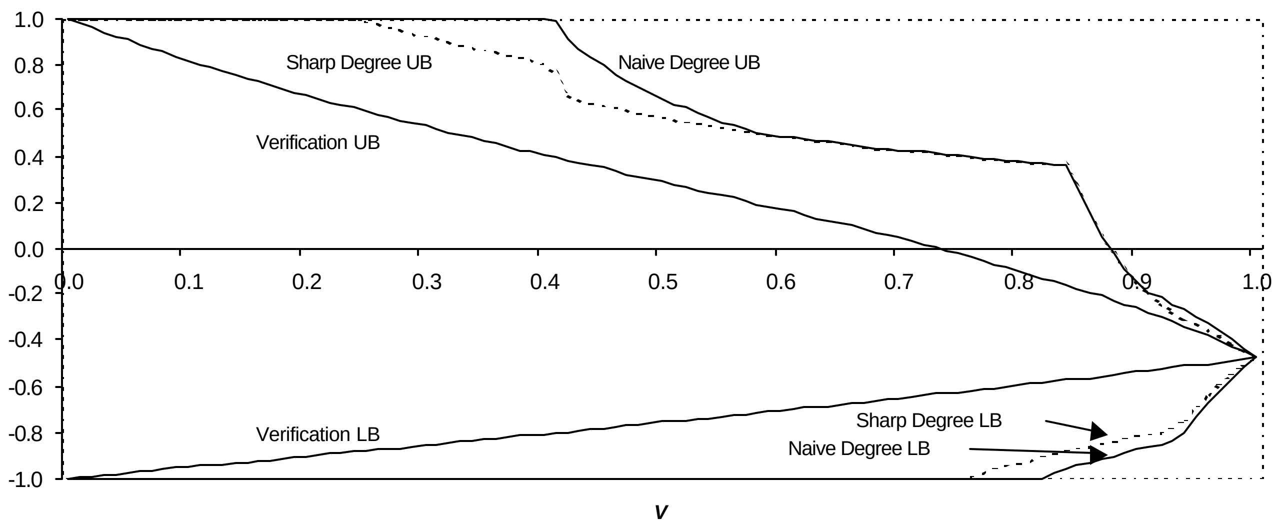

Table 1

Descriptive Statistics

Means and Standard Deviations of Primary Variables

	HRS (N=12,652)		SIPP (N=60,633)	
	mean	standard deviation	mean	standard deviation
Work-limited (self-reported)	.215	.411	.131	.337
Disability precludes work (self-reported)	.094	.291	.070	.255
'Yes' to either of the above (*X*=1)	.218	.413	.131	.337
Labor force participant (*L*=1)	.662	.473	.750	.434
Current receipt of disability income†	.100	.300	.030	.172
Age	55.6	5.66	40.4	13.6
Years of schooling	12.0	3.27	12.8	2.91
High school graduate	.708	.455	.850	.358
College graduate	.175	.380	.211	.408
Nonwhite race	.280	.449	.178	.382

Table 2

Conditional Employment Probabilities
by Self-reported Disability Status

	All		Work Limitation (self-reported)		No Work Limitation (self-reported)	
	HRS (N=12,652)	SIPP (N=60,633)	HRS (N=2764)	SIPP (N=7949)	HRS (N=9888)	SIPP (N=52,688)
All Respondents	.662	.749	.295	.358	.765	.808
Gender						
Men	.715	.823	.327	.391	.831	.885
Women	.617	.684	.265	.330	.709	.738
Age						
18-49	.753	.813	.478	.468	.623	.848
50-54	.737	.785	.341	.368	.828	.877
55-59	.662	.670	.286	.284	.779	.792
60-64	.555	.462	.245	.185	.662	.569
65-69	.316	.256	.119	.135	.409	.305
70+	.224	--	.087	--	.287	--

Table 3A

Bounds for P(L=1|W=1)[a] and
90% Confidence Intervals using the HRS (N=12,652)

Verified Group	v	Proposition 1 Degree Bounds	Proposition 2 Verification	Proposition 4 and Corollary Employment Monotonicity + Verification	Proposition 5 Age MIV in Employment[b] + Employment Monotonicity + Verification	Proposition 5 Corollary Age MIV in Disability[b] + Employment Monotonicity + Verification
beneficiaries	0.100	[0.000, 1.000]	[0.013, 0.897]	[0.013, 0.897]	[0.027, 0.886]	
		[0.000 1.000]	[0.010 0.902]	[0.010 0.902]	[0.018 0.895]	
wave 2 verification	0.267	[0.000, 1.000]	[0.044, 0.954]	[0.044, 0.838]	[0.077, 0.818]	
		[0.000 1.000]	[0.038 0.958]	[0.038 0.843]	[0.055 0.828]	
workers	0.662	[0.000, 1.000]	[0.160, 1.000]	[0.160, 0.662]	[0.187, 0.648]	[0.169, 0.561]
		[0.000 1.000]	[0.151 1.000]	[0.151 0.670]	[0.177 0.662]	[0.154 0.573]
claim no disability	0.782	[0.000, 1.000]	[0.000, 1.000]	[0.000, 0.783]	[0.000, 0.756]	
		[0.000 1.000]	[0.000 1.000]	[0.000 0.790]	[0.000 0.771]	
workers+ beneficiaries	0.745	[0.000, 1.000]	[0.164, 0.465]	[0.164, 0.465]	[0.189, 0.445]	[0.173, 0.438]
		[0.000 1.000]	[0.154 0.485]	[0.154 0.485]	[0.179 0.468]	[0.157 0.453]
wave 2 verification	0.816	[0.000, 1.000]	[0.234, 0.703]	[0.234, 0.662]	[0.264, 0.635]	[0.247, 0.561]
		[0.000 1.000]	[0.221 0.725]	[0.221 0.670]	[0.255 0.648]	[0.223 0.572]
claim no disability	0.846	[0.000, 1.000]	[0.295, 1.000]	[0.295, 0.662]	[0.333, 0.648]	[0.308, 0.561]
		[0.000 0.999]	[0.280 1.000]	[0.280 0.670]	[0.315 0.662]	[0.289 0.573]
all of the above	0.938	[0.014, 0.452]	[0.295, 0.413]	[0.295, 0.413]	[0.333, 0.391]	[0.308, 0.390]
		[0.000 0.462]	[0.280 0.431]	[0.280 0.431]	[0.315 0.415]	[0.289 0.417]

[a] Self-reported value: P(L=1|X=1) = 0.295
[b] 500 observations per age group

Table 3B

Bounds for P(L=1|W=1)[a] and
90% Confidence Intervals using the SIPP (N=60,633)

Verified Group	v	Proposition 1 Degree Bounds	Proposition 2 Verification	Proposition 4 and Corollary Employment Monotonicity + Verification	Proposition 5 Age MIV in Employment[b] + Employment Monotonicity + Verification	Proposition 5 Corollary Age MIV in Disability[b] + Employment Monotonicity + Verification
beneficiaries	0.030	[0.000, 1.000]	[0.012, 0.967]	[0.012, 0.967]	[0.025, 0.957]	
		[0.000 1.000]	[0.011 0.968]	[0.011 0.968]	[0.020 0.959]	
wave 2 verification	0.156	[0.000, 1.000]	[0.012, 0.981]	[0.012, 0.863]	[0.019, 0.823]	
		[0.000 1.000]	[0.010 0.982]	[0.010 0.865]	[0.014 0.827]	
workers	0.749	[0.000, 1.000]	[0.158, 1.000]	[0.158, 0.749]	[0.241, 0.717]	[0.167, 0.681]
		[0.000 1.000]	[0.154 1.000]	[0.154 0.752]	[0.221 0.722]	[0.159 0.691]
claim no disability	0.869	[0.000, 1.000]	[0.000, 1.000]	[0.000, 0.818]	[0.000, 0.756]	
		[0.000 1.000]	[0.000 1.000]	[0.000 0.820]	[0.000 0.771]	
workers+ beneficiaries	0.776	[0.000, 1.000]	[0.158, 0.644]	[0.158, 0.644]	[0.241, 0.670]	[0.167, 0.597]
		[0.000 1.000]	[0.154 0.656]	[0.154 0.656]	[0.223 0.680]	[0.160 0.611]
wave 2 verification	0.882	[0.000, 1.000]	[0.262, 0.765]	[0.262, 0.749]	[0.339, 0.717]	[0.277, 0.681]
		[0.000 1.000]	[0.255 0.776]	[0.255 0.752]	[0.316 0.722]	[0.267 0.691]
claim no disability	0.916	[0.000, 1.000]	[0.358, 1.000]	[0.358, 0.749]	[0.431, 0.717]	[0.367, 0.681]
		[0.000 1.000]	[0.350 1.000]	[0.350 0.752]	[0.417 0.722]	[0.354 0.691]
all of the above	0.954	[0.014, 0.551]	[0.358, 0.551]	[0.358, 0.551]	[0.431, 0.627]	[0.367, 0.539]
		[0.000 0.562]	[0.350 0.562]	[0.350 0.562]	[0.417 0.641]	[0.354 0.553]

[a] Self-reported value: P(L=1|X=1) = 0.358
[b] Age groups: 18-29, 30-31, 32-33, ..., 68-69

Table 4A

Bounds for $ = P(L=1|W=1)-P(L=1|W=0)[a]
and 90% Confidence Intervals using the HRS (N=12,652)

		Proposition 1	Proposition 2	Proposition 4 and Corollary	Proposition 5	Proposition 5 Corollary
Verified Group	v	Degree Bounds	Verification	Employment Monotonicity + Verification	Age MIV in Employment[b] + Employment Monotonicity + Verification	Age MIV in Disability[b] + Employment Monotonicity + Verification
beneficiaries	0.100	[-1.000, 1.000]	[-0.975, 0.847]	[-0.975, 0.000]	[-0.953, 0.000]	
		[-1.000 1.000]	[-0.978 0.856]	[-0.978 0.000]	[-0.960 0.000]	
wave 2 verification	0.267	[-1.000, 1.000]	[-0.793, 0.703]	[-0.793, 0.000]	[-0.740, 0.000]	
		[-1.000 1.000]	[-0.803 0.714]	[-0.803 0.000]	[-0.761 0.000]	
workers	0.662	[-1.000, 0.449]	[-0.840, 0.361]	[-0.840, 0.000]	[-0.813,-0.029]	[-0.475,-0.088]
		[-1.000 0.460]	[-0.849 0.368]	[-0.849 0.000]	[-0.823 -0.018]	[-0.491 -0.073]
claim no disability	0.782	[-1.000, 0.388]	[-0.783, 0.361]	[-0.783, 0.000]	[-0.756, 0.000]	
		[-1.000 0.395]	[-0.790 0.368]	[-0.790 0.000]	[-0.776 0.000]	
workers+ beneficiaries	0.745	[-1.000, 0.405]	[-0.823,-0.229]	[-0.823,-0.229]	[-0.788,-0.266]	[-0.471,-0.284]
		[-1.000 0.414]	[-0.833 -0.207]	[-0.833 -0.207]	[-0.796 -0.248]	[-0.487 -0.276]
wave 2 verification	0.816	[-1.000, 0.373]	[-0.591, 0.045]	[-0.591, 0.000]	[-0.539,-0.042]	[-0.397,-0.115]
		[-1.000 0.380]	[-0.607 0.069]	[-0.607 0.000]	[-0.549 -0.032]	[-0.414 -0.100]
claim no disability	0.846	[-0.953, 0.361]	[-0.470, 0.361]	[-0.470, 0.000]	[-0.404,-0.029]	[-0.335,-0.088]
		[-0.966 0.369]	[-0.487 0.368]	[-0.487 0.000]	[-0.414 -0.018]	[-0.353 -0.073]
all of the above	0.938	[-0.817,-0.257]	[-0.470,-0.296]	[-0.470,-0.296]	[-0.404,-0.333]	[-0.335,-0.359]
		[-0.839 -0.244]	[-0.487 -0.276]	[-0.487 -0.276]	[-0.414 -0.312]	[-0.353 -0.335]

[a] Self-reported value: P(L=1|X=1)-P(L=1|X=0) = -0.470
[b] 500 observations per age group

Table 4B

Bounds for $ = P(L=1|W=1)-P(L=1|W=0)[a] using the SIPP (N=60,633)

		Proposition 1	Proposition 2	Proposition 4 and Corollary	Proposition 5	Proposition 5 Corollary
Verified Group	v	Degree Bounds	Verification	Employment Monotonicity + Verification	Age MIV in Employment[b] + Employment Monotonicity + Verification	Age MIV in Disability[b] + Employment Monotonicity + Verification
beneficiaries	0.030	[-1.000, 1.000]	[-0.987, 0.964]	[-0.987, 0.000]	[-0.972, 0.000]	
		[-1.000 1.000]	[-0.988 0.966]	[-0.988 0.000]	[-0.977 0.000]	
wave 2 verification	0.156	[-1.000, 1.000]	[-0.851, 0.897]	[-0.851, 0.000]	[-0.804, 0.000]	
		[-1.000 1.000]	[-0.854 0.901]	[-0.854 0.000]	[-0.812 0.000]	
workers	0.749	[-1.000, 0.319]	[-0.842, 0.263]	[-0.842, 0.000]	[-0.759,-0.055]	[-0.565,-0.075]
		[-1.000 0.323]	[-0.846 0.266]	[-0.846 0.000]	[-0.779 -0.046]	[-0.574 -0.061]
claim no disability	0.869	[-0.951, 0.277]	[-0.818, 0.263]	[-0.818, 0.000]	[-0.758,-0.001]	
		[-0.956 0.280]	[-0.820 0.266]	[-0.820 0.000]	[-0.762 0.000]	
workers+ beneficiaries	0.776	[-1.000, 0.309]	[-0.841,-0.113]	[-0.841,-0.113]	[-0.755,-0.112]	[-0.564,-0.226]
		[-1.000 0.313]	[-0.845 -0.101]	[-0.845 -0.101]	[-0.773 -0.106]	[-0.573 -0.211]
wave 2 verification	0.882	[-0.935, 0.273]	[-0.594, 0.017]	[-0.594, 0.000]	[-0.476,-0.055]	[-0.454,-0.127]
		[-0.938 0.276]	[-0.601 0.028]	[-0.601 0.000]	[-0.501 -0.046]	[-0.463 -0.109]
claim no disability	0.916	[-0.894, 0.263]	[-0.450, 0.263]	[-0.450, 0.000]	[-0.317,-0.055]	[-0.364,-0.075]
		[-0.898 0.267]	[-0.458 0.266]	[-0.458 0.000]	[-0.335 -0.046]	[-0.379 -0.061]
all of the above	0.954	[-0.839,-0.217]	[-0.450,-0.217]	[-0.450,-0.217]	[-0.317,-0.163]	[-0.364,-0.326]
		[-0.855 -0.206]	[-0.458 -0.206]	[-0.458 -0.206]	[-0.335 -0.152]	[-0.379 -0.300]

[a] Self-reported value: P(L=1|X=1)-P(L=1|X=0) = -0.450

[b] Age groups: 18-29, 30-31, 32-33, ..., 68-69