IOWA STATE UNIVERSITY

Inferring Disability Status from Corrupt Data

Brent Kreider, John V. Pepper

March 2003

Working Paper # 03007

Department of Economics Working Papers Series

Ames, Iowa 50011

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

August 2006

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 find that inferences are highly sensitive to the maintained assumptions – especially to how one models potential inconsistencies between subjective self-assessments of work limitation and more objective measures of functional limitation. We estimate tight bounds under our strongest assumptions but then find that identification deteriorates rapidly as the assumptions are relaxed.
JEL classification: C14, I1

*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. A person's self-assessed degree of work capacity, in particular, may be influenced by a variety of economic, psychological, and social factors. Work disability, after all, is not a purely medical phenomenon; two individuals with identical medical pathologies may have different abilities to work in the labor market. The potential for large classification errors has been widely accepted as a central problem for social science research and for administrative purposes in defining eligibility for government assistance programs (e.g., U.S. General Accounting Office, 1997; 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 (1991a) vs. Parsons (1991)).[1]

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.

Many researchers are skeptical of the accuracy of these self-reports. Bound and Burkhauser

[1] Bound (1991b) provides a comprehensive analysis of the econometric issues surrounding the debates.

(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 may 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. The psychology literature discusses the potential medical role of "negative affectivity" in respondents' self-assessments of disability status (e.g., Watson and Clark, 1984).

Studies that have modeled and assessed the reliability of self-reported work limitations have not resolved these issues. 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 reporting errors. Kerkhofs and Lindeboom (1995), O'Donnell (1999), Kreider (1999, 2000), and Lindeboom and Kerkhofs (2004), 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 are difficult to reconcile. Most related studies impose what seem to be sensible restrictions on the reporting process. However, to address structural questions involving the simultaneous interactions between health status, government assistance programs, and labor market behaviors, these studies also impose strong parametric assumptions.

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 Administrations's (SSA) award decision. Under the identifying assumption that the SSA's definition of disability forms the social standard for what constitutes work incapacity (see Sections 2 and 4.3), they find that the disability self-reports are largely accurate. As they acknowledge, however, questions about the reliability of self-reported disability remain. Well-documented concerns about the reliability of SSA award decisions (U.S. General Accounting Office, 1997; Institute of Medicine, 2002) coupled with the possibility that self-reports may be influenced by the award outcome itself (possible for about a third of their sample for which self-reported disability status was recorded after the award decision) could lead both measures to be biased. More generally, even if self-reported work incapacity is unbiased within the pool of disability applicants, this result may not extend to nonapplicants or to less stringent notions of work disability of interest in other settings (see Section 2).

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. By narrowly focusing on the problem of inferring disability, we abstract away from the parametric assumptions used in the structural models, focusing instead on the identifying power of the more primitive assumptions about the reporting process.

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. We do not focus on providing point estimates of the true disability rates. Instead, in extending the nonparametric bounding methods developed by Horowitz and Manski (1995), we provide a unifying framework that allows us to explore 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 corresponding 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 developing these models, we extend the econometric literature by providing sharp bounds on the mean of a binary random variable when nothing is known about the accuracy of the classifications for an unverified portion of the sample while random errors may occur in the remaining portion. Our Proposition 1 bounds provide an explicit treatment of binary outcomes in the corrupt sampling setting.[2] Moreover, by allowing for random classification errors within verified subgroups, we depart from both the nonparametric (e.g., Manski, 1995; Dominitz and Sherman, 2004) and parametric (e.g., Kreider, 1999; McGarry, 2004) literatures that assume fully accurate reporting within certain subgroups. We also formally correct for the finite sample bias of the plug-in MIV estimator that arises from taking sups and infs over collections of estimates. Without this correction, the estimated MIV bounds would be too narrow.

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] Bollinger (1996) has previously bounded the mean regression in the classical errors-in-variables setting when the mismeasured regressor is binary. Kreider and Pepper (forthcoming) relax the independence assumption, allowing the binary regressor to come from a corrupt sampling process. To illustrate their methodology, they analyze what can be learned about employment outcomes among the disabled when disability status is unobserved.

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. Inferences are quite sensitive to the specific verification model. Under our strongest assumptions, we precisely identify the prevalence of disability and find some evidence of systematic misreporting. Under more conservative models, however, we can only bound the prevalence rate to lie within a wide (e.g., over 20 point) range which includes the accurate reporting rate.

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.[3] 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. 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.

Our analysis focuses 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

[3] We have examined the robustness of our main results using data from the 1996 panel of the Survey of Income and Program Participation (SIPP), a nationally representative sample of 36,800 households. For the SIPP, we use information from all 60,265 individuals between the ages of 18 and 69, the age range surveyed about the existence of work limitations. We further check robustness using data from the U.S. Census. Our primary results are consistent across these data sources, though we cannot replicate many of our models due to a lack of comparable data. These auxiliary results are available from the authors. See Maag and Wittenburg (2003) for discussion about disability measurement issues for the SIPP.

unable to work altogether.[4] 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 (especially for computationally expensive dynamic programming models) 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).[5] In particular, these direct disability questions capture the notion of both physical and mental limitations as well as the more elusive ideas involving social context. These ideas are reflected in Nagi's (1965) seminal work and espoused by both the World Health Organization (WHO) and framers of the 1990 Americans with Disabilities Act (ADA). Nagi (1965) relates disability to "the expression of a physical or a mental limitation" in a social context such as the workplace. The WHO framework is similar: "In the context of a health experience, a disability is any restriction or lack of ability (resulting from an impairment) to perform an activity in the manner or within the range considered normal for a human being" (World Health Organization, 1980). Under the ADA, disability requires the presence of "a physical or mental impairment that substantially limits one or more major life activities," such as performing tasks on the job (Americans with Disabilities Act, 1990). Each of these conceptualizations of disability allows for the possibility that substantial work capacity remains.[6]

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

[4] Focusing on long-term disability, we code a respondent reporting a work limitation as not disabled if the underlying health problem is reported to be only temporary and expected to last for less than three months (77 such cases). There were 22 missing values for these questions that we code as not disabled for purposes of presenting descriptive statistics. When estimating bounds on disability rates, however, we take worst cases and allow for the possibility that some or all of these respondents may be truly disabled.

[5] Using the HRS data, Benítez-Silva et al. (1999) find that self-reported disability status constitutes a powerful predictor of disability insurance applications and awards.

[6] See the Institute of Medicine (2002, Chapter 2) for further discussion of the conceptual issues in defining disability.

work" definition may be of more interest. For example, the Social Security Administration (U.S. Social Security Administration, 2006) 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." 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.

As elaborated in Section 4, our estimated bounds on the true disability rates combine these self-reports of work limitation with other information in the HRS that can potentially shed light on the reliability of these self-reports. 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 is 2.6 times higher among those reporting no work limitation compared with those reporting some limitation (78% compared with 30%). Likewise, just over half the respondents reporting work limitations and nearly four-fifths reporting being unable to work altogether had applied for federal disability benefits from SSDI or Supplemental Security Income (SSI), whereas very few respondents reporting no work limitation had applied for benefits. Because the HRS does not distinguish between SSDI and SSI applications, from this point on we refer to SSDI/SSI jointly as the SSA's Disability Insurance (DI) program.[7]

Although work disability is not synonymous with general health status, there is undoubtedly a close relationship given the potential impact of health conditions on work capacity. Our analysis exploits a wealth of information on a respondent's reported physical and mental health to aid in

[7] 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 limitation eligibility criteria for the two programs are identical, requiring a medically determinable impairment that prevents the applicant from engaging in any "substantial gainful activity." SSI benefits are means-tested and do not require prior work history, whereas SSDI benefits are set according to a recipient's prior earnings. Based on personal correspondence with an SSA official, it appears that SSDI applications are routinely screened for potential SSI eligibility and vice versa. In this light, applicants are effectively applying for benefits from both programs.

implementation of the verification and MIV assumptions. For example, a respondent reporting to be disabled but in excellent physical and mental health might not be verified as providing an accurate disability report.

In Table I, we display the means of these health-related variables by reported disability status. Table II displays a matrix of correlation coefficients for the numerous health and limitation measures used in this analysis. We observe three general categories of health-related information: subjective measures of general health, alternative measures of physical limitation, and specific health conditions.

At the most basic level, we exploit information from two generic questions about a respondent's physical and mental health status. About 42% of respondents reporting a work limitation claim to be in fair or poor general physical health, compared with only 12% among those reporting no work limitation. The correlation coefficient between these two measures is 0.52. Patterns are similar for reported emotional health status, although the correlation coefficient is only 0.31.

Another series of health-related questions provides information on other measures of limitation. To avoid using subjective self-reports of disability, some researches have relied on indirect summary measures like body mass (e.g., Gruber and Kubik, 1997) or subsequent mortality (e.g., Parsons, 1980).[8] In our sample, 3% of respondents died before the second interview could take place (Table I) and nearly 60% have a body mass outside the ideal range defined by Fahey et al. (1997).[9] 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%). While not particularly large, differences in the ideal body mass indicator

[8] Others, however, have criticized indirect measures as being poor indicators of disability status. Haveman and Wolfe (1984), for example, view mortality experience as a "weak and arbitrary" proxy for disability status. Many conditions affecting disability status (e.g., back problems) do not normally contribute to an early death, and many deaths occur for reasons unrelated to the source of a work limitation.

[9] They define ideal body mass to be 20-25 kilograms per meter. We calculate body mass for each respondent using information in the HRS on height and weight.

across reported disability status are also statistically significant at the 1% level. Nevertheless, these alternative measures appear to provide different information about limitations – the means vary from 0.03 for subsequent mortality to 0.209 for self-reported disability to 0.589 for non-ideal body mass, and the correlations range from only 0.02 to 0.13.

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. For each of these functional activities, a respondent can answer "not at all difficult," "a little difficult," "somewhat difficult," "very difficult/can't do," or "don't do."[10] Respondents were told to disregard any limitation expected to last less than three months.

Using definitions suggested by Loporest et al. (1995, p. S297), we aggregate this information into an index of functional limitations. The first step is to create four categories of functions: (I) *Basic functions* include the ability to (a) get in and out of bed without help, (b) bathe or shower without help, (c) eat without help, (d) dress without help, and (e) walk across a room; (II) *Sedentary work functions* include the ability to (a) sit for about two hours and (b) get up from a chair after sitting for a long period; (III) *Physical work functions* include the ability to (a) walk several blocks, (b) stoop/ kneel/crouch, (c) pick up a dime from a table, and (d) reach or extend arms above shoulder level; (IV) *Very physical work functions* include the ability to (a) climb several flights of stairs without resting, (b) lift or carry weights over 10 pounds, and (c) pull or push large objects like a living room chair. Given these four categories, the functional limitation index is then

[10] As pointed out by Loporest et al. (1995), the language "don't do" is somewhat problematic for interpretation since this response may not reflect an inability to perform the task. Nevertheless, we follow their approach and group the "can't do" and "don't do" responses together.

constructed based on the following outcomes:[11]

- Level 0: No functional limitation
- Level 1: Some difficulty with very physical work functions
- Level 2: Very difficult/can't do one of the very physical work functions
- Level 3: Some difficulty with physical or sedentary work functions
- Level 4: Very difficult/can't do one of the physical work or sedentary work functions
- Level 5: Some difficulty with any basic function
- Level 6: Very difficult/can't do one of the basic functions

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, nearly 14% of those reporting a work limitation and 5% of those reporting the inability to work altogether do not report difficulty with any of the activities. In Section 4, we examine how inferences on work disability depend on assumptions about the relationship between work disability and functional limitations.

Finally, we exploit self-reported information on the presence of specific clinical health conditions recorded in the HRS. Wallace and Herzog (1995) focus on a subset of 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. In most cases, responses to questions about the presence of a condition can be combined with responses to follow-up questions that help indicate the severity of the condition. Following Wallace and Herzog's definitions, Table I provides prevalence rates for six specific serious and objective conditions: diabetes (plus currently taking insulin), cancer (with treatment in the last 12 months), chronic lung disease (with reported activity limitations), a heart condition (congestive heart disease with prescribed medication or accompanied by shortness of breath), stroke (with reported health consequences), and psychiatric problems (currently taking medication or receiving treatment). We also report on the prevalence of arthritis and hypertension,

[11] Loporest et al. (1995) recognize some ambiguity in the index. For example, the severity of physical versus sedentary functions is unclear, as are comparisons between being unable to perform a physical activity versus having some difficulty with a very physical function. Nevertheless, we find their aggregation approach to represent a significant advancement over the usual approach of simply counting the number of reported ADLs and IADLs.

along with nine medical conditions associated with a major organ system: asthma, back problems, problems with legs or feet, kidney or bladder problems, stomach or intestinal ulcers, high cholesterol, the occurrence of a fracture since age 45, poor eyesight (with glasses), and poor hearing (with hearing aid).

While self-reports on health conditions are almost certainly more objective than self-reports on work disability, they do not specifically measure disability.[12] Across all of the 17 conditions listed in Table I, 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 disability questions are notably ambiguous, survey designers clearly have an expectation that respondents will be able to place these questions about work limitation 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, especially in confidential surveys, a much greater concern in the literature revolves around the possibility that social or psychological factors can lead to self-rationalization. Concerns over systematic misreporting are generally based on two distinct observations, one financial and one social. First, eligibility for government disability assistance programs is tied to both earnings and disability status (Bound and Burkhauser, 1999). Second, some people may feel social pressure to be working until normal retirement age (Bound, 1991b; Bowe, 1993). Thus, short of in-

[12] Also, despite the relative objectivity of specific conditions, the potential for misreporting remains. For example, conditions may be misdiagnosed, and respondents may be reluctant to disclose the presence of a condition. Conditions might often go undiagnosed for some time, especially among respondents with limited access to health care providers.

tentionally 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.

To evaluate the impact of invalid response, we introduce notation that distinguishes between self-reports and the truth. Let X be the self-reported measure, where $X = 1$ if the respondent reports a limitation 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:

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

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 can be assumed 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.

To narrow these identification bounds, we consider two different classes of assumptions linking observed covariates to the reporting process and the true disability rate. In Section 3.1, we consider verification assumptions that place more confidence in some responses than others. In Section 3.2, we consider the identifying power of monotonicity assumptions linking disability and observed covariates such as age, employment, and the presence of physician-diagnosed health conditions. In Section 3.3, we describe the estimator and, in particular, focus on a bias correction for the plug-in estimator of the nonparametric instrumental variable bound.

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:

$$P(W = 1) = P(W = 1|Y = 1)P(Y = 1) + P(W = 1|Y = 0)P(Y = 0). \tag{4}$$

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. So, in contrast to the existing literature, we allow for the possibility of exogenous response errors within the verified group such that there can be *partial verification*. 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: $P(Z = 1|Y = 1) \geq v_y \geq \frac{1}{2}$.[13] Let the reporting errors in the verified group be random so that $P(W = 1|Y = 1) = P(W = 1|Y = 1, Z)$.[14] No prior information is assumed about the validity of self-reports from the unverified cases. Then the following bounds follow (see Appendix A for a proof):[15]

[13] 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. In fact, however, the bounds derived in Proposition 1 extend to all $v_y \geq \min\{P(X = 1|Y = 1), P(X = 0|Y = 1)\}$. Otherwise, the Proposition 1 bounds that follow do not apply.

[14] We relax this independence assumption, termed *contaminated sampling* by Horowitz and Manski (1995), as part of the sensitivity analysis in Section 4.

[15] Molinari (2005) independently derives a similar result using a different approach that does not focus on partial verification. For a wide class of models, she shows that the relationship between the distribution of a true variable and its potentially mismeasured counterpart can be represented by a linear system of simultaneous equations involving a coefficient matrix of misclassification probabilities. She then shows how restrictions on this matrix (depending on the underlying assumptions of the model) can be used to partially identify regions for the true variable.

Proposition 1: Let $\eta_i = 1\left[v_y > P(X = i|Y = 1)\right]$, $i = 0, 1$ and $\theta = 1[P(X = 0|Y = 1) > P(X = 1|Y = 1)$. Then 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

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

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

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

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 provide accurate reports ($v_y = 1$)[16] then

$$P(X = 1, Y = 1) \leq P(W = 1) \leq P(X = 1, Y = 1) + P(Y = 0). \tag{5}$$

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. Thus, the width of this bound is the fraction of unverified cases, $P(Y = 0)$. Intuitively, for example, in the special case where all workers are known to provide accurate reports about limitation, then the true disability rate must be at least as high as the fraction of workers claiming limitation but no larger than this fraction plus the fraction of nonworkers.

3.2. MONOTONICITY ASSUMPTIONS

The Propositions 1 bounds 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.

[16] This model is precisely the case of censored outcomes considered by Manski (1995). The assumption of fully accurate reporting within certain groups was also evaluated by Lambert and Tierney (1997) and Dominitz and Sherman (2004) for the case of contaminated sampling.

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

$$\sup_{u_0 \geq u_1} LB(u_1) \leq P(W = 1|u = u_0) \leq \inf_{u_0 \leq u2} UB(u_2). \tag{6}$$

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. If the disability rate weakly increases in u, then

$$\sum_{u_0 \epsilon U} P(u = u_0)\{\sup_{u_0 \geq u_1} LB(u_1)\} \leq P(W = 1) \leq \sum_{u_0 \epsilon U} P(u = u_0)\{\inf_{u_0 \leq u_2} UB(u_2)\}.$$

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

Since the MIV assumption alone has no identifying power, we combine this assumption with the previous verification assumptions. In this case, the MIV can have identifying power if the verification probability or the observed disability rate is not monotonic in age.

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 (6) must be imposed over collections of various estimates. In finite samples, estimators that take sups and infs are systematically biased such that the estimated bounds will be too narrow.

To measure and correct for this bias, we present 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 θ 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 $\widehat{b} = E^*(T_n) - T_n$ where $E^*(\cdot)$ is the expectation operator with respect to the bootstrap distribution. A bootstrap bias-corrected estimator then follows as $T_n^c = T_n - \widehat{b} = 2T_n - E^*(T_n)$.[17]

For this application, let the parameter of interest, θ, be the MIV lower bound on $P(W = 1)$ (the upper bound case is analogous), let $LB_n(j)$ be the plug-in estimate of the MIV lower bound on $P(W = 1|u = j)$ for each age group $j = 1, ..., J$, and let T_n be the uncorrected MIV lower bound estimate across all age groups. To estimate these bounds, we divide the sample into 39 age groups containing 252 respondents per group (251 in four of the groups). Then for each cell, the verification bounds are estimated and the MIV restrictions in Equation (6) are applied. Finally, we compute the MIV lower bound, $T_n = \sum_{j \epsilon U} P_n(j)\{\sup_{j \geq j'} LB_n(j')\}$, where $P_n(j)$ is the fraction of respondents in age group j.

The bias b_n is estimated using the bootstrap sampling distribution of T_n. In particular, we randomly draw with replacement from the empirical distribution $10,000$ independent pseudo-samples of the original data. Then, using these samples, we compute a set of $10,000$ lower bound MIV estimates of $P(W = 1)$. Let T_n^k , $k = 1, ..., 10,000$, be the lower bound bootstrap estimate for the k^{th} pseudo-sample, and let $E^*(T_n) = \frac{1}{10,000} \sum_{k=1}^{10,000} T_n^k$ be the expected lower bound from the bootstrap distribution. Finally, we compute the estimated bias, $\widehat{b}$, and the bias-corrected MIV estimator, $T_n^c = 2T_n - E^*(T_n)$.

The bootstrap is also used to provide a tractable way to form confidence intervals for our estimates of 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

[17] The bootstrap bias correction effectively reduces finite sample bias (in monte-carlo simulations) and is asymptotically efficient at higher orders in a variety of different settings. See, for example, Parr (1983), Efron and Tibshirani (1993), Hahn et al. (2002), and Ramalho (2005). Kreider and Pepper (forthcoming) apply this method of correcting the MIV estimator in a different application.

the bootstrap distribution of the lower bound estimator and the 0.95 quantile of the distribution of the upper bound estimator.[18]

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 we use measures of functional limitation to corroborate subjective measures of work 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.

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; Lindeboom and Kerkhofs, 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. While most of the earlier research uses latent variable models to

[18] Horowitz and Manski (1998) and Manski and Pepper (2000 – see the longer NBER version) also use and discuss Bonferroni intervals to derive confidence intervals for bounds.

study more complex structural questions relating health, government assistance, and labor market behaviors, we focus attention on inferring disability alone. This setting allows us to strip away the parametric assumptions used in most of the previous literature and focus on exploiting the rich HRS data to assess the identifying power of different assumptions on the reporting process. Thus, 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 tailored to the work disability measure of interest. 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. Verification strategies for the narrower disability measure are presented in Section 4.3. 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 conditions listed in Table 1 (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.[19]

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 that deserves additional attention. The 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 limitation; 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"

[19] 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).

as follows: (1) proxy responses are never verified,[20] (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.

4.1.A VERIFICATION BOUNDS

Table III presents the estimated bounds for the true disability rate and their 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, a primary 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 1 contains various possible alternative measures of 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 (see Table 1). Alternative measures lying outside of this range can be rejected as measures of work disability. We see, for example that the incidence of non-ideal body mass (58.9%) and the subsequent mortality rate (3.0%), lie far outside

[20] In our sample, 5 percent of responses come from proxy reports. Lee et al. (2004) compare estimates of the number of disabled by respondent type in an environment in which self-response versus proxy was randomized. Among their primary findings, self-respondents and proxy respondents were equally likely to report disability during the initial interview, but proxy respondents were less likely to report disability in the second wave of the survey.

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. These proxies appear to be measuring other aspects of health.[21]

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. However, the width of the verification bounds increases by three-fold when we move from Model I, where 92% of respondents are verified, to Model II, where 78% 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 still provide only limited information on the true disability rate unless a large fraction of the caseload 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. Consistent with concerns raised by Benítez-Silva et al. (2004), these results suggest that conclusions about reporting errors based on latent variable models are driven largely by parametric assumptions. Moreover, we find that some alternative disability measures do not seem to resolve the identification problems. Instead, these measures appear to capture some other dimension of health or limitation.

4.1.B MIV BOUNDS

We can reduce uncertainty about the disability rate 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 models. First, we combine the assumption that true disability

[21] In a regression framework, these measures might still serve as important control variables for health and limitation, and perhaps as valid instrumental variables for the true disability rate.

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.[22] 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 listed in Table 1, 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, age, education, race, gender, marital status, veteran status, and asset level (details from this regression are available upon request).[23]

The bias-corrected MIV estimates are reported in Columns C and D of Table 3. 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. These bounds

[22] If only workers are verified, the monotonicity assumption is not informative on the upper bound. The lower bound is similar to the Proposition 1 bound except that it is not multiplied by the fraction of verified respondents, $P(Y=1)$. That is,

$$\eta_0 \left(\theta \frac{v_y - P(X=0|L=1)}{2v_y - 1} + (1-\theta) P(X=1|L=1) \right) \leq P(W=1). \tag{7}$$

In the special case that workers are known to provide fully accurate reports of work limitation, $v_y = 1$, the population disability rate is at least as large as the reported disability rate among workers, $P(X=1|L=1)$. Note that this assumption is equivalent to an assumption that the employment rate decreases with disability status: $P(L=1|W=1) \leq P(L=1|W=0)$.

[23] In the HRS, the incidence of moderate to severe functional limitation (e.g., Level 3 and above) is, up to sampling variability, monotonic in age, employment status, and the disability application index.

also do not contain the self-reported rate even if more than 10% of the verified respondents may misreport.[24] 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.

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 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 health (18.6%) 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 all reports are accurate, nor many different alternatives.

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. Arguably, for example, respondents with severe functional limitations who report being able-bodied should 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

[24] This result does not rely on strict independence between reporting errors and true disability status maintained in Proposition 1, $P(W = 1|Y = 1, Z = 0) = P(W = 1|Y = 1, Z = 1)$. To weaken this assumption, let the disability rate among inaccurate reporters be some unknown multiple of the disability rate among accurate reports: $P(W = 1|Y = 1, Z = 0) = \gamma P(W = 1|Y = 1, Z = 1)$ for some $\gamma \in [0, \infty)$, with $\gamma = 1$ under independence. If up to 10% of the verified respondents may misreport (i.e., $v_y = 0.90$), then accurate reporting of work limitation status continues to be rejected under the age-employment MIV assumption when $\gamma \leq 3.2$ – i.e., as long as the true disability rate among inaccurate reporters is no more than 3.2 times the true disability rate among accurate reporters. If up to half may misreport ($v_y = 0.50$), then accurate reporting is rejected as long as $\gamma \leq 1.9$.

strategy that uses self-reports of functional limitation to weaken the verification assumptions. Some researchers might be willing to assume that after accounting for self-reported disability, responses to questions about functional limitation provide no further evidence about work disability. In that case, the results presented in Table III apply. Otherwise, apparent inconsistencies between reports of functional limitation and work limitation serve to caution against verification.

Table IV displays the estimated bounds for verification Models I and II under the age-employment MIV assumption.[25] Column 1 presents the lower bound estimates when no respondents are verified if they report being disabled with a sufficiently low functional limitation index value (i.e., a value less than or equal to the particular level of θ). Column 2 presents the upper bound when no respondents are 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 θ). When we do not verify respondents claiming to be disabled but also claiming to have no functional limitations, 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 (Level 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 (Level 3), the upper bound increases to nearly 40%.

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 reliable, we provide evidence that indicators of work limitation and functional limitation are measuring markedly different aspects of impairment.

[25] Analogous results under the DI index MIV assumption are available in Appendix B Table I.

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 more 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 Models I and II above 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. In Model I, reported work incapacity ($X = 1$) is verified if the respondent is receiving disability benefits and reports one of the six aforementioned diagnosed 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. This percentage is slightly higher than the 91.9% obtained above for the "some work limitation" case because we verify the vast majority of responses reporting work capacity.

Under more conservative assumptions in Model II, we begin by unverifying reports from proxy respondents. Otherwise, reported work incapacity is verified only if the respondent reports "disabled" as current employment status (V2701), reports the current receipt of disability benefits, and reports one of the six diagnosed conditions. Self-reported work capacity is no longer verified for workers if the respondent reports some work limitation and either zero/missing labor hours or earnings. The reported ability to work among nonworkers is not verified if the respondent reports some work limitation and any of the following are indicated: the receipt of disability benefits, one of the six diagnosed conditions, or the person responded "disabled" as current employment status. Under these more conservative assumptions, self-reports of work capacity are verified for 85.4 percent of the sample.

Table V presents the base results. As before, we find that the verification and MIV restrictions confer substantial identifying power. Consider verification Model I, where nearly 94 percent of the respondents are verified as providing accurate reports. In this case, 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 under verification Model I do not contain the self-reported rate of 10.1%, a finding that is replicated under notable departures from full verification.[26] Under Verification Model II, however, the upper bound increases to 0.163, much higher than the self-reported rate.

These verification bounds decay further after incorporating information about reported functional limitations. In Table VI, we present estimates that require different levels of corroboration between measures of functional limitation and work limitation before verifying disability responses. When using Level 3 as our corroboration cutoff (some difficulty with physical or sedentary work functions), the Model I lower bound falls from 0.049 to 0.040 and the upper bound increases from 0.089 to 0.362. Under Model II, the lower bound decreases to 0.032 and the upper bound increases to 0.413.

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

[26] If up to 10% of the verified respondents may misreport (i.e., $v_y = 0.90$), then accurate reporting under the broad disability definition is rejected in the HRS as long as $\gamma \leq 11$ (see footnote 24). If up to half may misreport ($v_y = 0.50$), then accurate reporting is rejected as long as $\gamma \leq 5$.

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. That is, the applicant is found to be able to engage in substantial gainful activity. 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. Among the 9824 age-eligible respondents, 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 successful 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.[27] Of the 617 initially denied cases, 430 continued through the appeals process, and 263 of these appeals were successful.[28] 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. In Table VII, we compare data on self-reports and SSA award decisions for two different time windows. In Panel A, we consider the relatively large subgroup

[27] 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. None of the qualitative results depend on how we classify the relatively few cases for which there is some ambiguity.

[28] 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.

of all (age-eligible) HRS respondents who applied to receive DI benefits regardless of when the application was filed. In Panel B, 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. These two panels display the joint frequency distribution of the self-reports and DI beneficiary status. For both observation windows, self-reported work incapacity and DI beneficiary status generally concur, but this is clearly 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.[29]

Bounds on the true rate of work incapacity 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.

Table VIII presents the bounds on the true work incapacity rate for both observation windows. Bounds under Models I and II are provided in both cases, and the age-employment MIV bounds

[29] 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 are likely to 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. This allows them to observe new and repeat applications that are not included in our subsample of DI applicants. Second, 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). Third, Benítez-Silva et al. do not restrict the sample to age-eligible respondents nearing retirement age. 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. As shown in Appendix B Table II, the fraction of beneficiaries rises to 66% in the full sample and to 51% using the shorter horizon if we reclassify the 75 successful applicants no longer receiving benefits as beneficiaries.

are provided for the larger subpopulation of all age-eligible HRS applicants.[30] 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 Model I 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 nearly 75%.

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. In addition to random errors associated with self-reports associated with somewhat ambiguous questions, systematic 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 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 (ADA) in 1990, these measures have 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.

This paper provides and illustrates a methodology for partially identifying work disability rates.

[30] The MIV estimates are unreliable when using the smaller sample of 233 respondents.

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. Under strong assumptions, we are able to nearly identify the disability rate. Identification of this controversial parameter decays as some of 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 are 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 the end, however, our results do not imply that the use of self-reported disability should be abandoned. To the contrary, these self-reports seem to provide valuable information about work capacity beyond that captured in alternative measures of health (such as the existence of specific medical conditions, functional limitations, or proxy measures like body-mass and subsequent mortality). 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. Still, there are currently large gaps in our knowledge about the extent to which policy conclusions are being driven by untenable assumptions on the reporting error processes. Given this uncertainty, there exists a need for better information on the degree and nature of reporting errors on work limitation. 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.

References

[1] Americans with Disabilities Act of 1990. United States Public Law 101-336, 104 Stat. 327. [Online]. Available: http://www.usdoj.gov/crt/ada/pubs/ada.txt.

[2] 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.

[3] 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(2), 147-178.

[4] 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(6): 649-670.

[5] Bollinger, C. (1996). "Bounding Mean Regressions When A Binary Variable is Mismeasured." *Journal of Econometrics*, 73(2), 387-99.

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

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

[8] 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.

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

[10] 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(3), 319-338.

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

[12] Efron, B. and R. J. Tibshirani. (1993). An Introduction to the Bootstrap. Chapman and Hall. New York: NY.

[13] 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, p. 137.

[14] Gruber, J. and J. Kubik. (1997). "Disability Insurance Rejection Rates and the Labor Supply of Older Workers," *Journal of Public Economics*, 64, 1-23.

[15] Hahn, J., Kuersteiner, G., and Newey, W. (2002), "Higher Order Properties of the Bootstrap and Jackknife Bias Corrected Maximum Likelihood Estimators," MIT Working Paper.

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

[17] Horowitz, J. and Manski, C. (1995). "Identification and Robustness With Contaminated and Corrupted Data." *Econometrica*, 63(2), 281-02.

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

[19] Institute of Medicine. (2002). <u>The Dynamics of Disability: Measuring and Monitoring Disability for Social Security Programs. Committee to Review the Social Security Administration's Disability Decision Process Research</u>, G.S. Wunderlich, D.P. Rice and N.L. Amado, Editors, National Research Council, Washington, D.C., National Academy Press.

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

[21] 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.*

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

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

[24] Kreider, B. and J. Pepper (forthcoming). "Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors." *Journal of the American Statistical Association.*

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

[26] Lee, S., Mathiowetz, N. and Tourangeau, R. (2004), "Perceptions of Disability: The Effect of Self- and Proxy Response," *Journal of Official Statistics*, 20(4), 671-686.

[27] Lindeboom, M. and M. Kerkhofs. (2004). "Subjective Health Measures, Reporting Errors, and Endogeneity in the Relationship Between Health and Work," CEIS Tor Vergata Research Paper Series, Vol. 16, No. 46.

[28] Loporest, P., K. Rupp, and S.H. Sandell. (1995). "Gender, Disability, and Employment in the Health and Retirement Study," *Journal of Human Resources*, S293-S318.

[29] Maag, E. and Wittenburg, D. (2003). "Real Trends or Measurement Problems? Disability and Employment Trends from the Survey of Income and Program." mimeo, Urban Institute.

[30] Molinari, F. (2005). "Partial Identification of Probability Distributions with Misclassified Data." Cornell University, Working Paper.

[31] Manski, C. and Pepper, J. (2000). "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica*, 68(4): 997-1010

[32] Manski, C. (1995). Identification Problems in the Social Sciences. Harvard University Press, Cambridge, MA.

[33] McGarry, K. (2004). "Health and Retirement: Do Changes in Health Affect Retirement Expectations?." *Journal of Human Resources*, 39(3): 624-648

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

[35] Nagi, S. (1965). "Some Conceptual Issues in Disability and Rehabilitation" in Sociology and Rehabilitation. Sussman M.B., ed. Washington, DC: American Sociological Association.

[36] National Council on Disability. (2002). National Disability Policy: A Progress Report. Dec. 2000-Dec. 2001.

[37] O'Donnell, O. (1999). "Employment, Disability, and Work Incapacity." mimeo. University of Kent.

[38] Parr, W. (1983), "A Note on the Jackknife, the Bootstrap and the Delta Method Estimators of Bias and Variance," *Biometrika*, 70, 719-22.

[39] Parsons, D. (1980). "The Decline in Male Labor Force Participation," *Journal of Political Economy*, 88, 117-134.

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

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

[42] Ramalho, J. (2005) "Small Sample Bias of Alternative Estimation Methods for Moment Condition Models: Monte Carlo Evidence for Covariance Structures," *Studies in Nonlinear Dynamics and Econometrics*, 9, Article 3.

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

[44] U.S. General Accounting Office, "Social Security Disability: SSA Must Hold Itself Accountable for Continued Improvement in Decision-making." HEHS-97-102 August 12, 1997.

[45] 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). [online]: http://www.ssa.gov/OP_Home/ssact/title02/0223.htm

[46] 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.

[47] Watson, D. and L.A. Clark. (1984). "Negative Affectivity: The Disposition to Experience Negative Emotional States." *Psychological Bulletin*, 96, 465-490.

[48] World Health Organization. (1980). International Classification of Impairments, Disabilities, and Handicaps. Geneva.

APPENDIX A

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:

$$P(X = 1) = P(W = 1|Z = 1)P(Z = 1) + P(W = 0|Z = 0)P(Z = 0). \tag{8}$$

The independence assumption requires $P(W = 0|Z = 1) = P(W = 0|Z = 0)$. Substituting for $P(W = 0|Z = 0)$ in (**??**) 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 $P(Z = 1)$. $\square$

Table I. Means and Standard Deviations

	Full Sample N=9824		Reported Work Limitation? Yes N=2039	Reported Work Limitation? No N=7785	Reported Inability to Work Altogether? Yes N=992	Reported Inability to Work Altogether? No N=8832
	Mean	Std. Dev.	Mean	Mean	Mean	Mean
Age	56.0	3.18	56.4*	55.9	56.6*	55.9
Years of schooling	12.0	3.24	10.8*	12.3	10.3*	12.2
Female	0.532	0.499	0.538	0.530	0.526	0.532
Nonwhite	0.286	0.452	0.359*	0.267	0.435*	0.269
White collar occupation[a]	0.246	0.431	0.132*	0.276	0.106*	0.262
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
Reported fair/poor general health status[b]	0.186	0.389	0.423*	0.124	0.535*	0.147
Reported fair/poor emotional health status[b]	0.227	0.419	0.651*	0.115	0.824*	0.160
Often bothered by pain, at least moderate at its worst	0.219	0.414	0.573*	0.126	0.673*	0.168
Pain interferes with normal work	0.165	0.371	0.537*	0.067	0.665*	0.109
Died prior to second wave	0.030	0.171	0.075*	0.018	0.111*	0.021
Body mass index out of ideal range[c]	0.589	0.492	0.622*	0.580	0.628*	0.584
ADL/IADL functional limitation index (0-6)[d]	1.51	1.84	3.34*	1.04	3.98*	1.24
Level 0: No functional limitation	0.540	0.498	0.139*	0.645	0.047*	0.595
Level 1: Some difficulty with very physical work functions	0.055	0.229	0.043*	0.059	0.028*	0.058
Level 2: Very difficult/can't do one of the very physical work functions	0.031	0.175	0.038†	0.030	0.031	0.031
Level 3: Some difficulty with physical or sedentary work functions	0.177	0.382	0.220*	0.166	0.175	0.178
Level 4: Very difficult/can't do one of the physical work or sedentary work functions	0.141	0.348	0.337*	0.089	0.374*	0.114
Level 5: Some difficulty with any basic function	0.034	0.181	0.131*	0.008	0.194*	0.016
Level 6: Very difficult/can't do one of the basic functions	0.022	0.145	0.092*	0.003	0.150*	0.007

(continued)

Table 1, Continued

Specific health conditions[e]						
Diabetes (current and taking insulin)	0.083	0.276	0.172*	0.060	0.191*	0.071
Cancer (treatment in last 12 months)	0.037	0.190	0.063*	0.031	0.076*	0.033
Chronic lung disease (with activity limitations)	0.032	0.176	0.127*	0.007	0.184*	0.015
Heart condition (congestive heart disease with medication or accompanied by shortness of breath)	0.084	0.277	0.227*	0.046	0.290*	0.061
Stroke (with health consequences)	0.017	0.128	0.067*	0.004	0.108*	0.006
Psychiatric problem (current with medication or treatment)	0.104	0.306	0.246*	0.067	0.301*	0.082
Arthritis	0.344	0.475	0.600*	0.277	0.651*	0.309
Hypertension	0.153	0.360	0.275*	0.121	0.322*	0.134
Conditions by organ system[f]						
Asthma	0.061	0.240	0.120*	0.046	0.151*	0.051
Back problems	0.346	0.476	0.593*	0.281	0.628*	0.314
Problems with legs or feet	0.356	0.479	0.675*	0.273	0.751*	0.312
Kidney or bladder problems	0.107	0.309	0.237*	0.073	0.282*	0.088
Stomach or intestinal ulcers	0.095	0.294	0.193*	0.070	0.233*	0.080
High cholesterol	0.241	0.428	0.309*	0.223	0.339*	0.230
Fracture since age 45	0.138	0.345	0.201*	0.122	0.193*	0.132
Poor eyesight (with glasses)	0.032	0.177	0.101*	0.014	0.126*	0.022
Poor hearing (with hearing aid)	0.024	0.153	0.050*	0.017	0.058*	0.020
Number of reported conditions in previous two categories	2.08	1.92	4.02*	1.58	4.63*	1.78
Reported a severe condition[g]	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

[a] Occupation at onset of disability if reporting a work limitation and this information is available; current or most recent occupation otherwise.

[b] Other categories include excellent, very good, and good health status.

[c] Ideal body mass is defined as 20-25 kg/m^2 following Fahey et al. (1997).

[d] Following Lobrest et al. (1995, p. S297), we construct four categories of functions as described in the text: (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.

[e,f] Defined by Wallace and Herzog (1995, pS89 and Table 1); we additionally include poor eyesight (with glasses) and poor hearing (with hearing aid).

[g] Includes diabetes, cancer, chronic lung disease, heart condition, stroke, or psychiatric condition as defined by Wallace and Herzog (1995, Table 1).

Table II. Correlations Between Health Indicators

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	Self-reported work limitation										
(2)	Self-reported inability to work	0.65*									
(3)	Reported fair/poor general health status	0.52*	0.48*								
(4)	Reported fair/poor emotional health status	0.31*	0.30*	0.45*							
(5)	Died prior to second wave	0.13*	0.16*	0.15*	0.08*						
(6)	Body Mass Index out of ideal range	0.03*	0.03*	0.05*	0.04*	0.02†					
(7)	ADL/IADL functional limitation index (0-6)	0.51*	0.45*	0.47*	0.33*	0.12*	0.07*				
(8)	Often bothered by pain, at least moderate	0.44*	0.37*	0.39*	0.29*	0.05*	0.04*	0.46*			
(9)	Pain interferes with normal work	0.51*	0.45*	0.45*	0.33*	0.09*	0.04*	0.48*	0.75*		
(10)	Number of reported health conditions	0.52*	0.45*	0.54*	0.38*	0.13*	0.05*	0.54*	0.50*	0.52*	
(11)	Reported a severe condition	0.37*	0.33*	0.41*	0.30*	0.13*	0.03*	0.31*	0.23*	0.26*	0.50*

*,† significantly different from zero at the 1% and 10% levels, respectively

Table III. 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	
Verification Model I†					
		$v_y = 1$	$v_y = 0.9$	$v_y = 1$	$v_y = 0.9$
[0.127, 0.288][a]	[0.135, 0.215]	[0.178, 0.204][c]	[0.106, 0.204]	[0.149, 0.193]	[0.120, 0.193]
[0.121 0.298][b]	[0.129 0.223]	[0.155 0.207]	[0.086 0.207]	[0.143 0.209]	[0.112 0.209]
		+0.010 -0.016[d]	+0.016 -0.016	+0.009 -0.005	+0.006 -0.005
Verification Model II‡					
		$v_y = 1$	$v_y = 0.9$	$v_y = 1$	$v_y = 0.9$
[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

†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

‡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.

[a]point estimates of the population bounds
[b]bootstrapped 5th and 95th percentile bounds
[c]MIV point estimates, corrected for finite-sample bias
[d]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 IV. Sensitivity of Age and Employment MIV Bounds when Requiring Functional Limitation Corroboration

Work Limitation Case, Model I

Lower Bound: *X*=1 (reports work limitation) is never verified if ADL limitation index ≤ θ			Upper Bound: *X*=0 (reports no work limitation) is never verified if ADL limitation index ≥ θ		
θ=0:	no functional limitation	0.166[a] 0.152[b]	θ=6:	very difficult/can't do at least one basic function	0.205 0.210
θ=1:	some difficulty with at least one very physical work function	0.160 0.146	θ=5:	some difficulty with at least one basic function	0.209 0.217
θ=2:	very difficult/can't do at least one very physical work function	0.148 0.143	θ=4:	very difficult/can't do at least one physical or sedentary work function	0.271 0.279
θ=3:	some difficulty with at least one physical or sedentary work function	0.114 0.104	θ=3:	some difficulty with at least one physical or sedentary work function	0.398 0.411
θ=4:	very difficult/can't do at least one physical or sedentary work function	0.089 0.076	θ=2:	very difficult/can't do at least one very physical work function	0.422 0.431
θ=5:	some difficulty with at least one basic function	0.081 0.070	θ=1:	some difficulty with at least one very physical work function	0.482 0.495

(continued)

Note: Functional limitation index defined by Loprest et al. (1995). See discussion in text.

[a]MIV point estimates, corrected for finite-sample bias
[b]bootstrapped 5th and 95th percentile bounds

Table IV, Cont. Sensitivity of Age and Employment MIV Bounds when Requiring Functional Limitation Corroboration

Work Limitation Case: Model II

Lower Bound: X=1 (reports work limitation) is never verified if ADL limitation index ≤ θ			Upper Bound: X=0 (reports no work limitation) is never verified if ADL limitation index ≥ θ		
θ=0:	no functional limitation	0.122[a] 0.110[b]	θ=6:	very difficult/can't do at least one basic function	0.285 0.302
θ=1:	some difficulty with at least one very physical work function	0.119 0.107	θ=5:	some difficulty with at least one basic function	0.286 0.299
θ=2:	very difficult/can't do at least one very physical work function	0.110 0.101	θ=4:	very difficult/can't do at least one physical or sedentary work function	0.357 0.365
θ=3:	some difficulty with at least one physical or sedentary work function	0.093 0.085	θ=3:	some difficulty with at least one physical or sedentary work function	0.458 0.469
θ=4:	very difficult/can't do at least one physical or sedentary work function	0.089 0.076	θ=2:	very difficult/can't do at least one very physical work function	0.479 0.490
θ=5:	some difficulty with at least one basic function	0.078 0.065	θ=1:	some difficulty with at least one very physical work function	0.527 0.536

Note: Functional limitation index defined by Loprest et al. (1995). See discussion in text.

[a]MIV point estimates, corrected for finite-sample bias
[b]bootstrapped 5th and 95th percentile bounds

Table V. 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	(C) Age and Employment MIV Bounds		(D) Disability Application and Employment MIV Bounds	
		Verification Model I[†]			
		$v_y = 1$	$v_y = 0.9$	$v_y = 1$	$v_y = 0.9$
[0.036, 0.166][a]	[0.044, 0.109]	[0.049, 0.089][c]	[0.032, 0.089]	[0.048, 0.072]	[0.045, 0.072]
[0.033 0.174][b]	[0.041 0.114]	[0.042 0.097]	[0.025 0.097]	[0.042 0.084]	[0.038 0.084]
		+0.009 -0.008[d]	+0.011 -0.008	+0.003 -0.002	+0.003 -0.002
		Verification Model II[‡]			
		$v_y = 1$	$v_y = 0.9$	$v_y = 1$	$v_y = 0.9$
[0.000, 0.247]	[0.037, 0.184]	[0.043, 0.163]	[0.031, 0.163]	[0.042, 0.152]	[0.040, 0.152]
[0.000 0.256]	[0.034 0.190]	[0.038 0.175]	[0.022 0.175]	[0.037, 0.165]	[0.034, 0.165]
		+0.009 -0.013	+0.010 -0.013	+0.002 -0.007	+0.002 -0.007

[†]Model I (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.

[‡]Model II (v = 0.854): Proxy responses are never verified. Otherwise, reported work incapacity (X=1) is verified if the respondent receives disability benefits, reports a serious/objective diagnosed condition (see above), and checked “disabled” as current employment status. Reported work capacity of workers remains verified unless the respondent reports some work limitation and labor hours or earnings are zero/missing. Reported work capacity among nonworkers remains verified unless the respondent reports some work limitation and any of the following: (a) receipt of disability benefits, (b) a serious/objective condition as defined in Model I, or (c) checked “disabled” as current employment status.

[a]point estimates of the population bounds
[b]bootstrapped 5th and 95th percentile bounds
[c]MIV point estimates, corrected for finite-sample bias
[d]estimated finite-sample bias

Table VI. Sensitivity of Age and Employment MIV Bounds when Requiring Functional Limitation Corroboration

Unable to Work Case, Model I

Lower Bound: *X*=1 (reports work limitation) is never verified if ADL limitation index ≤ θ			Upper Bound: *X*=0 (reports no work limitation) is never verified if ADL limitation index ≥ θ		
θ=0:	no functional limitation	0.046[a] 0.042[b]	θ=6:	very difficult/can't do at least one basic function	0.099 0.108
θ=1:	some difficulty with at least one very physical work function	0.044 0.040	θ=5:	some difficulty with at least one basic function	0.113 0.119
θ=2:	very difficult/can't do at least one very physical work function	0.044 0.037	θ=4:	very difficult/can't do at least one physical or sedentary work function	0.199 0.210
θ=3:	some difficulty with at least one physical or sedentary work function	0.040 0.033	θ=3:	some difficulty with at least one physical or sedentary work function	0.362 0.374
θ=4:	very difficult/can't do at least one physical or sedentary work function	0.026 0.018	θ=2:	very difficult/can't do at least one very physical work function	0.396 0.402
θ=5:	some difficulty with at least one basic function	0.011 0.006	θ=1:	some difficulty with at least one very physical work function	0.453 0.466

(continued)

Note: Functional limitation index defined by Loprest et al. (1995). See discussion in text.

[a]MIV point estimates, corrected for finite-sample bias
[b]bootstrapped 5th and 95th percentile bounds

Table VI, cont. Sensitivity of Age and Employment MIV Bounds when Requiring Functional Limitation Corroboration

Unable to Work Case: Model II

Lower Bound: X=1 (reports work limitation) is never verified if ADL limitation index ≤ θ			Upper Bound: X=0 (reports no work limitation) is never verified if ADL limitation index ≥ θ		
θ=0:	no functional limitation	0.044[a] 0.038[b]	θ=6:	very difficult/can't do at least one basic function	0.170 0.177
θ=1:	some difficulty with at least one very physical work function	0.039 0.033	θ=5:	some difficulty with at least one basic function	0.179 0.186
θ=2:	very difficult/can't do at least one very physical work function	0.038 0.033	θ=4:	very difficult/can't do at least one physical or sedentary work function	0.261 0.272
θ=3:	some difficulty with at least one physical or sedentary work function	0.032 0.027	θ=3:	some difficulty with at least one physical or sedentary work function	0.413 0.426
θ=4:	very difficult/can't do at least one physical or sedentary work function	0.022 0.016	θ=2:	very difficult/can't do at least one very physical work function	0.440 0.450
θ=5:	some difficulty with at least one basic function	0.005 0.004	θ=1:	some difficulty with at least one very physical work function	0.486 0.498

Note: Functional limitation index defined by Loprest et al. (1995). See discussion in text.

[a]MIV point estimates, corrected for finite-sample bias
[b]bootstrapped 5th and 95th percentile bounds

Table VII. Self-reported Inability to Work and the Receipt of SSDI/SSI Benefits

A. All Age-Eligible Applicants

"Can't Work"	Receiving Benefits[a] 0	Receiving Benefits[a] 1	totals
0	193	103	296 (27.4%)
1	248	538	786 (72.6%)
totals	441 (40.8%)	641 (59.2%)	1,082

B. Age-Eligible Applicants with Most Recent Adjudication Within Six Months of the Interview Date

"Can't Work"	Receiving Benefits[a] 0	Receiving Benefits[a] 1	totals
0	37	13	50 (21.5%)
1	87	96	183 (78.5%)
totals	124 (53.2%)	109 (46.8%)	233

[a] Category 1 includes those awarded and still receiving benefits; category 0 includes rejected applicants and those no longer receiving benefits

Table VIII. 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 Model I†			
	No MIV	Age and Employment MIV	No MIV
	[0.381, 0.791]	[0.505, 0.751][a]	[0.335, 0.837][d]
	[0.357 0.811]	[0.441 0.790][b]	[0.288 0.871][b]
		+0.056 -0.062[c]	
Verification Model II‡			
	No MIV	Age and Employment MIV	No MIV
	[0.324, 0.894]	[0.432, 0.863]	[0.275, 0.906]
	[0.302 0.909]	[0.382 0.892]	[0.227 0.936]
		+0.051 -0.051	

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.

†,‡See definitions in text or previous table footnotes

[a]MIV point estimates, corrected for finite-sample bias
[b]bootstrapped 5th and 95th percentile bounds
[c]estimated finite-sample bias
[d]point estimates (no MIV assumption)

APPENDIX TABLES

Appendix Table Ia. Sensitivity of DI and Employment MIV Bounds when Requiring Functional Limitation Corroboration

Work Limitation Case, Model I

Lower Bound: *X*=1 (reports work limitation) is never verified if ADL limitation index ≤ θ			Upper Bound: *X*=0 (reports no work limitation) is never verified if ADL limitation index ≥ θ		
θ=0:	no functional limitation	0.146[a] 0.141[b]	θ=6:	very difficult/can't do at least one basic function	0.198 0.213
θ=1:	some difficulty with at least one very physical work function	0.144 0.139	θ=5:	some difficulty with at least one basic function	0.210 0.230
θ=2:	very difficult/can't do at least one very physical work function	0.129 0.125	θ=4:	very difficult/can't do at least one physical or sedentary work function	0.290 0.296
θ=3:	some difficulty with at least one physical or sedentary work function	0.101 0.095	θ=3:	some difficulty with at least one physical or sedentary work function	0.425 0.430
θ=4:	very difficult/can't do at least one physical or sedentary work function	0.067 0.059	θ=2:	very difficult/can't do at least one very physical work function	0.448 0.454
θ=5:	some difficulty with at least one basic function	0.045 0.039	θ=1:	some difficulty with at least one very physical work function	0.493 0.499

(continued)

Note: Functional limitation index defined by Loprest et al. (1995). See discussion in text.

[a]MIV point estimates, corrected for finite-sample bias
[b]bootstrapped 5th and 95th percentile bounds

Appendix Table Ia, cont. Sensitivity of DI and Employment MIV Bounds when Requiring Functional Limitation Corroboration

Work Limitation Case: Model II

Lower Bound: *X*=1 (reports work limitation) is never verified if ADL limitation index ≤ θ			Upper Bound: *X*=0 (reports no work limitation) is never verified if ADL limitation index ≥ θ		
θ=0:	no functional limitation	0.108[a] 0.100[b]	θ=6:	very difficult/can't do at least one basic function	0.321 0.328
θ=1:	some difficulty with at least one very physical work function	0.106 0.101	θ=5:	some difficulty with at least one basic function	0.324 0.330
θ=2:	very difficult/can't do at least one very physical work function	0.096 0.091	θ=4:	very difficult/can't do at least one physical or sedentary work function	0.373 0.379
θ=3:	some difficulty with at least one physical or sedentary work function	0.075 0.069	θ=3:	some difficulty with at least one physical or sedentary work function	0.481 0.489
θ=4:	very difficult/can't do at least one physical or sedentary work function	0.053 0.046	θ=2:	very difficult/can't do at least one very physical work function	0.499 0.507
θ=5:	some difficulty with at least one basic function	0.038 0.029	θ=1:	some difficulty with at least one very physical work function	0.537 0.547

Note: Functional limitation index defined by Loprest et al. (1995). See discussion in text.

[a]MIV point estimates, corrected for finite-sample bias
[b]bootstrapped 5th and 95th percentile bounds

Appendix Table Ib. Sensitivity of DI and Employment MIV Bounds when Requiring Functional Limitation Corroboration

Unable to Work Case, Model I

Lower Bound: X=1 (reports work limitation) is never verified if ADL limitation index ≤ θ			Upper Bound: X=0 (reports no work limitation) is never verified if ADL limitation index ≥ θ		
θ=0:	no functional limitation	0.046[a] 0.041[b]	θ=6:	very difficult/can't do at least one basic function	0.091 0.102
θ=1:	some difficulty with at least one very physical work function	0.045 0.040	θ=5:	some difficulty with at least one basic function	0.118 0.126
θ=2:	very difficult/can't do at least one very physical work function	0.043 0.037	θ=4:	very difficult/can't do at least one physical or sedentary work function	0.230 0.236
θ=3:	some difficulty with at least one physical or sedentary work function	0.037 0.032	θ=3:	some difficulty with at least one physical or sedentary work function	0.388 0.395
θ=4:	very difficult/can't do at least one physical or sedentary work function	0.020 0.016	θ=2:	very difficult/can't do at least one very physical work function	0.415 0.422
θ=5:	some difficulty with at least one basic function	0.009 0.006	θ=1:	some difficulty with at least one very physical work function	0.467 0.473

(continued)

Note: Functional limitation index defined by Loprest et al. (1995). See discussion in text.

[a]MIV point estimates, corrected for finite-sample bias
[b]bootstrapped 5th and 95th percentile bounds

Appendix Table Ib, cont. Sensitivity of DI and Employment MIV Bounds when Requiring Functional Limitation Corroboration

Unable to Work Case: Model II

Lower Bound: *X*=1 (reports work limitation) is never verified if ADL limitation index ≤ θ			Upper Bound: *X*=0 (reports no work limitation) is never verified if ADL limitation index ≥ θ		
θ=0:	no functional limitation	0.041[a] 0.036[b]	θ=6:	very difficult/can't do at least one basic function	0.164 0.178
θ=1:	some difficulty with at least one very physical work function	0.039 0.034	θ=5:	some difficulty with at least one basic function	0.188 0.197
θ=2:	very difficult/can't do at least one very physical work function	0.037 0.031	θ=4:	very difficult/can't do at least one physical or sedentary work function	0.284 0.291
θ=3:	some difficulty with at least one physical or sedentary work function	0.032 0.028	θ=3:	some difficulty with at least one physical or sedentary work function	0.427 0.436
θ=4:	very difficult/can't do at least one physical or sedentary work function	0.018 0.015	θ=2:	very difficult/can't do at least one very physical work function	0.452 0.460
θ=5:	some difficulty with at least one basic function	0.008 0.006	θ=1:	some difficulty with at least one very physical work function	0.502 0.508

Note: Functional limitation index defined by Loprest et al. (1995). See discussion in text.

[a]MIV point estimates, corrected for finite-sample bias
[b]bootstrapped 5th and 95th percentile bounds

Appendix Table II. Self-reported Inability to Work and the SSDI/SSI Award Decision

A. All Age-Eligible Applicants

"Can't Work"	Granted Benefits[a] 0	Granted Benefits[a] 1	totals
0	155	141	296 (27.4%)
1	211	575	786 (72.6%)
totals	366 (33.8%)	716 (66.2%)	1,082

B. Age-Eligible Applicants with Most Recent Adjudication Within Six Months of the Interview Date

"Can't Work"	Granted Benefits[a] 0	Granted Benefits[a] 1	totals
0	35	15	50 (21.5%)
1	79	104	183 (78.5%)
totals	114 (48.9%)	119 (51.1%)	233

[a] Category 1 includes those awarded benefits, whether still receiving them or not; category 0 includes rejected applicants