

4-2008

# Inferring disability status from corrupt data

Brent Kreider

*Iowa State University*, [bkreider@iastate.edu](mailto:bkreider@iastate.edu)

John Pepper

*University of Virginia*

Follow this and additional works at: [https://lib.dr.iastate.edu/econ\\_las\\_pubs](https://lib.dr.iastate.edu/econ_las_pubs)

 Part of the [Behavioral Economics Commons](#), [Econometrics Commons](#), [Health Economics Commons](#), and the [Social Welfare Commons](#)

The complete bibliographic information for this item can be found at [https://lib.dr.iastate.edu/econ\\_las\\_pubs/638](https://lib.dr.iastate.edu/econ_las_pubs/638). For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

---

This Article is brought to you for free and open access by the Economics at Iowa State University Digital Repository. It has been accepted for inclusion in Economics Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

---

# Inferring disability status from corrupt data

## **Abstract**

In light of widespread concerns about the reliability of self-reported disability, we investigate what can be learned about the prevalence of work disability under various assumptions on the reporting error process. Developing a nonparametric bounding framework, we provide tight inferences under our strongest assumptions but then find that identification deteriorates rapidly as the assumptions are relaxed. For example, we find that inferences are highly sensitive to how one models potential inconsistencies between subjective self-assessments of work limitation and more objective measures of functional limitation. These two indicators appear to measure markedly different aspects of health status.

## **Disciplines**

Behavioral Economics | Econometrics | Health Economics | Social Welfare

## **Comments**

This is the peer reviewed version of the following article: Kreider, Brent, and John Pepper. "Inferring disability status from corrupt data." *Journal of Applied Econometrics* 23, no. 3 (2008): 329-349, which has been published in final form at doi: [10.1002/jae.979](https://doi.org/10.1002/jae.979). This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions.

## Inferring Disability Status from Corrupt Data

Brent Kreider

Department of Economics  
Iowa State University  
bkreider@iastate.edu

John Pepper\*

Department of Economics  
University of Virginia  
jvpepper@virginia.edu

February 2007

**Abstract.** In light of widespread concerns about the reliability of self-reported disability, we investigate what can be learned about the prevalence of work disability under various assumptions on the reporting error process. Developing a nonparametric bounding framework, we provide tight inferences under our strongest assumptions but then find that identification deteriorates rapidly as the assumptions are relaxed. For example, we find that inferences are highly sensitive to how one models potential inconsistencies between subjective self-assessments of work limitation and more objective measures of functional limitation. These two indicators appear to measure markedly different aspects of health status.

*JEL classification:* C14, I1

---

Acknowledgments: The authors received valuable comments from the editor, John Rust, two anonymous referees, Christopher Bollinger, John Bound, Helle Bunzel, Charles Manski, Francesca Molinari, Steven Stern, Justin Tobias, and from seminar participants at Cornell, Pennsylvania, Michigan, Iowa State, Syracuse, Virginia and Virginia Commonwealth University. They also benefited from discussions at meetings of the Econometric Society, Allied Social Science Association, International Health Economics Association, Northwestern/University of Chicago Joint Center for Poverty Research, Midwest Economic Association, and Southern Economic Association. This research was supported in part by the Joint Center for Poverty Research, the Sandell Grant Program, the W.E. Upjohn Institute for Employment Research, and the Bankard Fund for Political Economy.

\*Corresponding author: Department of Economics, University of Virginia, 2015 Ivy Road, Rm 417, Charlottesville, VA 22904, 434-924-3402 (ph), 434-982-2904 (fax)

## 1. INTRODUCTION

Health status has long been recognized as a crucial determinant of many important economic decisions, including choices about whether to participate in the labor force or enroll in public transfer programs. Yet there exists widespread concern about the reliability of self-reported health and disability in survey datasets (e.g., Institute of Medicine, 2002). Ongoing debates about measuring the presence of work-limiting disabilities, the effects of health on labor market decisions, and the influence of Social Security Disability Insurance (SSDI) policy on declining labor force participation rates have all emphasized issues regarding the reliability of self-reported disability information (e.g., Haveman and Wolfe (1984) vs. Parsons (1984); Bound (1991) vs. Parsons (1991)).

This paper focuses on the problem of drawing inferences on the prevalence of long-term work disability using self-reports of work capacity. Numerous studies measure disability status based on subjective self-reports of limitation, such as responses to questions of the form: “Do you have a health impairment that limits the kind or amount of work you can perform?” We examine the prevalence of “true disability” among respondents in the Health and Retirement Study (HRS), a survey of persons nearing retirement commonly used to evaluate the effects of disability on the work behavior of older persons. In the HRS, nearly 21% of the respondents report having a long-term work limitation caused by a medical problem; about half of these respondents report being unable to work altogether.<sup>1</sup>

Many researchers are skeptical of the accuracy of these self-reports. Bound and Burkhauser (1999, p. 3446), for example, suggest the possibility that “those who apply for SSDI and especially those who are awarded benefits tend to exaggerate the extent of their work limitations (relative to those who do not apply)...” Eligibility for disability transfers is specifically tied to diminished work capacity. Others (e.g., Bowe, 1993) have argued that the threshold for claiming disability may be lower for those who find themselves out of the labor force, either by choice or through involuntarily unemployment. Some who have withdrawn from the labor force prior to normal retirement age might rationalize their employment status as driven mostly by their health conditions instead of by other factors, such as high preferences for leisure or unlucky labor market outcomes.

Studies that have modeled and assessed the reliability of self-reported work limitations have not come to any consensus. Using a variety of parametric latent variable models to assess the impact of health on labor market outcomes, several researchers have found evidence of systematic disability

---

<sup>1</sup>In a companion paper (Kreider and Pepper, forthcoming), we consider the consequences of arbitrary classification errors when disability status is a potentially mismeasured conditioning variable.

reporting errors. Kerkhofs and Lindeboom (1995), and Kreider (1999, 2000), for example, estimate large reporting errors that are related to labor force status. In contrast, Stern (1989) and Dwyer and Mitchell (1999) accept the hypothesis that labor market outcomes do not affect reporting behavior. These conflicting findings have proved difficult to reconcile. Most related studies impose what seem to be sensible restrictions on the reporting process. As noted by Benítez-Silva et al. (2004), however, most of the earlier approaches required strong parametric assumptions and behavioral restrictions.

To disentangle these issues, Benítez-Silva et al. (2004) isolate the problem of inferring disability status. Using an innovative approach that focuses on a subsample of applicants for federal disability benefits, they compare self-reports of work incapacity to the Social Security Administration’s (SSA) award decision. Under the identifying assumption that the SSA’s definition of disability forms the social standard for what constitutes work incapacity, they find that disability self-reports are unbiased.

Given the ongoing debates about measuring work limitations, we similarly focus on the narrow but complex problem of inferring disability rates from self-reported survey data. In contrast to Benítez-Silva et al. (2004), we assess disability among the general population of individuals nearing retirement age and thus do not observe an alternative direct measure of work limitation. Instead, we develop and apply a nonparametric bounding methodology that allows us to assess the identifying power of some basic assumptions about the reporting process that have been applied in the literature.

We describe the data and different measures of limitation in Section 2. In Section 3, we develop a methodological framework to infer disability in corrupt data in which we assume, initially, that nothing is known about the patterns of reporting errors. Extending the partial identification methods developed by Horowitz and Manski (1995), we consider what can be learned under different restrictions on the reporting process. This framework allows one to assess the sensitivity of inferences about work disability to the strength of the identifying assumptions. Two classes of assumptions are considered: first, we consider “verification” assumptions that formalize the notion of placing more confidence in some responses than others (e.g., depending on corroborating medical evidence); and second, we consider monotone instrumental variable (MIV) assumptions that specify monotonic relationships between the true disability rate and certain observed covariates, such as labor force participation and age.

In Sections 4 and 5, we present results and draw conclusions. We first study what can be learned

about the prevalence of work disability in the general population. We then turn to inferences for the subsample of disability insurance applicants, the group studied by Benítez-Silva et al. (2004). Since we observe no objective measure of true work capacity, there invariably will be questions about the credibility of any verification or MIV assumption. Thus, a primary objective is to assess how inferences vary under different seemingly reasonable restrictions. To do so, we exploit the wealth of information available in the HRS on health and labor market status to motivate and assess the identifying power of different assumptions. For example, we might have more confidence that a respondent truly has a significant work limitation if the respondent also reports a serious, objectively diagnosed health condition that is known to be associated with disability (e.g., having had a stroke).

## 2. DATA

Our analysis uses data from the Health and Retirement Study, a nationally representative survey of 7608 households whose heads were nearing retirement age (aged 51-61) at the time of the initial interview in 1992-93. The HRS has become an especially popular data source for studying the effects of health status and public policy on the work behavior of older persons because of its detailed information about health and disability, work history, and participation in public transfer programs. The first wave is comprised of 12,652 respondents (heads and other adult household members). As common in micro analyses of the HRS data, we restrict our sample to the 9,824 age-eligible respondents born between 1931 and 1941.

We focus on inferring long-term disability rates in the first wave of the survey using responses to direct questions on work limitation. HRS respondents were asked, “Do you have any impairment or health problem that limits the kind or amount of paid work you can do?” Those who answered in the affirmative to this broad disability question were asked the more narrow question: “Does this limitation keep you from working altogether?” Of the 9824 respondents, 2039 (20.8% of the sample) reported a long-term work limitation, and 992 (10.1% of the sample) reported being unable to work altogether. We also use information from the second wave, conducted two years after the first wave, to help resolve uncertainty about pending applications for federal disability benefits.

Responses to these work limitation questions provide convenient summary measures of disability and are often viewed to be more informative about work capacity than more objective yet indirect proxies, such as the presence of specific health conditions or functional limitations (e.g., Haveman and Wolfe, 1984).<sup>2</sup> In particular, these direct disability questions capture the notion of both

---

<sup>2</sup>Using the HRS data, Benítez-Silva et al. (1999) find that self-reported disability status constitutes a powerful

physical and mental limitations as well as the more elusive ideas involving social context. These ideas are reflected in Nagi’s (1965) seminal work relating disability to “the expression of a physical or a mental limitation” in a social context such as the workplace.<sup>3</sup>

Since employment and disability are not mutually exclusive, researchers interested in studying the impact of disability on labor market behaviors have relied largely on the broader measure of “some limitation” in work capacity. In some contexts, however, the more restrictive “inability to work” definition may be of more interest. For example, the SSA requires recipients of federal disability insurance benefits to demonstrate “the inability 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 which has lasted or can be expected to last for a continuous period of at least 12 months.”<sup>4</sup> In 2006, substantial gainful activity is defined as earnings exceeding \$860 per month (\$500 per month during the time of the HRS survey). We estimate bounds on the disability rates for both the broad and restrictive measure of disability.

To draw inferences on the true disability rate, we combine these self-reports of work limitation with other information that can potentially shed light on their reliability. Table I displays means and standard deviations for selected variables used in our analysis. As expected, labor market and disability insurance status vary substantially with reported disability status. For example, the employment rate among those reporting no work limitation is 78.4%, nearly 2.6 times higher than the 29.6% employment rate among those reporting some limitation. Likewise, just over half the respondents reporting work limitations and nearly four-fifths reporting being unable to work altogether had applied for federal disability insurance (DI) benefits from the SSA, whereas very few respondents reporting no work limitation had applied for benefits.<sup>5</sup>

Although work disability is not synonymous with general health status, there is undoubtedly a close relationship. Our analysis exploits a wealth of information on a respondent’s reported physical and mental health to aid in implementation of the verification and MIV assumptions. In Table I, we display the means of selected health-related variables by reported disability status. At the most

---

predictor of disability insurance applications and awards.

<sup>3</sup>This broad conceptualization of work limiting disability has been espoused by both the World Health Organization and the framers of the American with Disabilities Act (ADA). See the Institute of Medicine (2002, Chapter 2) for further discussion of the conceptual issues in defining disability.

<sup>4</sup>U.S. Social Security Administration, 2006. *Compilation of the Social Security Laws Including the Social Security Act, as Amended, and Related Enactments, Social Security Act*, 42 U.S.C. § 423(d), Section 223(d).

<sup>5</sup>The federal government provides cash and medical benefits to the disabled through the Social Security Disability Insurance (SSDI) and Supplemental Security Income (SSI) programs. The formal medical eligibility criteria for the two programs are identical. Because the HRS does not distinguish between SSDI and SSI applications, we refer to SSDI/SSI jointly as the SSA’s Disability Insurance (DI) program.

basic level, we exploit information from two generic questions about a respondent’s physical and mental health status. For example, we see that 42.3% of respondents reporting a work limitation claim to be in fair or poor general physical health, compared with only 12.4% among those reporting no work limitation.

Another series of health-related questions provides information on other more objective measures of limitation. To proxy for disability status, some researches have relied on indirect summary measures like body mass (e.g., Gruber and Kubik, 1997) or subsequent mortality (e.g., Parsons, 1980). In our sample, 3.0% of respondents died before the second interview, and 58.9% have a body mass outside the ideal range defined by Fahey et al. (1997). These measures of limitation are clearly associated with self-reports of disability. The mortality rate, for example, is more than four times higher among those reporting a work limitation (7.5% compared with 1.8%) and more than five times higher among those reporting an inability to work at all (11% compared with 2.1%).

Beyond these indirect proxy measures of limitation, the HRS includes a battery of direct questions related to a respondent’s ability to perform basic functions. Activities of daily living indicators (ADLs) are intended to measure the ability to undertake basic self-care functions such as eating or dressing without help. Instrumental activities of daily living (IADLs) are intended to measure capabilities relevant to independent living, such as the ability to travel beyond walking distance. Such limitations do not directly measure work disability, but they may often contribute to difficulties in performing job-related tasks. Using definitions suggested by Loporest et al. (1995, p. S297), we aggregate this information into an index of functional limitations ranging from *Level 0*: no functional limitation, to *Level 6*: cannot do one of the basic life functions (see Table III for details). As expected, respondents reporting work limitations are more inclined to report functional limitations. Still, these measures seem to reflect different aspects of reported impairment. For example, 13.9% of those reporting a work limitation and 4.7% of those reporting the inability to work altogether do not report difficulty with any of the activities.

Finally, we exploit self-reported information on the presence of specific clinical health conditions. Following Wallace and Herzog (1995), we focus on a subset of seventeen reported conditions in the HRS that are expected to be the most prevalent among middle-aged and elderly persons and/or most likely to result in work disability. Across all of the 17 conditions we consider (see Table I, notes d and e), respondents who report a work limitation report an average of 4.02 conditions compared with 1.58 conditions among those reporting no work limitation.



### 3. CLASSIFICATION ERROR MODEL

While the work limitation questions are notably ambiguous, survey designers clearly have an expectation that respondents will be able to place these questions in a reasonable social context. When a survey asks whether a respondent is “unable to work altogether,” for example, it is understood that the respondent might reasonably answer “yes” even though hypothetically it might be possible to perform some small amount of work. The threshold for answering in the affirmative depends on current social norms for what constitutes an inability to work (see, e.g., Kapteyn et al., forthcoming).

The problem is that some respondents might use a different threshold for assessing disability. While it seems unlikely that a significant number of survey respondents are prone to willfully misrepresent their work capacity, a much greater concern in the literature revolves around the possibility that social or psychological factors can lead to self-rationalization (see, e.g., Bound and Burkhauser, 1999). Concerns over systematic misreporting are generally based on two observations, one financial and one social. First, eligibility for government disability assistance programs is tied to both earnings and disability status. Second, some people may feel social pressure to be working until normal retirement age. Thus, short of intentionally misreporting, some nonworkers or disability insurance applicants might have a different threshold for equating a health condition with a work limitation. To help rationalize a nonemployment spell, for example, nonworkers might be more prone than workers to interpret a particular medical problem (e.g., a bad back of a given severity) as a work limitation. At the same time, other respondents might not wish to admit that they are having difficulty coping with a health condition, so they might claim to be able-bodied despite having a substantial work limitation.

Formally, let  $X = 1$  if the respondent reports a disability and 0 otherwise. Let  $W = 1$  indicate that the individual is truly disabled relative to the intent of the survey question, with  $W = 0$  otherwise. Finally, let  $Z$  indicate whether a respondent provides accurate information, with  $Z = 1$  if  $W = X$  and  $Z = 0$  otherwise. We are interested in making inferences on the unobserved true disability rate,  $P(W = 1)$ .

Some fraction,  $P(X = 1, Z = 0)$ , inaccurately report being disabled (false positives) while others,  $P(X = 0, Z = 0)$ , inaccurately report being nondisabled (false negatives). Thus, the true and reported disability rates are related as follows:

$$P(W = 1) = P(X = 1) + P(X = 0, Z = 0) - P(X = 1, Z = 0). \quad (1)$$

The observed disability rate equals the true disability rate if the fraction of false negative reports exactly offsets the fraction of false positive reports. The data, however, only identify the fraction of the population that self-reports disability,  $P(X = 1)$ . The sampling process cannot identify the fraction of false negative or false positive reports.

As a starting point, it is useful to evaluate what can be inferred about the disability rate  $P(W = 1)$  given prior information on the fraction of respondents who provide valid self-reports. In particular, suppose

$$P(Z = 1) \geq v \tag{2}$$

where  $v$  is a known lower bound on the accurate reporting rate. By varying the value of  $v$ , we can consider the wide range of views characterizing the debate on inaccurate reporting. Those willing to assume fully accurate reporting can set  $v = 1$ , in which case the sampling process identifies the disability rate. Those uncomfortable with placing any lower bound on the fraction of accurate responses (e.g., Myers, 1982; Bowe, 1993) can set  $v = 0$ , in which case the sampling process is uninformative. Middle ground positions are evaluated by setting  $v$  somewhere between 0 and 1.

Given the restriction that no more than some fraction,  $1 - v$ , of the population misreports disability status, we know from (1) that

$$\max\{P(X = 1) - (1 - v), 0\} \leq P(W = 1) \leq \min\{P(X = 1) + (1 - v), 1\}. \tag{3}$$

These bounds are derived by Horowitz and Manski (1995, Proposition, Corollary 1.2). Henceforth, we will refer to these bounds as the HM bounds. Intuitively, the bounds narrow as the upper bound misreporting rate,  $1 - v$ , declines.

In the HRS sample, 20.8% of respondents report some work limitation. The bounds in (3) reveal that this self-reported disability measure provides only modest information about the true disability rate unless  $v$  is large. In fact, the HM bounds remain completely uninformative unless it is known that the accurate reporting rate exceeds 20.8%; the lower bound is zero unless it is known that at least 79.2% of responses are accurate.

### 3.1. PARTIAL VERIFICATION OF OBSERVED SUBGROUPS

Short of assuming fully accurate reporting, a number of researchers combine distributional restrictions with assumptions of fully accurate disability self-reports within particular groups of respondents. Kreider (1999) and McGarry (2004), for example, explicitly assume that workers provide fully accurate responses, remaining agnostic about the self-reports from nonworkers. In

the spirit of these ideas, we evaluate what can be inferred about the true disability rate when prior information is brought to bear on the degree of misreporting within certain observed subgroups. For now, we focus on basic notation. Our specific verification strategies are presented in Section 4.

To formalize the notion of partial verification, let  $Y = 1$  indicate that a respondent belongs to a verified subgroup, with  $Y = 0$  otherwise. Using the law of total probability, we can decompose the true disability rate by subgroups:

$$P(W = 1) = P(W = 1|Y = 1)P(Y = 1) + P(W = 1|Y = 0)P(Y = 0). \quad (4)$$

Although respondents in the verified subgroups might have few incentives to misreport, there may remain random errors: respondents may make mistakes in assessing the disability threshold, valid reports can be miscoded, and so forth. We allow for the possibility of exogenous response errors within the verified group such that there can be *partial verification*.<sup>6</sup> Formally, let  $v_y$  be the known lower bound fraction of accurate reporters in the verified subgroup and assume that at least half of the verified group reports accurately:  $P(Z = 1|Y = 1) \geq v_y \geq \frac{1}{2}$ .<sup>7</sup> Let the reporting errors in the verified group be random so that  $P(W = 1|Y = 1) = P(W = 1|Y = 1, Z)$ . No prior information is assumed about the validity of self-reports from the unverified cases. Then the following bounds apply (see the appendix for a proof):<sup>8</sup>

**Proposition 1:** *If  $P(Z = 1|Y = 1) \geq v_y \geq \frac{1}{2}$  for a known  $v_y$  and  $P(W = 1|Z, Y = 1) = P(W = 1|Y = 1)$ , it follows that*

$$\eta_0 \left[ \theta \frac{v_y - P(X = 0|Y = 1)}{2v_y - 1} + (1 - \theta)P(X = 1|Y = 1) \right] P(Y = 1) \leq P(W = 1) \leq \quad (5)$$

$$\left( \eta_1 \left[ \theta P(X = 1|Y = 1) + (1 - \theta) \frac{v_y - P(X = 0|Y = 1)}{2v_y - 1} \right] + (1 - \eta_1) \right) P(Y = 1) + P(Y = 0).$$

where  $\eta_i = 1 [v_y > P(X = i|Y = 1)]$ ,  $i = 0, 1$  and  $\theta = 1 [P(X = 0|Y = 1) > P(X = 1|Y = 1)]$ .

By varying the value of  $v_y$ , we can assess the sensitivity of the bounds to the strength of the verification assumption. In the special case that all respondents in verified groups are known to

<sup>6</sup>Dominitz and Sherman (2004) consider the case of full verification for certain subpopulations.

<sup>7</sup>The assumption that at least half the reports are accurate is applied by Bollinger (1996) and others and seems consistent with the notion of verification.

<sup>8</sup>Molinari (2005) independently derives a similar result using a different approach. She shows that the relationship between the distribution of a true variable and its mismeasured counterpart can be represented by a system of equations involving a coefficient matrix of misclassification probabilities. Restrictions on this matrix can be used to partially identify regions for the true variable. Kreider (2005) provides graphical motivation for these bounds by appealing to restrictions on the patterns of false positives and false negatives.

provide accurate reports ( $v_y = 1$ ) then

$$P(X = 1, Y = 1) \leq P(W = 1) \leq P(X = 1, Y = 1) + P(Y = 0). \quad (6)$$

In this informational setting, the true disability rate is intuitively no less than the reported rate among verified cases and no greater than this rate plus the fraction of unverified cases.

### 3.2. MONOTONICITY ASSUMPTIONS

The bounds in Equation (5) can be further narrowed when combined with monotonicity assumptions linking disability and observed covariates. Consider, for example, age and disability. The incidence of many debilitating health conditions rises with age, and many health conditions are persistent once developed. The resulting tendency for individuals to accumulate health problems over time suggests that the population disability rate is nondecreasing in age.

To formalize the age monotonicity assumption, let  $u$  measure the age of the respondent and let  $LB(u)$  and  $UB(u)$  be the known lower and upper bounds, respectively, given the available information on the true disability rate,  $P(W = 1|u)$ . Age is a monotone instrumental variable (MIV) if the true disability rate weakly increases with  $u$ . Under this restriction, Manski and Pepper (2000, Proposition 1 and Corollary 1) show that

$$\sup_{u_0 \geq u_1} LB(u_1) \leq P(W = 1|u = u_0) \leq \inf_{u_0 \leq u_2} UB(u_2). \quad (7)$$

There are no other restrictions implied by the MIV assumption. The MIV bound on the unconditional disability rate,  $P(W = 1)$ , is easily obtained using the law of total probability (Manski and Pepper, 2000).

### 3.3. ESTIMATION

The Proposition 1 bounds are functions of various nonparametrically estimable probabilities and thus can be consistently estimated by “plugging-in” the sample analogs. Estimation of the MIV bounds, however, is complicated by the fact that the monotonicity restrictions in Equation (7) must be imposed over collections of various estimates. For example, in our application, we divide the sample into 39 age groups containing 252 respondents per group (251 in four of the groups). In finite samples, plug-in estimators that take supremums and infimums are systematically biased. In our application, the estimated bounds will be too narrow because the lower bound estimate is upward-biased and the upper bound estimate is downward-biased.

To measure and correct for this bias, we employ a modified estimator that uses a nonparametric bootstrap bias correction. The basic idea is straightforward. Let  $T_n$  be a consistent analog estimator of some unknown parameter  $\theta$  such that the bias of this estimator is  $b_n = E(T_n) - \theta$ . Using the bootstrap distribution of  $T_n$ , one can estimate this bias as  $\hat{b} = E^*(T_n) - T_n$  where  $E^*(\cdot)$  is the expectation operator with respect to the bootstrap distribution.<sup>9</sup> A bootstrap bias-corrected estimator then follows as  $T_n^c = T_n - \hat{b} = 2T_n - E^*(T_n)$ . See Efron and Tibshirani (1993) for a general description of the bootstrap bias correction and Kreider and Pepper (forthcoming) for further details.

The bootstrap is also used to provide a tractable way to form confidence intervals for our estimated bounds on the disability rate. To do this, we first apply the percentile-bootstrap method (bias-corrected) to derive 90% confidence intervals for the upper and lower bounds (see Efron and Tibshirani, 1993). The interval on the lower bound, for example, is defined by the 0.05 and 0.95 quantiles of the bootstrap distribution of the estimated bound. A Bonferroni joint confidence interval with a level of at least 90% is then derived by taking by the 0.05 quantile from the bootstrap distribution of the lower bound estimator and the 0.95 quantile of the distribution of the upper bound estimator.

#### 4. SPECIFIC STRATEGIES AND RESULTS

In this section, we provide details about our specific verification strategies and present empirical results. Throughout, we report estimated HM bounds, verification bounds, and MIV bounds. We begin by considering the problem of drawing inferences on the broader definition of disability involving some limitation in the kind or amount of work that can be performed. In Section 4.1, we bound the true disability rate under two different sets of verification assumptions. In Section 4.2, we assess identification decay when requiring that the subjective disability response must be corroborated by more objective measures of functional limitation. For example, one could decide not to verify work limitation among respondents who report no functional limitation. Finally, in Section 4.3 we consider drawing inferences on work incapacity, the more restrictive definition of work disability. Focusing on the subpopulation of disability insurance applicants, we assess whether the data provide any evidence of bias in the SSA award decision.

---

<sup>9</sup>For our application, we randomly draw with replacement from the empirical distribution 10,000 independent pseudo-samples of the original data.

## 4.1 VERIFICATION STRATEGIES

Many researchers have argued that the propensity to provide inaccurate reports of work limitation may be linked to particular observed groups of respondents. Researchers have argued that the extent of response errors is likely to vary by employment status (e.g., Stern, 1989; Kreider, 1999; McGarry, 2004), applications to and participation in government disability insurance programs (Bound and Burkhauser, 1999; Kreider, 2000), reported disability status (Institute of Medicine, 2002), and other observed covariates. Following this theme, we evaluate what can be learned about the true disability rate when certain observed groups are assumed to provide accurate responses, or at least to provide some lower bound degree of accurate reporting.

When formulating verification strategies, we borrow from the basic ideas contained in the existing literature but, at the same time, thoroughly examine the extensive health and labor market information available in the HRS. We aim to verify self-reported disability status for cases that appear to be the most credible and to not verify cases that involve some type of ambiguity or inconsistency. For example, previous studies have verified the self-reports of workers under the premise that workers face few incentives to misreport. But of the 6503 respondents reporting to be gainfully employed, 733 report elsewhere in the survey either zero hours, zero earnings, or being nonemployed. Given these labor market inconsistencies, we do not verify the work limitation responses of such individuals based on employment status alone (they might be verified based on other information). Likewise, we do not verify the responses of the 58 individuals who claimed to be able-bodied in one part of the survey but disabled or receiving disability benefits in another part of the survey.

Given the inherent uncertainty about which responses should be verified, we present two different models of partial verification. Model I involves relatively strong verification assumptions, some of which are relaxed in Model II. We begin with the broad measure of disability involving some work limitation. For Model I, we treat disability status reports of  $X$  as verified (with discussion below) for:

1. those currently working for pay (HRS variable V2717=1) except those who (a) report that they receive disability benefits from any program, (b) did not check the “working” box in question F1a (variable V2701) for current employment status, or (c) do not report positive labor hours and positive earnings (i.e., either value is zero or missing);
2. those reporting no work limitation ( $X = 0$ ), except those who (a) report receiving disability benefits or (b) checked “disabled” in box F1d (variable V2701) for current employment status;

3. those reporting a work limitation ( $X = 1$ ) if they also report being unable to work altogether due to one of the six serious health conditions identified by Wallace and Herzog (1995, p.S90);
4. disability beneficiaries (reporting  $X = 1$ ), except those who report that they are (a) currently working or (b) able to work.<sup>10</sup>

In Model I, 91.9% of the sample is verified. Borrowing from the existing literature, we verify the responses of most workers and of most respondents reporting to be able-bodied. In both cases, there appear to be few economic or psychological factors that would lead to misreporting. However, in each case exceptions are made for potentially conflicting information. The responses of workers who receive disability benefits are not verified, nor does employment status confer verification if there exists contradictory information on labor hours or earnings. Similarly, we verify  $X = 0$  cases except in the face of contradictory evidence that the respondent is receiving disability benefits or reports being disabled earlier in the survey. We verify the presence of at least some work limitation,  $X = 1$ , if the respondent reports complete work incapacity caused by a health condition that is known to often be debilitating and associated with relatively few false positive diagnoses.

Verification of disability beneficiaries is a more subtle matter. Our maintained assumption is that, in the absence of labor force participation, the receipt of disability benefits among those claiming to be unable to work at all corroborates the existence of at least some work limitation. Many have raised concerns that beneficiaries are inclined to exaggerate the extent of their limitations and that disability awards are prone to classification errors. Verifying some work limitation among this subset of beneficiaries, however, does not imply that the awards process is without error or that beneficiaries do not exaggerate the extent of disability; it only requires that adjudication errors are not so extreme that beneficiaries who report complete work incapacity are not work-limited at all.

Model II relaxes some of these assumptions. In particular, in Model II responses are “unverified” as follows: (1) proxy responses are never verified (5% of our sample), (2)  $X = 0$  cases are not verified if the respondent (a) reports pain of at least moderate severity, at its worst, that makes activities difficult or (b) has one of the six serious medical conditions and reports being limited in housework or other activities besides paid work, and (3)  $X = 1$  cases are no longer verified based on reporting one of the six serious health conditions. Under these more conservative assumptions, at least 78.4% of the sample is known to provide accurate responses.

---

<sup>10</sup>For this purpose, beneficiaries include all respondents who reported receiving disability benefits from any public or private program. Respondents were queried about the receipt of disability benefits from a variety of programs (e.g., SSDI, SSI, Veterans’ Disability, “State disability program,” Employer/union plan).

### 4.1.A VERIFICATION BOUNDS

Table II presents the estimated bounds for the true disability rate and their conservative 90% confidence intervals. Column A provides results under the corrupt sampling assumption alone. Under both Models I and II, the bounds reflect much uncertainty about the true disability rate. If at least 91.9% of respondents are known to provide accurate reports, for example, the HM bounds constrain the true disability rate to lie within  $[0.127, 0.288]$ . Without additional information about the reporting process, the disability rate may lie anywhere within this 16 point range.

In this setting, one function of the bounds is to test the validity of alternative measures of disability and models of the reporting error process. If the verification assumptions are correct, estimates lying significantly outside the bounds cannot be valid measures of true disability. Table I contains various possible measures of work limiting disability. Most notably, the self-reported disability rate of 20.8% lies within the 16 point range and thus cannot be rejected as being an accurate measure of true disability. Neither, however, can we reject the possibility that the fraction of respondents reporting to be in fair or poor physical health (18.6%), the fraction reporting to be in poor mental health (22.7%), or the fraction reporting to have one of the six serious medical conditions (27.3%), are valid measures of work disability. In contrast, the incidence of non-ideal body mass (58.9%) and the subsequent mortality rate (3.0%), lie far outside of the estimated bounds. Thus, given the assumption that at least 91.9% of respondents provide accurate self-reports, we find that these alternative measures do not reveal the incidence of work disabilities. Still, in a regression framework, these measures might serve as important control variables for health and limitation, and perhaps as valid instrumental variables for the true disability rate.

Column B displays estimated bounds under the assumption that all verified respondents provide accurate self-reports of disability. The verification assumption provides substantial identifying power, but the specifics are quite sensitive to the underlying model. Under Model I, the bounds narrow to the seven point range of  $[0.135, 0.215]$ , a 50% reduction in the width of the bounds. In this case, the fraction reporting to be in poor mental health and the fraction reporting a serious medical condition lie outside of the bounds, and thus cannot be valid measures of work limiting disability if the model is correct. However, the width of the verification bounds increases by three-fold when we move from Model I, where 91.9% of respondents are verified, to Model II, where 78.4% of respondents are verified. These changes in the underlying assumptions about the nature and extent of reporting errors generate large changes in the uncertainty about the disability rate.

Although the verification bounds can be substantially more informative than the HM bounds,



they provide only limited information on the true disability rate unless a large fraction of the sample is verified to provide completely accurate information. When we relax the parametric restrictions applied in much of the literature and isolate the identifying power of the verification assumptions, there remains much uncertainty about the true disability rate. These results support concerns raised by Benítez-Silva et al. (2004) that conclusions about reporting errors based on latent variable models (e.g., Kreider, 1999; McGarry, 2004) could be driven largely by parametric assumptions. Even in light of this uncertainty, however, we find that some alternative disability measures appear to capture different dimensions of health or limitation.

#### 4.1.B MIV BOUNDS

Uncertainty about the disability rate can be reduced at the cost of imposing stronger assumptions. In this section, we combine verification assumptions with MIV restrictions and illustrate how inferences vary across the different assumptions. First, we combine the assumption that true disability weakly increases with age, as discussed above, with the restriction that the disability rate is no higher among the employed than among the nonemployed:  $P(W = 1|L = 0) \geq P(W = 1|L = 1)$ , where  $L$  indicates whether a respondent participates in the labor market. Second, we combine this employment monotonicity assumption with an assumption that fitted values from an ordered probit model of federal disability applications comprise an MIV. In particular, a natural MIV can be constructed as the outcome of a respondent’s Disability Insurance application decision. Let this variable equal 0 if the respondent has not applied for disability benefits, 1 if a disability application was rejected, 2 if an application was accepted after appeal, and 3 if an application was accepted initially. Using this variable, we constructed an MIV as the fitted values from an ordered probit model that exploits information from attributes expected to influence work disability. The specification includes indicators for each of the 17 health conditions identified by Wallace and Herzog (1995), indicators for the functional limitation index (Levels 1-6), the indicator for subsequent mortality (died before wave 2), the indicator for ideal body mass, the indicator for being often bothered by pain, along with age, education, race, gender, marital status, veteran status, and asset level (details from this regression are available upon request).

The MIV estimates are reported in Columns C and D of Table II. These MIV assumptions have substantial identifying power. Under the age-employment MIV assumption, the Model I bounds on the work limitation rate, for example, collapse to the three point range [0.178, 0.204], while the DI-employment MIV shrinks the bounds to [0.149, 0.193]. In these cases, the self-reported disability rate, 0.208, lies outside of these bounds for the true disability rate and just on the edge of the upper

bound of the conservative 90% confidence interval. Thus, if the MIV assumptions are valid, these estimates provide some evidence of misreporting. In particular, since the unverified group consists primarily of nonworkers who claim to be disabled, we find some support for suggestions in the literature that members of this group systematically over-report disability.<sup>11</sup>

As before, however, the identification bounds decay rapidly as we relax the verification restrictions. The width of the age-employment MIV bound, for example, increases from the three point range in Model I to a nearly 16 point range,  $[0.129, 0.285]$ , in Model II. Thus, under the Model II verification assumptions, there is much uncertainty about the true disability rate. In this case, the self-reported disability rate of 20.8% lies within the estimated bounds, but so too does the fraction of respondents reporting to be in fair or poor physical or emotional health (18.6% or 22.7%) and the fraction reporting to have one of the six serious medical conditions (27.3%). The estimated bounds are quite sensitive to the underlying assumptions; we generally cannot reject the possibility that self-reports are unbiased.

## 4.2 FUNCTIONAL LIMITATIONS

We now investigate the sensitivity of the estimated bounds to assumptions linking work disability to functional limitation. Measures of physical limitation in the HRS might corroborate verification assumptions on self-reported disability. As noted in Section 2, disparities between these health-related measures do not imply that either measure is invalid. Still, inconsistencies might argue against verification. For example, respondents with severe functional limitations who report being able-bodied might not be verified as providing accurate reports of disability.

To study the sensitivity of the estimated bounds, we trace out the implications of a corroboration strategy that uses self-reports of functional limitation to weaken the verification assumptions. If a researcher believes that responses to questions about functional limitation provide no further evidence about work disability, then the results presented in Table II apply. Otherwise, apparent inconsistencies between reports of functional limitation and work limitation serve to caution against verification.

Table III displays the estimated bounds for verification Models I and II under the age-employment MIV assumption. The lower bound decreases the case where respondents are not verified if they

---

<sup>11</sup>Bound and Waidmann (2002), for example, observe that the fraction of working-aged men who identify themselves as work-limited nonworkers closely tracks the fraction receiving SSDI benefits (rising in the 1970s, falling in the 1980s, and rising again in the 1990s). Finding no parallel changes in reported disability among respondents older than 65 (who would be ineligible for SSDI benefits), they argue that exogenous changes in the availability of disability benefits appear to induce changes in disability reporting behavior.

report being disabled with a sufficiently low functional limitation index value (i.e., a value less than or equal to the particular level of  $\theta$ ), and the upper bound increases when respondents are not verified if they report being able-bodied with a sufficiently high functional limitation value (i.e., a value greater than or equal to the specified level of  $\theta$ ). When we do not verify respondents claiming to be disabled but also claiming to have no functional limitations ( $\theta = 6$ ), for example, the lower bound decreases from 0.178 to 0.166. The lower bound falls further to 0.114 when respondents with some difficulty with work functions ( $\theta = 3$ ) are not verified. When we do not verify respondents claiming to be able-bodied yet also having a severe functional limitation ( $\theta = 6$ ), the upper bound barely increases from 0.204 to 0.205. However, when we do not verify cases involving at least some difficulty with basic work functions ( $\theta = 3$ ), the upper bound increases to 0.398.

Inferences are clearly sensitive to how one models and assesses the relationship between reports of functional and work limitation. Identification decays rapidly if disparities between these two measures are taken to cast doubt on the validity of the self-reports of work limitation. Stated differently, to the extent that self-reported limitation responses are believed to be mostly reliable, we provide evidence that indicators of work limitation and functional limitation are measuring very different aspects of impairment.

### 4.3 WORK INCAPACITY AND THE SSA AWARD PROCESS

Thus far, we have focused on the problem of inferring the prevalence of impairment that limits the kind or amount of work that can be undertaken. While this conceptualization is widely utilized in research applications, the more restrictive definition of work limitation is of interest in some settings. For example, Benítez-Silva et al. (2004) take the SSA’s definition of disability – “the inability to engage in any substantial gainful activity” – (see Section 2) as the basis for the social standard for what constitutes work incapacity. In our HRS sample, 10.1% of respondents report they are unable to work altogether. In this section, we first use the methods developed above to place bounds on the true fraction of respondents nearing retirement age who are incapable of work. We then turn our attention to the subsample of DI applicants to assess whether SSA award outcomes are consistent with this conceptualization of disability.

Our verification assumptions for the “unable to work altogether” case are similar to those described in Model I for “some work limitation,” with several notable differences. Given the restrictive nature of this disability conceptualization, we impose stronger standards for verifying disability and impose weaker standards for verifying nondisability. A self-report of work incapacity ( $X = 1$ ) is verified if the respondent is receiving disability benefits and reports one of the six aforementioned

diagnosed serious conditions. The self-reported ability to work is verified unless the respondent reports being nonemployed, having some work limitation, and receiving disability benefits. Under these assumptions, self-reports of work capacity are verified for 93.5% of the sample.

Table IV presents the base results. As before, we find that the verification and MIV restrictions confer substantial identifying power. The HM bounds confine the disability rate to the 13 point range  $[0.036, 0.166]$ , whereas the verification bounds lie within the 6 point range  $[0.044, 0.109]$ . These bounds shrink further to the four point range  $[0.049, 0.089]$  under the age-employment MIV assumption. As before, the MIV bounds do not contain the self-reported rate of 10.1%, a finding that is robust to some departure from full verification. These verification bounds, however, decay further after requiring some corroborating information about functional limitation. When using function limitation Level 3 as our corroboration cutoff (some difficulty with physical or sedentary work functions), the lower bound falls from 0.049 to 0.040 and the upper bound increases from 0.089 to 0.362.

The restrictive definition of work disability is particularly germane for the subpopulation of DI applicants who, to be awarded benefits, must demonstrate the inability to engage in substantial gainful activity. By focusing on this group of respondents, our bounding approach can supplement insights into the validity of the DI award process. To obtain disability benefits, applicants provide detailed medical, income, and asset information to a federal SSA office. Eligibility is strict, and many applicants are denied benefits on the grounds that they do not meet the medical severity criteria. The accuracy of this process has been the subject of both political and academic debate.

Using HRS data on DI applications, awards, and receipt, we compare the fraction of beneficiaries to the estimated bounds on the true prevalence of work incapacity. Bounds on the true rate of work incapacity among this subgroup may provide evidence about the accuracy of SSA award decisions. If the award process accurately determines the rate at which applicants are unable to engage in gainful activity, then the fraction of beneficiaries should lie within the estimated bounds on the true disability rate. If the fraction of beneficiaries instead lies outside of the bounds, then we can reject the joint hypothesis that the SSA award process is accurate and forms the basis for the social definition of work incapacity. Like Benítez-Silva et al.’s (2004) test of Rational Unbiased Reporting, we test for accurate award decisions on average, not for a particular individual.

Among the 9824 age-eligible respondents in the HRS, 1082 had applied for DI benefits prior to the first interview. The ultimate award decision, which can take a few months to a few years to be resolved, is discerned using information from the first two waves of the HRS. For success-

ful applicants, we also document whether the respondent was receiving (or scheduled to receive) benefits during the Wave 1 interview. This allows us to compare self-reported disability status with concurrent determination of DI eligibility. Of the 1082 disability applicants, 452 were initially awarded benefits and 617 were initially denied. The award decision was not available in Wave 1 for an additional 13 cases, but Wave 2 information indicates that only one of these applications was ultimately successful.<sup>12</sup> Of the 617 initially denied cases, 430 continued through the appeals process, and 263 of these appeals were successful.<sup>13</sup> Of those awarded benefits, 75 recipients were no longer participating in the program by Wave 1 of the survey. Therefore, we find that 641 respondents (59.2% of the applicant pool) were receiving or scheduled to receive benefits at the time the questions about work limitation were asked.

Since the HRS collects disability status information at discrete times that do not necessarily coincide with the time of the application and award decisions for DI benefits, an important issue is the relevant window of observation. We compare data on self-reports and SSA award decisions for two different time windows. First, we consider the subgroup of all (age-eligible) HRS respondents who applied to receive DI benefits regardless of when the application was filed. Second, we focus on the much smaller subsample of 233 applicants whose most recent SSA adjudication date lies within six months of the Wave 1 interview date. For both observation windows, self-reported work incapacity and DI beneficiary status generally concur, but this is not always true. Nearly 33% of respondents in the larger sample and 43% of respondents in the smaller sample provide self-reports that differ from the DI outcome. A relatively small number of respondents report that they can work despite receiving benefits. A larger number report that they cannot work and are not receiving DI benefits. Thus, a notably larger fraction of applicants classify themselves as being unable to work – 73% in the full sample and 79% using the shorter horizon – than report the current receipt of disability benefits – 59% and 47%, respectively.<sup>14</sup>

---

<sup>12</sup>Of the remaining 12 cases, five respondents indicated in Wave 2 that they had been denied benefits. We classified the other seven cases as denied as well: none reported receiving benefits in either wave, and we found no indication of pending decisions.

<sup>13</sup>By Wave 1 of the survey, 259 appeals were successful and 162 were not successful. For the remaining nine cases, we used Wave 2 information to classify four applications as ultimately successful and the rest unsuccessful. The decision whether to appeal was unavailable for three applicants; we classified each case as ultimately rejected based on evidence from Wave 2.

<sup>14</sup>Benítez-Silva et al. (2004) find the marginal distribution of the ultimate DI award outcome to be very similar to the marginal distribution of self-classified work incapacity status. There are several notable differences between the sampling frames and assignment rules that may explain these differences. First, whereas we focus on respondents in the first wave of the survey, Benítez-Silva et al. use the first three waves. Second, Benítez-Silva et al. do not restrict the sample to age-eligible respondents nearing retirement age. Third, Benítez-Silva et al. define an observation window that restricts attention to individuals who applied for DI benefits within a one-year window surrounding the interview date (6 months before and after). Finally, whereas we classify outcomes based on the current receipt of DI benefits, Benítez-Silva et al. classify outcomes based on whether the applicant was approved to receive benefits.

Table V presents the bounds on the true work incapacity rate for both observation windows. Verification bounds are provided in both cases, and the age-employment MIV bounds are provided for the larger subpopulation of all age-eligible HRS applicants. The MIV estimates are unreliable when using the smaller sample of 233 respondents. In all cases, the estimated bounds are rather wide, and in all cases the bounds include the DI beneficiary rate. Consider the tightest bounds found under the MIV assumptions. For the subsample of all age-eligible HRS applicants, we estimate the true work incapacity rate to lie within  $[0.505, 0.751]$ . Since the bounds overlap with the fraction of applicants that was deemed eligible for assistance (59%), we find no evidence of bias of the SSA award decision under the maintained assumptions. Thus, without additional information on the reporting process, we cannot reject the possibility that the true work incapacity rate equals the DI beneficiary rate of around 60%. Nor, however, can we reject the possibility the true rate equals the self-reported work incapacity rate of 73%.

## 5. CONCLUSION

While questions have been raised about the validity of many self-reported measures, surveys of disability have been especially controversial. Quantifying disability is conceptually difficult, and there is no commonly accepted gold standard for its measurement. Systematic response errors can arise if a person’s self-assessed disability status is influenced by economic or psychological factors. The nature and extent of these errors has been debated in the academic literature for more than two decades since Anderson and Burkhauser (1984) characterized disability measurement problems in micro survey datasets as “the major unsettled issue in the empirical literature on the labor supply of older workers.” Today, especially since the passage of the Americans with Disabilities Act in 1990, the use of these self-reports in guiding public policy has become a matter of growing public concern. The National Council on Disability (NCD, 2002), for example, argues that the use of self-reported disability information can lead to dangerous public policy decisions. The Council goes so far as to suggest that the federal government should not support the dissemination of self-reported work limitation data due to a lack of acceptable methods for assessing disabilities (see also Myers, 1982 and Bowe, 1993).

Yet despite these concerns, these self-reports seem to provide valuable information about work capacity beyond that captured in alternative measures of health. More “objective” measures may be less prone to classification error, yet they may also contain far less information about work capacity than responses to direct questions about a person’s ability to work. Respondents have an opportunity to place their health limitations in a useful social context. Some of our results suggest,

for example, that indicators of work limitation and functional limitation are measuring markedly different aspects of health status.

This paper provides a methodology for partially identifying work disability rates using self-reports of limitation. Our framework allows us to explore the identifying power of a range of different assumptions that bridge the gap between completely discarding the data (e.g., as suggested by the NCD) and taking all of the data at face value. Our results help reveal the nature and extent to which our knowledge about the prevalence of disability is limited by our lack of understanding of reporting errors. Under strong assumptions, we are able to nearly identify the disability rate. Identification deteriorates as the identifying assumptions are relaxed.

The patterns of identification decay are striking. Without strong prior information on the nature and degree of accurate reporting, the bounds can be frustratingly wide. Moreover, the bounds can be sensitive to relatively minor changes in the underlying classification error models. The results are especially sensitive to how one models potential inconsistencies between the subjective self-assessments of work limitation and more objective measures of functional limitation. In cases where we can only bound the parameter to lie within a wide range, we must either accept a large degree of uncertainty or be willing to impose stronger identifying assumptions.

The Institute of Medicine (2002) has called for more methodological research on these measurement issues. We hope that our nonparametric bounding framework can be used as a stepping stone for resolving the uncertainty about how best to measure work limitations and model disability, labor supply, and the receipt of public transfers.

## APPENDIX

### Proof of Proposition 1

To simplify notation, let the conditioning on the verified subgroup,  $Y = 1$ , be implicit. Then the law of total probability implies:

$$P(X = 1) = P(W = 1|Z = 1)P(Z = 1) + P(W = 0|Z = 0)P(Z = 0). \quad (8)$$

The independence assumption requires  $P(W = 0|Z = 1) = P(W = 0|Z = 0)$ . Substituting for  $P(W = 0|Z = 0)$  in (8) and using the fact that  $W$  is binary, it follows that  $P(X = 1) = P(W = 1|Z = 1)[2P(Z = 1) - 1] + [1 - P(Z = 1)]$ . Therefore,

$$P(W = 1) = P(W = 1|Z = 1) = \frac{P(X = 1) - 1 + P(Z = 1)}{2P(Z = 1) - 1} = \frac{P(Z = 1) - P(X = 0)}{2P(Z = 1) - 1}.$$

Although  $P(Z)$  is unknown, we know that  $v_y \leq P(Z = 1) \leq 1$ . Thus, we can bound the disability rate by assessing  $P(W = 1|Z = 1)$  across the possible values of  $P(Z = 1)$ . It follows that if  $v_y \leq P(X = 0)$ , the lower bound on the true disability rate is zero. Likewise, if  $v_y \leq P(X = 1)$ , the upper bound is one. Otherwise, differentiating this equation with respect to  $P(Z)$  reveals that if  $P(X = 0) > P(X = 1)$ , then  $P(W = 1)$  is increasing in  $P(Z = 1)$  for all conjectured values of  $P(Z = 1) > P(X = 0)$ . Otherwise, it is decreasing in  $P(Z = 1)$ .  $\square$



## References

- [1] Anderson, K. and Burkhauser, R. 1984. "The Importance of the Measure of Health in Empirical Estimates of the Labor Supply of Older Men." *Economics Letters*, **16**: 375-80.
- [2] Benítez-Silva, H., M. Buchinsky, H.M. Chan, J. Rust, and S. Sheidvasser. 1999. "An Empirical Analysis of the Social Security Disability Application, Appeal, and Award Process." *Labour Economics*, **6**: 147-178.
- [3] Benítez-Silva, H., M. Buchinsky, H.M. Chan, S. Cheidvasser, and J. Rust. 2004. "How Large is the Bias in Self-Reported Disability?" *Journal of Applied Econometrics*, **19**: 649-670.
- [4] Bollinger, C. 1996. "Bounding Mean Regressions When A Binary Variable is Mismeasured." *Journal of Econometrics*, **73**: 387-99.
- [5] Bound, J. 1991 "The Health and Earnings of Rejected Disability Insurance Applicants: Reply." *American Economic Review*, **81**: 1427-1434.
- [6] Bound, J. and Burkhauser, R. 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.
- [7] Bound, J., and T. Waidmann. 2002: "Accounting for Recent Declines in Employment Rates among Working-Aged Men and Women with Disabilities," *Journal of Human Resources*, **37**: 231-50.
- [8] Bowe, F. 1993. "Statistics, Politics, and Employment of People with Disabilities." *Journal of Disability Policy Studies*, **4**: 84-91.
- [9] Dominitz, J. and Sherman, R. 2004. "Sharp Bounds Under Contaminated or Corrupted Sampling With Verification, With an Application to Environmental Pollutant Data," *Journal of Agricultural, Biological & Environmental Statistics*, **9**: 319-338.
- [10] Dwyer, D. and Mitchell, O. 1999. "Health Problems as Determinants of Retirement: Are Self-Rated Measures Endogenous?" *Journal of Health Economics*, **18**: 173-93.
- [11] Efron, B. and R. J. Tibshirani. 1993. An Introduction to the Bootstrap. Chapman and Hall. New York: NY.
- [12] Fahey, T., Insel, P., and Roth, W. 1997. Fit and Well: Core Concepts and Labs in Physical Fitness and Wellness, 2nd Edition, Mayfield Publishing Company, Mountain View, CA: 137.
- [13] Gruber, J. and J. Kubik. 1997. "Disability Insurance Rejection Rates and the Labor Supply of Older Workers," *Journal of Public Economics*, **64**: 1-23.
- [14] Haveman, Robert, and Barbara Wolfe. 1984. "The Decline in Male Labor Force Participation: Comment." *Journal of Political Economy*, **92**: 532-41.
- [15] Horowitz, J. and Manski, C. 1995. "Identification and Robustness With Contaminated and Corrupted Data." *Econometrica*, **63**: 281-302.
- [16] Institute of Medicine. 2002. The Dynamics of Disability: Measuring and Monitoring Disability for Social Security Programs. Committee to Review the Social Security Administration's Disability Decision Process Research, G.S. Wunderlich, D.P. Rice and N.L. Amado, Editors, National Research Council, Washington, D.C., National Academy Press.

- [17] Kerkhofs, M. and Lindeboom, M. 1995. "Subjective Health Measures and State Dependent Reporting Errors." *Health Economics*, **4**: 221-35.
- [18] Kapteyn, A., J. Smith, and A. van Soest. forthcoming. "Vignettes and Self-Reports of Work Disability in the U.S. and the Netherlands," *American Economic Review*.
- [19] Kreider, B. 1999. "Latent Work Disability and Reporting Bias." *Journal of Human Resources*, **34**: 734-69.
- [20] \_\_\_\_\_. 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.
- [21] \_\_\_\_\_. 2005. "Partially Identifying the Prevalence of Health Insurance Given Contaminated Sampling Response Error," Working Paper, Iowa State University.
- [22] Kreider, B. and J. Pepper. Forthcoming. "Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors." *Journal of the American Statistical Association*.
- [23] Loporest, P., K. Rupp, and S.H. Sandell. 1995. "Gender, Disability, and Employment in the Health and Retirement Study," *Journal of Human Resources*, **30**: S293-S318.
- [24] Molinari, F. 2005. "Partial Identification of Probability Distributions with Misclassified Data." Cornell University, Working Paper.
- [25] Manski, C. and Pepper, J. 2000. "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica*, **68**: 997-1010
- [26] McGarry, K. 2004. "Health and Retirement: Do Changes in Health Affect Retirement Expectations?" *Journal of Human Resources*, **39**: 624-48
- [27] Myers, R. 1982. "Why Do People Retire from Work Early?" *Social Security Bulletin*. September: 10-14.
- [28] Nagi, S. 1965. "Some Conceptual Issues in Disability and Rehabilitation" in Sociology and Rehabilitation. Sussman M.B., ed. Washington, DC: American Sociological Association.
- [29] National Council on Disability. 2002. National Disability Policy: A Progress Report. Dec. 2000-Dec. 2001.
- [30] Parsons, D. 1980. "The Decline in Male Labor Force Participation," *Journal of Political Economy*, **88**: 117-134.
- [31] \_\_\_\_\_. 1984. "Disability Insurance and Male Labor Force Participation: A Response to Haveman and Wolfe." *Journal of Political Economy*, **92**: 542-549.
- [32] \_\_\_\_\_. 1991. "The Health and Earnings of Rejected Disability Insurance Applicants: Comment." *American Economic Review*, **81**: 1419-1426.
- [33] Stern, S. 1989. "Measuring the Effect of Disability on Labor Force Participation." *Journal of Human Resources*, **24**: 361-95.
- [34] Wallace, R. B. and A.R. Herzog. 1995. "Overview of the Health Measures in the Health and Retirement Study," *Journal of Human Resources*, **30**: S84-S107.

Table I. Selected Means and Standard Deviations

	Full Sample N=9824		Reported Work Limitation?		Reported Inability to Work Altogether?	
			Yes N=2039	No N=7785	Yes N=992	No N=8832
	Mean	Std. Dev.	Mean	Mean	Mean	Mean
Age	56.0	3.18	56.4	55.9	56.6	55.9
Female	53.2	0.499	53.8	53.0	52.6	53.2
Nonwhite	0.286	0.452	0.359	0.267	0.435	0.269
Currently working for pay	0.683	0.465	0.296*	0.784	0.000*	0.759
Ever applied for SSDI/SSI benefits	0.110	0.313	0.528*	0.001	0.792*	0.034
Currently receive SSDI/SSI benefits	0.065	0.246	0.311*	0.000	0.537*	0.012
Currently receive disability benefits from any program	0.073	0.260	0.351*	0.000	0.586*	0.016
<b>Health Status</b>						
Reported fair/poor general health status <sup>a</sup>	0.186	0.389	0.423*	0.124	0.535*	0.147
Reported fair/poor emotional health status <sup>a</sup>	0.227	0.419	0.651*	0.115	0.824*	0.160
Died prior to second wave	0.030	0.171	0.075*	0.018	0.111*	0.021
Body mass index out of ideal range <sup>b</sup>	0.589	0.492	0.622*	0.580	0.628*	0.584
<u>ADL/IADL functional limitation index (0-6)<sup>c</sup></u>						
Level 0: No functional limitation	0.540	0.498	0.139*	0.645	0.047*	0.595
Level 6: Very difficult/can't do one of the basic functions	0.022	0.145	0.092*	0.003	0.150*	0.007
Number of reported debilitating health conditions <sup>d,e</sup>	2.08	1.92	4.02*	1.58	4.63*	1.78
Reported a severe condition <sup>e</sup>	0.273	0.446	0.593*	0.190	0.712*	0.224

\*† Significant difference between the “yes” and “no” responses at the 1% and 5% levels, respectively

<sup>a</sup> Other categories include excellent, very good, and good health status.; <sup>b</sup> Ideal body mass is defined as 20-25 kg/m<sup>2</sup> following Fahey et al. (1997).

<sup>c</sup> Following Lobrest et al. (1995, p. S297), we construct four categories of functions: (I) *basic functions*, (II) *sedentary work functions*, (III) *physical work functions*, and (IV) *very physical work functions*. For each activity, a respondent can answer “not at all difficult,” “a little difficult,” “somewhat difficult,” “very difficult/can’t do,” or “don’t do.” The last two categories are grouped together. Respondents were told to exclude any limitation expected to last less than three months. The functional limitation index takes on values 0-6 as defined by Level 0 – Level 6 in the table.

<sup>d</sup> Defined by Wallace and Herzog (1995, pS89 and Table 1) to including conditions listed in (e) below and asthma, back, leg or feet problems, kidney or bladder problems, stomach or intestinal ulcers, high cholesterol, and fractures. We additionally include poor eyesight (with glasses) and poor hearing (with hearing aid).

<sup>e</sup> Includes diabetes, cancer, chronic lung disease, heart condition, stroke, or psychiatric condition as defined by Wallace and Herzog (1995, Table 1). We additionally include arthritis and hypertension.

Table II. Corrupt Sampling, Partial Verification, and MIV Bounds on  $P(W=1)$   
Work Limitation Case

(A) HM Corrupt Sampling Bounds*	(B) Proposition 1 Verification Bounds	(C) Age and Employment MIV Bounds		(D) Disability Application and Employment MIV Bounds	
<b>Verification Model I<sup>†</sup></b>					
		$v_y = 1$	$v_y = 0.9$	$v_y = 1$	$v_y = 0.9$
[0.127, 0.288] <sup>a</sup>	[0.135, 0.215]	[0.178, 0.204] <sup>c</sup>	[0.106, 0.204]	[0.149, 0.193]	[0.120, 0.193]
[0.121, 0.298] <sup>b</sup>	[0.129, 0.223]	[0.155, 0.207]	[0.086, 0.207]	[0.143, 0.209]	[0.112, 0.209]
		+0.010 -0.016 <sup>d</sup>	+0.016 -0.016	+0.009 -0.005	+0.006 -0.005
<b>Verification Model II<sup>‡</sup></b>					
		$v_y = 1$	$v_y = 0.9$	$v_y = 1$	$v_y = 0.9$
[0.000, 0.423]	[0.103, 0.318]	[0.129, 0.285]	[0.065, 0.285]	[0.110, 0.320]	[0.089, 0.320]
[0.000, 0.434]	[0.098, 0.326]	[0.115, 0.300]	[0.051, 0.300]	[0.104, 0.325]	[0.080, 0.325]
		+0.013 -0.013	+0.013 -0.013	+0.008 -0.008	+0.006 -0.008

<sup>†</sup>For Model I ( $v = 0.919$ ), work limitation status  $X$  (but not work incapacity status) is treated as verified for members of the following groups:

- (1) disability beneficiaries (reporting  $X=1$ ) unless currently working or report able to work
- (2) those currently working for pay ( $V2717=1$ ) unless (a) receiving disability benefits, (b) did not check the “working” box in question F1a ( $V2701$ ) for current employment status, (c) labor hours are zero/missing, or (d) earnings are zero/missing
- (3) those reporting no work limitation ( $X=0$ ) unless also report receiving disability benefits or checked “disabled” as current employment status
- (4) those reporting work limitation ( $X=1$ ) if report unable to work due to one of the six serious diagnosed conditions highlighted by Wallace and Herzog (1995): treated for cancer in the last 12 months, diabetic taking insulin, chronic lung disease that limits activities, congestive heart disease with treatment or shortness of breath, stroke with health consequences, or current psychiatric/emotional problem with medication or other treatment

<sup>‡</sup>Model II ( $v = 0.784$ ) differs from Model I in that: (1) proxy responses are never verified, (2)  $X=1$  cases are not verified based on specific medical conditions, and (3)  $X=0$  cases are never verified if the respondent (a) reports pain of at least moderate severity at its worst that makes activities difficult or (b) has a serious/objective medical condition defined in Model I and reports being limited in housework or other activities.

<sup>a</sup>point estimates of the population bounds

<sup>b</sup>bootstrapped 5<sup>th</sup> and 95<sup>th</sup> percentile bounds

<sup>c</sup>MIV point estimates, corrected for finite-sample bias

<sup>d</sup>estimated finite-sample bias

\*There are 22 missing values for reported work limitation  $X$ ; the estimated bounds conservatively take worst case scenarios for these missing values.

Table III. Sensitivity of Age and Employment MIV Bounds when Requiring Functional Limitation Corroboration  
Work Limitation Case

Lower Bound			Upper Bound				
	$X=1$ (reports work limitation) is never verified if ADL limitation index $\leq \theta$	Model I	Model II		$X=0$ (reports no work limitation) is never verified if ADL limitation index $\geq \theta$	Model I	Model II
$\theta=0$	no functional limitation	0.166 <sup>a</sup> 0.152 <sup>b</sup>	0.122 <sup>a</sup> 0.110 <sup>b</sup>	$\theta=6$ :	very difficult/can't do at least one basic function	0.205 <sup>a</sup> 0.210 <sup>b</sup>	0.285 <sup>a</sup> 0.302 <sup>b</sup>
$\theta=1$	some difficulty with at least one very physical work function	0.160 0.146	0.119 0.107	$\theta=5$ :	some difficulty with at least one basic function	0.209 0.217	0.286 0.299
$\theta=2$	very difficult/can't do at least one very physical work function	0.148 0.143	0.110 0.101	$\theta=4$ :	very difficult/can't do at least one physical or sedentary work function	0.271 0.279	0.357 0.365
$\theta=3$	some difficulty with at least one physical or sedentary work function	0.114 0.104	0.093 0.085	$\theta=3$ :	some difficulty with at least one physical or sedentary work function	0.398 0.411	0.458 0.469
$\theta=4$	very difficult/can't do at least one physical or sedentary work function	0.089 0.076	0.089 0.076	$\theta=2$ :	very difficult/can't do at least one very physical work function	0.422 0.431	0.479 0.490
$\theta=5$	some difficulty with at least one basic function	0.081 0.070	0.078 0.065	$\theta=1$ :	some difficulty with at least one very physical work function	0.482 0.495	0.527 0.536

Note: Functional limitation index defined by Loprest et al. (1995):

- Level 1: Some difficulty with very physical work functions (e.g., climb several flights of stairs without resting)
- Level 2: Very difficult/can't do one of the very physical work functions
- Level 3: Some difficulty with physical or sedentary work functions (e.g., walk several blocks or sit for two hours)
- Level 4: Very difficult/can't do one of the physical work or sedentary work functions
- Level 5: Some difficulty with any basic function (e.g., get in and out of bed without help)
- Level 6: Very difficult/can't do one of the basic functions

<sup>a</sup>MIV point estimates, corrected for finite-sample bias

<sup>b</sup>bootstrapped 5<sup>th</sup> and 95<sup>th</sup> percentile bounds

Table IV. Corrupt Sampling, Partial Verification, and MIV Bounds on  $P(W=1)$

Unable to Work Case

(A) HM Corrupt Sampling Bounds*	(B) Proposition 1 Verification Bounds <sup>†</sup>	(C) Age and Employment MIV Bounds		(D) Disability Application and Employment MIV Bounds	
		$v_y = 1$	$v_y = 0.9$	$v_y = 1$	$v_y = 0.9$
[0.036, 0.166] <sup>a</sup>	[0.044, 0.109]	[0.049, 0.089] <sup>c</sup>	[0.032, 0.089]	[0.048, 0.072]	[0.045, 0.072]
[0.033, 0.174] <sup>b</sup>	[0.041, 0.114]	[0.042, 0.097]	[0.025, 0.097]	[0.042, 0.084]	[0.038, 0.084]
		+0.009 -0.008 <sup>d</sup>	+0.011 -0.008	+0.003 -0.002	+0.003 -0.002

<sup>†</sup>  $v = 0.935$ : Reported work incapacity ( $X=1$ ) is treated as verified if the respondent receives disability benefits and reports one of the six serious diagnosed conditions highlighted by Wallace and Herzog (1995): treated for cancer in the last 12 months, diabetic taking insulin, chronic lung disease that limits activities, congestive heart disease with treatment or shortness of breath, stroke with health consequences, or current psychiatric/emotional problem with medication or other treatment. Reported work capacity ( $X=0$ ) is verified for workers ( $L=1$ ). For nonworkers, work capacity is verified unless the respondent reports some work limitation and the receipt of disability benefits.

Results for Model II, which relaxes some of these verification assumptions, are available from the authors.

<sup>a</sup>point estimates of the population bounds when  $v = 0.935$ , the verification rate in Column B

<sup>b</sup>bootstrapped 5<sup>th</sup> and 95<sup>th</sup> percentile bounds

<sup>c</sup>MIV point estimates, corrected for finite-sample bias

<sup>d</sup>estimated finite-sample bias

Table V. Age and Employment MIV Bounds on Work Incapacity  
among SSDI/SSI Applicants

P(Unable to Work) Among SDI/SSI Applicants		
A. All Applicants		B. Applicants with the Most Recent Adjudication Date Within Six Months of the Interview Date
Verification, No MIV <sup>†</sup>	Verification Plus Age and Employment MIV	Verification, No MIV
[0.381, 0.791]	[0.505, 0.751] <sup>a</sup>	[0.335, 0.837] <sup>d</sup>
[0.357, 0.811]	[0.441, 0.790] <sup>b</sup>	[0.288, 0.871] <sup>b</sup>
	+0.056 -0.062 <sup>c</sup>	

Note: Case A (all applicants) imposes the age and employment MIV assumption. We do not impose the MIV assumption for Case B due to insufficient sample sizes.

<sup>†</sup>See verification definition in previous table footnote

<sup>a</sup>MIV point estimates, corrected for finite-sample bias

<sup>b</sup>bootstrapped 5<sup>th</sup> and 95<sup>th</sup> percentile bounds

<sup>c</sup>estimated finite-sample bias

<sup>d</sup>point estimates (no MIV assumption)