

ORES Working Paper Series

Number 90

Counting the Disabled: Using Survey Self-Reports to
Estimate Medical Eligibility for Social Security's Disability Programs

Debra Dwyer,* Jianting Hu,** Denton R. Vaughan,*** and Bernard Wixon**

Division of Economic Research

January 2001

Social Security Administration
Office of Policy
Office of Research, Evaluation, and Statistics

- * State University at Stony Brook
Department of Economics, SUNY Stony Brook, Rm. S625, Social Behavioral Science Bldg.,
Stony Brook, NY 11794-4384
- ** Social Security Administration, Office of Policy
9th floor, ITC Building, 500 E Street SW, Washington, DC 20254-0001
- *** Bureau of the Census, Housing and Household Economics Statistics Division
FOB 3, Room 1473, Washington, DC 20233

Working Papers in this series are preliminary materials circulated for review and comment. The views expressed are the authors' and do not necessarily represent the position of the Social Security Administration. The papers have not been cleared for publication and should not be quoted without permission.

Abstract

This paper develops an approach for tracking medical eligibility for the Social Security Administration's (SSA's) disability programs on the basis of self-reports from an ongoing survey. Using a structural model of the disability determination process estimated on a sample of applicants, we make out-of-sample predictions of eligibility for nonbeneficiaries in the general population. This work is based on the 1990 panel of the Survey of Income and Program Participation. We use alternative methods of estimating the number of people who would be found eligible if they applied, considering the effects of sample selection adjustments, sample restrictions, and several methods of estimating eligibility/ineligibility from a set of continuous probabilities. The estimates cover a wide range, suggesting the importance of addressing methodological issues. In terms of classification rates for applicants, our preferred measure outperforms the conventional single variable model based on the "prevented" measure.

Under our preferred estimate, we find that 4.4 million people—2.9 percent of the nonbeneficiary population aged 18-64—would meet SSA's medical criteria for disability. Of that group, about one-third have average earnings above the substantial gainful activity limit. Those we classify as medically eligible are similar to allowed applicants in terms of standard measures of activity limitations.

I. Introduction

The purpose of this paper is to develop methodological tools needed to track potential growth in the disability programs administered by the Social Security Administration (SSA). Specifically, we simulate medical eligibility for disability benefits for members of the general population who are not receiving such benefits, using data from the 1990 Survey of Income and Program Participation (SIPP). Employing a structural model of the disability determination process developed in Hu and others (1997), we estimate those who would qualify under SSA's definition of disability, as implemented by state Disability Determination Service (DDS) agencies. Eligibles are estimated on the basis of their SIPP responses to questions on health, work, activity limitations, and socioeconomic characteristics. Using that approach, we estimate that 2.9 percent of the general population aged 18-64—4.4 million people—were medically eligible but were not receiving disability benefits as of early 1992.

Simulations of program eligibility are undertaken routinely for social insurance and welfare programs that are not targeted toward the disabled. For example, a projection of the number of people old enough to take Social Security retirement benefits—a straightforward simulation of the nonfinancial element of eligibility—frequently provides the intellectual backdrop for discussions of Social Security reform. In general, eligibility criteria for many programs such as Supplemental Security Income for the Aged (SSI/Aged) and Aid to Families with Dependent Children (AFDC) are based on income, assets, work behavior, and demographic

Acknowledgments. The authors wish to thank several colleagues for their helpful comments: Sharmila Choudhury, Kajal Lahiri, Joyce Manchester, Scott Muller, Cheryl Neslusan, David Pattison, Kalman Rupp, Robert Weathers, and, especially, Benjamin Bridges and Michael V. Leonesio. The authors thank Pat Cole for editing support. This analysis was completed while Denton R. Vaughan was an employee of the Social Security Administration.

characteristics—information reliably observed in national surveys. This permits estimates of the pool of eligibles (see, for example, Blank and Ruggles 1996), giving policymakers a means of evaluating implications of changes in eligibility policy.

Prospects for reliable estimates of disability eligibles have always been far less promising, for two reasons. First, medical eligibility for disability programs depends on true health status and ability to work, neither of which is directly observable in surveys. In fact, the survey information that is collected is, in many instances, self-evaluative and subjective. Second, the disability determination process, which compares the applicant's impairment severity and functional capacity to program standards, is also somewhat judgmental. Due to these limitations, it is difficult to assess medical eligibility among nonapplicants with precision. This poses a handicap for policymakers because eligibility is the primary means by which they control the size and targeting of any public program.

The eligibility simulation presented in this paper builds on our prior research (Hu and others 1997). For that work we matched SSA records on disability applications to SIPP survey information, thus identifying survey sample members who applied for disability around the time of the survey and establishing whether their applications were allowed or denied. Using that sample of applicants, we estimated a statistical model of SSA's allow/deny decision based on survey responses on self-reported health, activity limitations, work, and socioeconomic characteristics. In the current study we apply a reestimated version of that model to nonbeneficiaries to predict whether they would be found medically eligible if they were to apply for benefits. We incorporate a sample selection correction to adjust for the fact that disabled people who choose to apply for benefits may not be a random sample of the disabled in the

general population. In both studies, the matching of disability records to survey data has permitted us to frame the estimation of medical eligibility as an empirical issue. In effect, this approach represents an effort to interpret survey self-reports on health in the light of SSA's evaluations of respondents' health.

This paper makes both methodological and policy contributions. Methodologically, it represents the first attempt to estimate medical eligibility for SSA's disability programs using information from a recurring, nationally representative survey in conjunction with a statistical model of SSA's disability determination process. The three appendices to this report explain our methodology in detail to facilitate its use by other analysts. Our policy contributions include estimating the size of the eligible pool and providing a brief sketch of its characteristics. That estimate suggests the potential for additional program growth as of the time of the survey; in addition, it will serve as a baseline for future estimates. We note, however, that the estimates are preliminary in several respects. The remaining sections of the paper provide some background and a discussion of methodological issues, followed by results and conclusions.

II. Background

SSA administers two disability programs that pay cash benefits to persons unable to work due to a serious impairment, although the programs have distinct policy objectives. Disability Insurance (DI) is a social insurance program. DI benefits are paid to workers who become disabled and who meet the work requirements of the program. Supplemental Security Income (SSI) employs the same medical criteria as DI, but it is a means-tested program providing cash benefits to those disabled or aged who have income and assets below defined thresholds. In

contrast to DI recipients, SSI beneficiaries typically have limited work experience. In recent years the programs have experienced substantial growth: total annual expenditures for the two programs grew from \$26 billion to \$68 billion between 1985 and 1997—an increase of over 150 percent in current dollars.

Recent program growth can be understood in terms of changes in eligibility criteria, changes in incentives to apply, and interaction effects. Research has suggested numerous factors that may affect applications. Application decisions are strongly related to the size of program benefits. Also, program interactions likely play a role. For example, the increasing difficulty in obtaining private health insurance, particularly for the disabled, makes disability benefits more valuable because beneficiaries not only receive cash benefits, but also typically become eligible for Medicare or Medicaid. Moreover, demographic trends have an impact on the size of the applicant pool. For example, with the aging of the general population we expect to see deteriorating health. In addition, the decision to apply for benefits is related to general economic conditions and the state of the labor market, as well as to circumstances within specific households. Much of the literature to date has focused on such aspects of the individual's decision to apply for benefits (Benitez-Silva and others 1999; Bound and others 1995; Halpern and Hausman 1986; Haveman, Wolfe, and Wallich 1988; Kreider 1998; Rupp and Stapleton 1995; Stapleton and others 1994; and Yelowitz 1998). By contrast, the government's decision on eligibility has received much less attention, although eligibility has not only had a role in recent program growth but also represents a direct means of controlling the size and targeting of the programs.

Nonetheless, research suggests that take-up rates for SSI and DI are less than 100 percent, as with most social insurance and assistance programs.¹ The resulting pool of nonparticipating eligibles represents the potential for program growth resulting from recessions or other contingencies that might influence the application decision. Moreover, both changes in eligibility rules and variation in the strictness or leniency with which the rules are applied can also affect the number of potential eligibles. Monitoring changes in the pool of eligibles ensuing from (real or hypothetical) trends or policy initiatives would add much to our understanding of the disability programs. Developing the tools to monitor such changes is the rationale for this study. In the past, such estimates had not been feasible with respect to disability programs for the following reason: surveys represent the main source of information on the health of the general population, yet the relationship between survey self-reports on health and SSA's disability definition had not been subjected to empirical analysis. However, Hu and others (1997) and Lahiri, Vaughan, and Wixon (1995) have recently developed and tested a sequential model of the complex and judgmental disability determination process. That work was based on a data set for a sample of applicants that linked self-reports from the 1990 Survey of Income and Program Participation with administrative data on SSA's disability determination decisions.² We use that model to simulate the pool of persons medically eligible for SSI or DI among nonbeneficiaries in the general population.

There are five steps in establishing the medical eligibility of disability applicants, as

¹ The take-up rate, also called the participation rate, is the percentage of program eligibles who take benefits.

² A major new data collection effort by SSA—the National Study of Health and Activity (NSHA)—will also combine self-reports with medical examinations. Advantageous features of the NSHA are noted in the conclusions.

implemented by state Disability Determination Services (DDSs). Those steps are illustrated in Chart 1. Steps 1 through 3 are screens: the first is on earnings and the next two are medical. Applicants are denied benefits at step 1 if they earn more than the maximum substantial gainful activity (SGA) amount—\$500 per month during the period represented by the data (late 1991 to early 1992). Activities are considered “substantial” if they involve significant physical or mental activities and “gainful” if done for pay or profit. In step 2, impairments are assessed to determine whether they are severe. If not, the applicant is denied. The severity test is based on the ability to perform common work-related activities such as walking, lifting, seeing, speaking, and understanding simple instructions. A duration test is also used, typically at step 2. The duration test requires that impairments have lasted or are expected to last at least 12 months or that the impairment is expected to result in death. Applicants are allowed on the rolls at step 3 if the impairment satisfies codified clinical criteria called the Listings of Impairments. Applicants not allowed at step 3 are severely impaired, but their impairments do not “meet the listings.” Such applicants are evaluated at the last two steps of the determination process, which involve an assessment of their residual capacity to work. At step 4, those found able to perform their past work are denied. After step 4, remaining applicants are allowed in step 5 if they are found unable to do any work in the economy; otherwise, they are denied. For a more detailed description of the process, see Lahiri, Vaughan, and Wixon (1995).

Because we are focusing on medical eligibility, we ignore the first step of the process and model the last 4 steps, which, following convention, we refer to as steps 2 through 5 of the

determination process.³ In this paper we simulate neither step 1, the SGA test, nor the broader financial criteria for the programs, although eligibility associated with these criteria will be estimated in subsequent work. However, a more fundamental reason for not simulating the SGA test as an integral step in the process is the contingent event of interest to policymakers. More specifically, one policy objective is to estimate the potential program growth resulting from economic or household events, such as loss of a job by the sample member or a spouse. To do that, one must estimate the medical and financial elements of eligibility independently, to permit estimation of the number of working disabled who would be eligible if they lost their jobs.

The steps of the decision process have distinct criteria. For that reason, Hu and others (1997) modeled each step separately and then linked them sequentially, reflecting the structure of the administrative process. Health plays an important role in steps 2 and 3, while occupational and demographic characteristics dominate later (conditional on having passed the health screens). Lahiri, Vaughan, and Wixon (1995) and Hu and others (1997) showed that reduced form models that evaluate the final allow/deny decision as a single decision are not as informative in that they downplay the role of factors that demonstrably influence decisions at particular steps. For example, variables such as activities of daily living (ADLs), mental conditions, age, education, and skill level were found to be major factors in the four-step structural model but not relevant in the one-equation reduced-form model. The reason appears to

³ “Medically eligible,” as used here, refers to eligibility under steps 2 through 5 of the sequential determination process, even though the process also involves vocational and demographic criteria in some steps. That is, the phrase refers to the nonfinancial elements of eligibility.

be that they are important in certain steps of the process but not in others. Following Hu and others (1997), we use the sequential model to estimate the factors that determine medical eligibility for the pool of applicants. We use those estimates of conditional probabilities to simulate eligibility at each step of the determination process for the general population.

III. Methodology

The Disability Determination Model

Hu and others (1997) modeled steps 2 through 5 of the disability determination process using SSA administrative records on disability determinations matched to four waves of the 1990 SIPP. They estimated effects of such factors as health conditions, job characteristics and worker skills, district office and state agency differences, and demographic traits at each step. We use those estimates, derived on the basis of the actual experience of applicants, to simulate the eligibility status of a sample of persons representing nonbeneficiaries in the general population.⁴ The four decision nodes of the determination process, shown in Chart 2, result in five outcomes, as follows:

d_2 = denial at step 2 based on nonseverity of medical impairment(s),

a_3 = allowance at step 3 based on listed impairment(s),

d_4 = denial at step 4 based on residual capacity for past work,

a_5 = allowance at step 5 based on residual incapacity for any work in the economy, and

⁴ For the period represented by the data, approximately one-quarter of all allowances were based on appeals beyond the DDS level. Our analysis is based on a model of DDS decisions, including DDS reconsiderations; that is, the estimates we report are those implied by DDS medical standards.

d_5 = denial at step 5 based on residual capacity for work in the economy.

Each outcome at nodes k , l , m , and n takes a value of one if the favorable outcome from the standpoint of the applicant is realized, that is, an allowance or pass on to the next step. We model the probability of a denial at the second step as follows:

$$\Pr(d_2) = P_{k=0} = 1 - F(\alpha' W_k),$$

where $P_{k=0}$ is the probability of denial at step 2 based on a logit regression, W_k is the vector of explanatory variables, and α is the parameter vector to be estimated.

Similarly,

$$\Pr(a_3) = P_{l=1} = P_{l=1|k=1} \cdot P_{k=1} = F(\beta' X_l)F(\alpha' W_k),$$

where $P_{l=1|k=1}$ is the probability of allowance at step 3, conditional on not being denied at step 2 (node k), X_l is the vector of explanatory variables for step 3, and β is the parameter vector to be estimated for step 3.

Likewise, we represent equations for the remaining decisions as follows:

$$\begin{aligned} \Pr(d_4) &= P_{m=0} = P_{m=0|k=1,l=0} \cdot P_{l=0|k=1} \cdot P_{k=1} \\ &= [1 - F(\gamma' Y_m)][1 - F(\beta' X_l)]F(\alpha' W_k), \end{aligned}$$

$$\begin{aligned} \Pr(a_5) &= P_{n=1} = P_{n=1|k=1,l=0,m=1} \cdot P_{m=1|k=1,l=0} \cdot P_{l=0|k=1} \cdot P_{k=1} \\ &= F(\delta' Z_n)F(\gamma' Y_m)[1 - F(\beta' X_l)]F(\alpha' W_k), \end{aligned}$$

$$\begin{aligned} \Pr(d_5) &= P_{n=0} = P_{n=0|k=1,l=0,m=1} \cdot P_{m=1|k=1,l=0} \cdot P_{l=0|k=1} \cdot P_{k=1} \\ &= [1 - F(\delta' Z_n)]F(\gamma' Y_m)[1 - F(\beta' X_l)]F(\alpha' W_k), \end{aligned}$$

where $P_{m=0|k=1,l=0}$ is the probability of denial at step 4 (node m) conditional on being passed on at step 2 and at step 3. At step 4, γ is the parameter vector to be estimated, and Y_m are the explanatory variables. At the last step (node n), $P_{n=1|k=1,l=0,m=1}$ is the probability of allowance conditional on not being denied at step 2, not being allowed at step 3, and not being denied at step 4. Here, δ is the parameter vector to be estimated and Z_n are the explanatory variables. Analogously, $P_{n=0|k=1,l=0,m=1}$ represents the probability of a denial at the last step (node n). Individuals can be allowed at steps 3 or 5 so that the overall allowance probability using the conditional probabilities is calculated with the following formula:

$$\begin{aligned} \Pr(a_3) + \Pr(a_5) &= P_{l=1} + P_{n=1} \\ &= F(\beta' X_l)F(\alpha' W_k) + F(\delta' Z_n)F(\gamma' Y_m)[1 - F(\beta' X_l)]F(\alpha' W_k). \end{aligned}$$

Parameter vectors, α , β , γ , and δ , are estimated sequentially over surviving subsamples using logit regressions. Those estimates are then used to simulate the number of eligibles in our sample of nonbeneficiaries. Each candidate is assigned conditional probabilities for surviving each decision node as well as an overall unconditional probability.⁵

A number of methodological issues arise when making predictions for the general population. For example, should we simulate eligibility for all nonbeneficiaries or only for those who report a health problem or work limitation? How much of a difference does it make to

⁵ Estimates at each step are conditional on having survived the previous node; therefore, using these estimates produces conditional probabilities. To determine the overall probability of allowance, we therefore calculate the unconditional probability from the probabilities at each step of the process. See Hu and others (1997) for details.

assume zero probabilities for those who report no health problems? Do we have a sample selection concern because we are simulating probabilities of allowance for nonbeneficiaries based on estimates for a group of applicants? Once we have a conditional probability of allowance for everyone in our sample, how do we define a cutoff so as to estimate a population of eligible individuals? These issues and others regarding the data are addressed below.

The Data and the Sample

We use data from waves 2, 3, 6, and 7 of the 1990 SIPP panel to develop a sample representing nonbeneficiaries in the general population. Our sample consists of 25,525 men and women between the ages of 18 and 64 (during wave 7 of the 1990 panel) who responded in all four waves and for whom there is a successful match to the SSA Summary Earnings Record (78 percent of the wave 7 core public-use file for January-April, 1992).⁶ The wave 3 and 6 interviews include modules covering work limitations, functional status, Activities of Daily Living/Instrumental Activities of Daily Living (ADL/IADLs), and mental and physical health conditions.

We have administrative records on disability determinations only for applicants. To use estimates from Hu and others (1997), we need reliable, survey-based proxies for a few

⁶ The earnings restriction was introduced to simplify both missing data problems and simulation applications. For example, the restriction permits us to define a sample of persons whose fully insured status and disability insured status can be estimated based on past earnings. However, the restriction might affect the representativeness of the remaining sample. A probit of attrition (0/1 included in our sample or not) based on unweighted sample cases suggests that the restricted sample per se may underrepresent persons with functional limitations. To consider that possibility, we employed a public-use weight that was adjusted to closely reproduce the corresponding public-use file population estimates by age. The resulting distributions by detailed health, work disability, and functional status closely agree with SIPP-based estimates published by McNeil (1993) for approximately the same period. The weights are adjusted to represent the civilian noninstitutional population as of early 1992.

administrative variables used as independent variables in that study, most of which are from the SIPP (including all of the health variables). For the remaining variables, in many cases we use substitutes in both our reestimation of the disability determination model and our eligibility predictions for the nonbeneficiary population. In a few cases we dropped variables. As a result of these changes, our parameter estimates are slightly different from those in Lahiri, Vaughan, and Wixon (1995) and Hu and others (1997). The details of our choice of variables and reasons for selecting them are described in Appendix A. The parameter estimates used for the prediction appear in Appendix B.

Should we restrict the sample to persons most likely to be found medically eligible—to a subpopulation we consider most “at-risk”? For example, we expect that the medically eligible will be drawn from those with some kind of health problem. We also expect that, if our model performs well, only persons with health problems will be estimated to be eligible, permitting us to use the full sample to make predictions, regardless of the health status of the respondents. However, such model performance hinges on the ability to accurately assess disability using survey data. Unfortunately, true health status is a latent variable and survey measures do not perfectly reflect disability under SSA’s definition. Since health is the driving force in these models, any weakness in the health measures will have a major impact on the model’s performance. More specifically, because the survey data do not measure severity as accurately as we might like, differences in functional capacity between nonapplicants and allowed applicants may be underestimated.

In light of such concerns about survey health measures, we explicitly consider how well the model performs with respect to people with no health problems by testing alternative

approaches. We run the simulation on the full sample, and then, for the sake of comparison, on a restricted sample. Under the full-sample approach, even if a survey respondent is not likely to be medically eligible, we permit the model to decide. We then compare those results with estimates for a restricted sample. Under the restricted-sample approach, a probability of zero is assigned to those respondents who report no health problems. If the model performs well, the results under the full sample and the restricted sample should be similar.

Under our restricted-sample approach, we limit the sample to those who report at least one health problem, because a sample defined in that way is likely to capture potential applicants. We then estimate eligibility among sample members with health problems. This raises a concern about how to restrict the sample, particularly since the restriction must be based on the imperfectly measured health variables. We choose the least restrictive criterion—having at least one health problem reported in any of the four waves.

Correcting for Selectivity

We use a sample of applicants who have gone through the disability determination process between 1989 and 1993 to make predictions for nonbeneficiaries in the general population.⁷ However, those who apply for the programs may not be a random sample of the disabled in the population. In other words, disabled candidates may self-select into the applicant

⁷ See Hu and others (1997) or Lahiri, Vaughan, and Wixon (1995) for a complete description of the administrative data used in the analysis.

pool based on their own knowledge of the severity of their disabilities which, in turn, informs their expectations about the outcome of the decision process.

Moreover, such self-selection may occur in ways that are unobservable to the analyst. For example, applicants with severe impairments may appear similar to nonapplicants with milder limitations based on self-reports. If that is the case, survey data may not permit us to observe the true range of variation in severity. Our sample of applicants—for reasons unobservable in the data—would therefore be more likely to be eligible than nonapplicants with similar observed characteristics, causing us to overestimate nonapplicant eligibles. Fortunately, we have information from SSA administrative records that identifies applicants and permits us to adjust for selectivity.

While severity of impairment drives both application decisions and eligibility decisions, opportunity costs also affect applications. It would take a less severe impairment to induce someone with lower opportunity costs to apply, but self-reported health indicators may not pick up such differences in severity. If so, then some groups with higher incentives to apply may do so with less severe impairments, but the severity differentials may not be fully observable. In fact, Hu and others (1997) find socioeconomic factors influential in step 2 of the determination process. Moreover, step 2 is the only step at which such socioeconomic factors are observed to have unexpected effects. Because step 2 is a medical screen intended to filter out persons with less severe impairments, economic factors should not influence the determination outcome at that step, although they would affect incentives to apply. This suggests that something unobserved that is correlated with underlying economic status is left out. Controlling for these economic factors by linking a model of the disability decision with a model of the decision to

apply—even in a preliminary way—would pick up some of the unobserved differences in severity.

We model step 2 of the determination process and the decision to apply simultaneously as follows:

$$z_{i1} = \beta'_1 X_{i1} + \varepsilon_{i1}, \quad y_{i1} = \begin{cases} 1 & \text{if } z_{i1} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$z_{i2} = \beta'_2 X_{i2} + \varepsilon_{i2}, \quad y_{i2} = \begin{cases} 1 & \text{if } z_{i2} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$(\varepsilon_{i1}, \varepsilon_{i2}) \sim \text{BVN}(0, 0, 1, 1, \lambda)$$

where z_{i1} = the propensity to be passed on at step 2 (latent),

X_{i1} = factors that determine eligibility at step 2,

y_{i1} = observed indicator of eligibility at step 2,

z_{i2} = the propensity to apply for disability (latent),

X_{i2} = factors explaining the decision to apply,

y_{i2} = observed indicator measuring the decision to apply,

BVN = the bivariate normal distribution, and

λ = the correlation between the decision to apply and eligibility at step 2.

The equations are estimated simultaneously to allow for the presence of λ , making the estimates more efficient (than a two-step sample-selection model). Without selection we would

observe y_{i1} and y_{i2} for everyone, so that the log likelihood function would consist of four sets of probabilities for the four possible outcomes.⁸

In the present analysis, because there is selection, we observe y_1 only if $y_2 = 1$. The log likelihood for the bivariate probit model with selection is:

$$\begin{aligned} \log L = & \sum_{y_2=1, y_1=1} \Phi_2[\beta_1 X_{i1}, \beta_2 X_{i2}, \lambda] + && \text{(allowed applicant)} \\ & \sum_{y_2=1, y_1=0} \Phi_2[-\beta_1 X_{i1}, \beta_2 X_{i2}, -\lambda] + && \text{(denied applicant)} \\ & \sum_{y_2=0} \Phi_1[-\beta_2 X_{i2}], && \text{(non - applicant)} \end{aligned}$$

where Φ_2 is a bivariate standard normal cumulative density function (CDF) and Φ_1 is a univariate standard normal CDF. These relationships account for whether or not the respondent applies and how that decision factors into the initial medical screen of the determination process. Our model controlling for sample selection uses this bivariate probit specification for step 2 and univariate probits for the remaining three steps.

Defining a Pool of Eligibles: Alternative Methods

Policymakers are interested not only in the size of the eligible pool but also its characteristics. That is, they want to know which subpopulations are targeted by current (or proposed alternative) program criteria. We considered several techniques for counting and identifying eligibles, given the probability of allowance estimated for each sample member.

⁸ The four outcomes are: allowed applicant ($y_2=1, y_1=1$), denied applicant ($y_2=1, y_1=0$), eligible nonapplicant ($y_2=0, y_1=1$), and ineligible nonapplicant ($y_2=0, y_1=0$).

One technique is to sum the weighted probabilities predicted for sample respondents. That method gives us a count but does not assign eligibility status to each individual and, hence, does not result in a discrete pool of eligibles that could be conveniently described for policymakers.⁹ To assign individual eligibility status, we categorize respondents as eligible or ineligible in two ways. First, we use the random number generator approach recommended by Giannarelli and Young (1992). To determine eligibility we draw a number between zero and one from the uniform distribution for each respondent. The respondent is eligible if the random number is less than or equal to the respondent's eligibility probability; otherwise, he or she is ineligible. The expected number and composition of eligibles yielded by this approach are designed to approximately reproduce results obtained by summing probabilities, except that it allows us to categorize each individual as either eligible or ineligible.

Another common approach involves using 0.5 as a cutoff, designating an individual as eligible if his or her predicted probability exceeds 0.5 and otherwise designating the individual as ineligible. In many contexts, the use of a 0.5 cutoff makes sense in that a probability of 0.5 represents the point at which an event is equally likely to occur or not occur. In the present case, however, the distribution of probabilities is not centered at 0.5, so arbitrarily picking 0.5 has no empirical basis. Instead, we use information available from the distribution of applicants' allowance probabilities to determine the cutoff. We follow other researchers (Hosmer and Lemeshow 1989; Lemeshow and others 1988) by using a weighted average of allowance probabilities for both allowed and denied applicants as a cutoff. Cramer (1997) evaluates effects

⁹ Technically, if probabilities can be summed to estimate the number of eligibles, then the probabilities can also be used as weights and the pool can be described. However, this approach is somewhat complex for routine use.

of that approach in dealing with an unbalanced sample.¹⁰ Simulations using logits on unbalanced samples tend to underestimate those in the nondominant group. In our case, that would imply a slight underestimate of eligibility because there are more denied applicants than allowed applicants. We refer to this approach as a dynamic cutoff in that the cutoff varies with model specification and sample. Since there is no established cutoff methodology, we produce several estimates of the medically eligible pool, allowing tests for robustness among a range of estimates.

IV. Classification Rates for the “Prevented” Measure: A Benchmark

Is a complex, multivariate approach necessary? It might be instructive to consider how well one can estimate medical eligibility by using a single question: “Does [your] health or condition prevent [you] from working . . .?” For lack of an empirically estimated alternative, the response to that question is sometimes used to estimate medical eligibility (Haveman and others, 1994; Benitez-Silva and others 1999). Intuitively, it is reasonable to expect that being “prevented from working” might be linked to the programmatic criterion of being work disabled. Our framework, which involves a sample of individuals for whom we know both survey responses and application outcomes, allows us to consider the extent to which this question permits successful classification of applicants.¹¹

¹⁰ Unbalanced samples are those that are classified into a 0/1 category with unequal groupings. For example, in our sample of applicants, more are denied than allowed.

¹¹ Associated with the “prevented” measure is a problem of endogeneity—a problem that limits the usefulness of the measure. Even if the measure classifies applicants well, it would not be useful in identifying those in the general population who, despite serious impairments, continue to work.

We find that only 55.6 percent of applicants are correctly classified based on the “prevented” measure.^{12,13} These findings suggest the limits of the single-variable “prevented” measure. They also serve as a caution, we think, against relying on intuition in defining medical eligibility based on survey self-reports. We use these findings as a benchmark in evaluating our multivariate approach.

V. Results

Correcting for Sample Selectivity

As expected, applicants report more health problems, significantly worse general health, less education, and lower earnings than nonapplicants (see Table 1). These results support our expectation that applicants are substantially different from nonapplicants. However, Table 1 also illustrates that a number of survey self-reports on health may have considerable explanatory power in distinguishing people with serious impairments. That finding alleviates some of the concern, discussed earlier, about using survey self-reports. Nonetheless, those reports may not reflect the full extent of differences in severity between applicants and nonapplicants.

The results of our simulations using the alternative methodologies appear in Table 2. (See Appendix B for regression estimates used in the simulation.) We compare models with and without sample selection, using both the full sample and a restricted sample (people reporting health problems) and alternative cutoff procedures. The estimates of eligibility range from under

¹² The classification success rates were well balanced between allowed applicants (53.0 percent) and denied applicants (57.7 percent). The prevented measure used here is from wave 7 of the 1990 SIPP panel.

¹³ The success rate for the broader “limited” measure—based on whether the respondent’s health or condition limits the kind or amount of work—was lower (48.9 percent).

1 percent of the nonbeneficiary population to 32 percent. We consider neither of those extreme values to be credible, but they do allow us to evaluate alternative methodologies. For example, given that the mean probability of allowance for applicants falls below 0.5, a cutoff at that value seriously underestimates eligibility (0.6 percent). The ceiling of 32.2 percent is high because of the problems associated with sample selection and measurement error in health variables. Hence, a credible range of values would be much narrower, for reasons discussed below.

When we adjust for sample selection by modeling eligibility and application behavior simultaneously, we find a strong correlation across equations, suggesting the presence of selectivity bias. (See Appendix B for results from this analysis.) In comparing results from Table 2 with and without selection controls, the dramatic differences suggest that the model with no sample-selection control is exaggerating eligibility, qualifying many people for reasons other than health.

Even the crude adjustment for selection used here has a substantial effect on this problem (see Charts 3 and 4, in which allowance probabilities are plotted for the full sample by health status for the models with and without sample selection). With sample-selection, as shown in Chart 3, both distributions are centered on 0.2, but the distribution is wider for respondents with work limitations. Only those with work limitations would have probabilities higher than an illustrative cutoff value of 0.4, for example. By contrast, Chart 4 demonstrates that without sample selection the probabilities are centered near 0.4 for both groups. That result is not surprising because we expect exaggerated probabilities without sample-selection controls.¹⁴

¹⁴ In fact, sample selection reduced the mean observed probability from .334 to .191 for the sample as a whole.

Moreover, in the model with no sample-selection controls, the probabilities of many sample members both with and without work limitations would exceed cutoff values in the middle of the distribution—at 0.4, for example. For such cutoff values, sample-selection controls permit us to distinguish much more effectively between people with work limitations and those without.

Charts 5 and 6 illustrate how sample selection alters the distribution of allowance probabilities for the restricted sample. Again we see that the model with sample selection does a better job of identifying people with severe health limitations. More specifically, Chart 5 shows that a probability cutoff of 0.4 would distinguish those with work limitations and high allowance probabilities (those most severely impaired) from those with no work limitations. And, as with the full sample, that distinction cannot be drawn as efficiently without sample-selection (see Chart 6).

Note that the contrasts in the distributions are not as pronounced for members of the restricted sample (Charts 5 and 6) as for members of the full sample (Charts 3 and 4). That is because there is less variation in health among members of the restricted sample. Still, the presence of a work limitation seems to be more correlated with allowance probabilities for the model with sample selection than for the model without. Hence, the sample-selection model performs consistently well across the two samples.

Table 3 makes a similar point about the sample selection model; specifically, it offers evidence on how well sample-selection estimates target by health status under our preferred specification. As discussed below, our preferred method uses sample-selection correction, the dynamic cutoff, and the full sample. Using that specification, we find that sample selection performs better in identifying those with health problems as eligible. Furthermore, when we

correct for selectivity, the differences in the frequency of health problems between eligibles and ineligibles are substantial.

Table 4 reports classification success rates for models with and without sample selection. That is, it reports how well our estimation methodologies classify a group of applicants by comparing our eligible/ineligible estimates with the allow/deny findings of the DDS agencies. We assume the DDS findings to be correct; that is, any discrepancies are assumed to result from errors in our estimates. Even our most accurate methodologies misclassify about 30 percent of applicants.

What accounts for these errors? First, the process we simulate—DDS determinations— involves considerable scope for judgment; in fact, many determinations are eventually overturned. Second, the survey data involve errors of various types. Timing discrepancies may explain inconsistencies between survey and administrative data in some cases. Also, because survey questions on health often call for self-evaluation, two respondents with identical symptoms may characterize their health differently. For example, a small number of disability applicants report no health problems in the survey. In addition, little is known about the survey responses of those with mental impairments. An advantage of our approach is that we juxtapose the survey responses and DDS findings for a group of applicants, permitting us to analyze how to best interpret and use the survey information. Finally, errors in model specification will also contribute to observed misclassifications. For example, future improvements in selection controls would probably improve the estimates reported here.

As indicated in Table 4, the sample-selection models are better able to identify the eligibility status of applicants, with overall success rates from 60 percent to 68 percent, whereas

models without selection controls predict less accurately under most methodologies. In particular, the sample-selection model—when used with the full sample and the dynamic cutoff—has the highest overall classification success rate (68.4 percent).¹⁵

We see other evidence that the sample-selection model does a better job of identifying potential eligibles when we compare results for the full and the restricted samples. If the models are performing well in identifying persons who meet SSA’s criteria, there should be no significant and unexplained differences in the estimates of eligibility when using the full sample, as opposed to the restricted sample.¹⁶ As Table 2 illustrates, that is true only for the sample-selection model and, furthermore, only for certain cutoff methods. Without selection, the results differ considerably under all cutoff methods. For example, using a cutoff at the weighted mean for applicants, 24.3 percent are eligible for the full-sample model, as opposed to 10.1 percent for the restricted-sample model. In contrast, the same estimates are 2.9 percent and 2.7 percent, respectively, for the model with sample selection.

Cutoff Methods

Having decided to rely on the sample-selection model, which cutoff procedure is preferable? Both theory and empirical results factor into our decision. First, there is a theoretical justification for the use of cutoffs because there are thresholds implicit in the decision process at each step of the disability determination. In step 2, for example, adjudicators establish

¹⁵ The classification success rate is defined as the percentage of predicted outcomes (eligible/ineligible) that agree with actual DDS decisions (allow/deny) for applicants in our sample.

¹⁶ The estimates from models without sample selection should be similar for the full and restricted samples if the model does a good job of assigning eligibility probabilities. The same should be true for sample-selection models.

whether the impairment is severe and at step 3 they determine whether the impairment qualifies under a more exacting standard, the listings.

As shown in Table 2, the use of a full or restricted sample makes a considerable difference in the estimates, but only under some cutoff methods. Two cutoff methods that use sample selection—summing the probabilities and the random number generator cutoff—yield quite different results between the full and restricted samples. By contrast, the two remaining cutoff methods—the dynamic cutoff and the 0.5 cutoff—show almost no difference between sample-selection estimates under the full and restricted samples. What accounts for these differences in performance under alternative cutoff methods?

Our disability determination model assigns low probabilities of allowance to some sample members, such as those with impairments of marginal severity. Low probabilities may also be assigned to sample members who are healthy, relative to other members, but who have key demographic characteristics. The alternative cutoff methods differ in their treatment of sample members with low probabilities. For two methods—summing the probabilities or using the random number cutoff—the distribution of medical eligibles by health replicates the distribution of the underlying probabilities. Under these two methods, some sample members with low probabilities are estimated as eligible. But summing low probabilities for many people can make a substantial difference in the estimates. Likewise, drawing randomly from a uniform distribution causes some respondents with low probabilities to be eligible, given that the sample includes members with low probabilities.

By contrast, the dynamic cutoff and the 0.5 cutoff do not replicate the distribution of the underlying probabilities; rather, they censor low-probability sample members through the use of a threshold. If most sample members in relatively good health have low probabilities, then these cutoff methods will limit eligibility primarily to sample members with health problems. In fact, under our preferred specification for the dynamic cutoff, over 98 percent of those estimated as eligible have a health problem.¹⁷ Hence, the dynamic cutoff, like the somewhat arbitrary sample restriction, ensures that those estimated as eligible are drawn from those who report a health problem.

However, in comparing estimates produced by using the dynamic cutoff and the sample restriction (linked to a model that replicates the distribution of the probabilities), there is an important difference—a difference that strengthens the empirical case for cutoffs. Table 5 juxtaposes estimates from our preferred model (the dynamic cutoff, see column 1) and a model using the restricted sample and the random number generator (column 2). Despite the restricted sample, the “alternative” estimate is more than twice as large as the preferred estimate, because the preferred estimate censors not only those with no health problem, but also those with probabilities below the dynamic cutoff—mainly those who are less sick or less impaired. So we expect the dynamic cutoff estimate to include not only a smaller group of eligibles but also a

¹⁷ Future work should include an investigation of the low-probability observations censored by this approach. For such observations, the appropriate cutoff method depends on the nature of the error. For example, if some respondents seriously understate the severity of their impairments in surveys, then a portion of them should be included as eligible (although trying to select the correct observations may lower success rates). That may occur for some respondents with mental conditions, for example. For such groups, a sample restriction may also censor inappropriately. On the other hand, to the extent that low-probability cases occur as the result of timing discrepancies—the respondent was healthy at the time of the survey but was seriously impaired months later at the time of application—censoring is appropriate.

more impaired group. Table 5 confirms that expectation. Table 5 also permits us to compare the health profile of eligibles with that for applicants allowed by DDSs (see columns 1 and 3). In terms of standard survey health measures, eligibles under our preferred method are quite similar to allowed applicants, while eligibles under the alternative estimate are substantially less impaired.

Moreover, of the methods we use that censor low probabilities, the success rates in Table 4 are highest for the dynamic cutoff. The frequently used 0.5 cutoff does a poor job of identifying allowed applicants (success rates from 28 percent to 30 percent under the sample-selection model) and therefore may underestimate eligibility in the general population. By comparison, a random number generator cutoff does a better job of identifying allowed applicants (success rates from 37 percent to 40 percent) but a worse job of identifying denied applicants. When comparing sample-selection models, the dynamic cutoff not only yields the highest overall success rates but also has the best balance between identifying allowed and denied applicants (65 percent to 70 percent of each).¹⁸ Moreover, the overall rate compares favorably with that for the “prevented” measure (55.6 percent).

To further demonstrate why we prefer the dynamic cutoff with the sample-selection model, let us consider the distribution of allowance probabilities for nonapplicants and for applicants as shown in Chart 7. For nonapplicants, the distribution of allowance probabilities is

¹⁸ As an alternative, one could maximize the denied success rate, on grounds that it is important to correctly classify millions of nonapplicants in the general population. However, a methodology that does well in terms of classifying those denied may not classify allowed applicants well. This may result in an underestimate of eligibles, as it probably did under the 0.5 cutoff. Conversely, maximizing the allowed success rate could result in overestimates of eligibility. We focused on the overall rate as a compromise. However, these issues should be revisited when NSHA data permit us to calibrate model estimates using direct measures of eligibility.

centered at 0.2 and there is little variation, whereas for the applicant pool the distribution is much more uniform. That suggests that a cutoff at 0.5 might be too restrictive and, as the success rates show (Table 4), would miss most allowed applicants. A cutoff based on the weighted mean for applicants captures more of the right tail of the probability distribution for applicants, where allowed applicants are likely to be concentrated. That is, it captures applicants with the highest allowance probabilities who we expect to be in the poorest health. This suggests why the dynamic cutoff yields the highest success rate for allowed applicants.

Policy Findings

Given this analysis, our estimates suggest that 4.4 million people, or 2.9 percent of the nonbeneficiary population aged 18-64, would meet SSA's medical criteria for disability. Of that group, 3 million (or 2.0 percent of the population studied) had average earnings below the maximum SGA amount (\$500 per month) in the prior year; that is, they are estimated to be eligible in terms of both SSA's medical criteria and the SGA test.¹⁹ The balance, 1.4 million, are medically eligible but have average earnings above the SGA limit. Some of the latter group would end up on the rolls in the event of a recession.

How do those we estimate to be medically eligible compare to ineligible? Based on conventional survey health measures, eligibles are much more impaired and much more frequently work limited (see Table 6). Moreover, they are more likely to be older, unmarried,

¹⁹ Applying the financial criteria for the two programs—disability insurance requirements under DI or the SSI income and asset limits—would further limit the size of the eligible pool. Those criteria will be implemented in later work.

less educated, poor, and low earners. However, the estimates also suggest that more than one-third of eligibles have a mental condition. That may be a vestige of past policies on deinstitutionalization; certainly, it highlights the importance of current policies on medical care and encouraging work for that subgroup.

VI. Conclusions

The purpose of this paper is to develop an approach for tracking medical eligibility for SSA's disability programs. Using a structural model of the disability determination process estimated on a sample of applicants, we make out-of-sample predictions of eligibility for nonbeneficiaries in the general population. We use several methods to develop a range of estimates of the number who would be found eligible if they applied. Our estimates range roughly from 1 percent to 32 percent of the general population, suggesting the importance of addressing underlying methodological issues so as to narrow the range.

The highest estimate (32 percent) comes from a simulation using estimates from a model with no sample selection, no sample restriction, and summed probabilities. Those estimates assign eligibility to many respondents who report no health impairments. That assignment partly reflects limitations in survey health measures, specifically, that self-reports do not depict sufficient variability in health. As a result, health may become less important in the model, relative to socioeconomic characteristics. The problem represented by respondents with low probabilities of allowance is pervasive and is addressed in different ways by the methodologies tested here. The underlying objective is to design a methodology that best deals with limitations in survey self-reports relating to health.

On the one hand, we adjusted for sample selectivity by simultaneously modeling the decision to apply and the medical eligibility decision. Such a model controls for unobserved differences in the severity of health among applicants and nonapplicants. Our preliminary attempt at such a model was successful in terms of predictive power, and the correlation coefficient representing sample selection is significant. A more extensive model of the decision to apply may substantially improve those eligibility estimates.

We also considered how to define a discrete pool of eligibles on the basis of the probabilities estimated for members of the general population. The main rationale for converting a continuous variable to a binary eligible/ineligible code is to facilitate describing eligibles. Policymakers are interested not only in the *number eligible* but also in the *subpopulations targeted* under program criteria. Of the methods tested, the frequently used 0.5 cutoff does a poor job of classifying allowed applicants, but both summing the probabilities and the random number generator result in too many with low probabilities (and relatively good health) being estimated as eligible. The dynamic cutoff, which employs a cutoff specific to the distribution of applicants, represents our preferred approach for several reasons.

First, the dynamic cutoff, when used with sample selection and the full sample, ensures that those with no health problems are estimated to be ineligible. The sample restriction achieves the same result but involves some arbitrariness in defining a health problem. Second, we prefer this estimate because it yields a group of eligibles whose health characteristics closely approximate those of allowed applicants. Finally, our preferred estimate gives the highest overall classification success rate for applicants, 68.4 percent. That result represents a clear

improvement over the rate for the conventional single-variable model based on the “prevented” measure (55.6 percent).

Using our preferred estimate, we find that 4.4 million persons (2.9 percent of the general population not receiving disability benefits and aged 18-64 in 1991) satisfy SSA’s definition of disability. Of those, 3.0 million meet SSA’s SGA test and 1.4 million have earnings exceeding the SGA. The size of those groups suggests that there is substantial potential for future program growth. In addition, it underscores the importance of studying incentives to apply for disability as well as policy options that alter incentives, such as early intervention and workplace accommodation.

Future work should include several topics. First, the characteristics of those we estimate to be medically eligible should be examined in detail. That study would investigate how subpopulations of special interest are affected by program criteria; in addition, it would serve as a baseline for similar estimates using future data or alternative eligibility criteria. Second, a number of methodological advances can be made. We expect to estimate a more detailed model of the decision to apply for benefits, allowing refinement of the preliminary sample-selection adjustment employed in this paper. Also, sample members with no health problems or, more generally, with low estimated probabilities of allowance should be analyzed further. In the long run, SSA’s new National Study of Health and Activity will offer several avenues for improvement, including medical and functional examinations for nonapplicants, simultaneous administration of surveys and medical exams, survey questions tailored to measurement of medical eligibility, and an opportunity to calibrate survey-based estimates.

Taken together, these methodological improvements should advance the effort to close a long-standing information gap—the inability to credibly estimate the number of persons who are medically eligible for disability benefits. In the context of that effort, the short-term contribution of this study is to offer a baseline eligibility estimate developed through a framework that assesses survey self-reports in the light of SSA’s medical evaluations of the respondents. Perhaps our long-run contribution is the conceptual framework we provide for benchmarking future methodological advances in estimating medical eligibility.

Table 1.
Frequencies of selected characteristics for applicants and nonapplicants (in percent unless otherwise indicated).

Characteristic	Applicants	Nonapplicants
Total population aged 18-64 (thousands)	5,174	149,479
Health		
Health status fair or poor	58.5	8.4
Work limitation status (wave 7, total)	100.0	100.0
Not limited	25.6	89.0
Limited but not prevented	27.4	7.0
Prevented	47.0	4.0
Functional limitations (total)	100.0	100.0
None	40.9	90.0
One or more limitations	59.1	10.0
One or more severe limitations	31.5	3.0
One or more ADL	19.5	1.6
One or more IADL	26.8	2.6
Medical conditions	100.0	100.0
Mental	14.3	2.8
Musculoskeletal	27.2	4.3
Neurological, sensory	6.0	1.0
Cardiovascular, respiratory	10.6	1.5
Other	12.7	1.5
None reported	28.9	88.8
Demographic and financial		
Age (total)	100.0	100.0
18-44	47.6	69.6
45-54	26.2	17.1
55-64	26.2	13.3
Educational attainment (in years, total)	100.0	100.0
Less than 12	41.2	14.8
12 or more	58.8	85.2
Average monthly earnings (total) ^a	100.0	100.0
Greater than \$500	26.2	63.6
\$1 to \$500	22.2	15.1
Zero	51.7	21.3

SOURCE: SSA/OP/ORES/DER/Disability Modeling Group, January 20, 2000.

NOTE: Data are for early 1992, based on the 1990 Survey of Income and Program Participation.

^a Based on Social Security Administration earnings data for 1991.

Table 2.
Results of simulating medical eligibility for a sample representing the general population aged 18-64 (estimates weighted).

	No sample selection		Sample selection	
	Full sample	Restricted sample	Full sample	Restricted sample
Summing the probabilities				
Number eligible (thousands)	48,194	16,424	27,366	9,575
Percentage of general population	32.2	11.0	18.3	6.4
Random number generator cutoff				
Number eligible (thousands)	47,334	16,284	26,511	9,655
Percentage of general population	31.6	10.9	17.7	6.4
Cutoff at weighted mean for applicants ^a				
Number eligible (thousands)	36,329	15,081	4,393	4,020
Percentage of general population	24.3	10.1	2.9	2.7
Cutoff at 0.5				
Number eligible (thousands)	15,054	8,096	1,029	880
Percentage of general population	10.1	5.4	0.7	0.6

SOURCE: SSA/OP/ORES/DER/Disability Modeling Group, February 3, 2000.

NOTE: The sample excludes those receiving disability benefits. The data are for early 1992 and are based on the 1990 Survey of Income and Program Participation.

^aThe cutoffs used were the mean probabilities of allowance, weighted by proportion allowed versus denied. The numerical values for the cutoffs were .411 for estimates with no sample selection and .333 for those with selection.

Table 3.
Frequencies of selected health measures by eligibility status (percent)

Health Measures	No sample selection		Sample selection	
	Eligible	Ineligible	Eligible	Ineligible
Work limitations				
Limited (including prevented)	16.1	8.8	67.1	8.9
Prevented	6.0	2.1	37.1	2.0
One or more functional limitations	15.5	7.9	51.5	8.4
One or more severe functional limitations	5.1	2.0	28.5	2.0
One or more ADL limitations	2.9	0.9	18.3	0.9
One or more IADL limitations	4.5	1.1	31.3	1.0
Mental condition^a	6.1	1.0	34.5	1.3

SOURCE: SSA/OP/ORES/DER/Disability Modeling Group, January 20, 2000.

NOTE: The sample excludes those receiving disability benefits. The data are based on wave 7 of the 1990 Survey of Income and Program Participation. The estimates are based on the full sample with dynamic cutoff.

^aFirst condition as reason for work limitation is condition code 01, 17, 19, 20, or 23. Also includes persons having one or more mental or emotional problems irrespective of the presence of a reported work limitation.

Table 4.
Classification success rates for allowed and denied applicants using
alternative simulation methodologies (percent of applicants, weighted).

Cutoff method	No sample selection		Sample selection	
	correction		model	
	Full sample	Restricted sample	Full sample	Restricted sample
Random number generator cutoff				
Overall success rate	59.1	57.2	61.6	60.1
Allowed success rate	45.6	45.1	39.6	37.3
Denied success rate	69.7	66.6	78.7	77.8
Cutoff at weighted mean for applicants				
Overall success rate	63.7	63.5	68.4	67.8
Allowed success rate	57.8	55.5	67.2	65.0
Denied success rate	68.2	69.8	69.4	70.0
Cutoff at 0.5				
Overall success rate	64.1	64.0	63.6	63.4
Allowed success rate	42.2	41.4	30.1	27.5
Denied success rate	81.2	81.5	89.6	91.3

SOURCE: SSA/OP/ORES/DER/Disability Modeling Group, January 20, 2000.

NOTE: The estimates are based on data from the 1990 Survey of Income and Program Participation as well as Social Security Administration information on disability determinations.

Table 5.
Health profile of eligibles under two alternatives, as compared to
allowed applicants (in percent unless otherwise indicated).

Characteristic	Eligibles^a		Allowed applicants
	Preferred (1)	Alternative (2)	(3)
Total population aged 18-64 (thousands)	4,393	9,655	2,263
Work limitation status (wave 7, total)	100.0	100.0	100.0
Not limited	32.9	62.3	23.4
Limited but not prevented	30.1	23.1	23.6
Prevented	37.0	14.6	53.0
One or more functional limitations	51.5	34.3	59.8
One or more severe functional limitations	28.5	13.1	32.9
One or more ADL limitations	18.3	7.3	20.3
One or more IADL limitations	31.3	11.5	30.2

SOURCE: SSA/OP/ORES/DER/Disability Modeling Group, January 20, 2000.

NOTE: The sample excludes those receiving disability benefits. Data are for early 1992 based on the 1990 Survey of Income and Program Participation.

^aThe preferred estimate uses the dynamic cutoff, the full sample, and sample selection. The alternative uses the random number generator, the restricted sample, and sample selection.

Table 6.
Profile of disability eligibles in the general population (in percent
unless otherwise indicated)

Characteristic	Eligible	Ineligible
Total population aged 18-64 (thousands)	4,393	145,356
Health		
Work limitation status (wave 7, total)	100.0	100.0
Not limited	32.9	91.1
Limited but not prevented	30.1	6.8
Prevented	37.0	2.0
One or more functional limitations	51.5	8.4
One or more severe functional limitations	28.5	2.0
One or more ADL limitations	18.3	0.9
One or more IADL limitations	31.3	1.0
Mental Condition	34.5	1.3
Demographic		
Age (total)	100.0	100.0
18-44	51.0	70.3
45-54	14.3	17.5
55-64	34.7	12.2
Female	54.7	51.5
Not married	56.3	37.5
Less than high school education	36.6	14.0
Black	12.8	10.7
Hispanic origin (any race)	6.9	7.8
Financial		
Poor	22.1	8.0
Poor or near poor	30.8	11.2
Average monthly earnings (total) ^a	100.0	100.0
Greater than \$500	31.6	65.5
\$1 - \$500	17.8	15.3
Zero	50.6	19.2

SOURCE: SSA/OP/ORES/DER/Disability Modeling Group, January 20, 2000.

NOTE: The sample excludes those receiving disability benefits. The data are for early 1992, based on the 1990 Survey of Income and Program Participation. The estimates use the full sample, the dynamic cutoff, and sample selection.

^aBased on SSA earnings data for 1991.

Chart 1.---SSA disability determination process

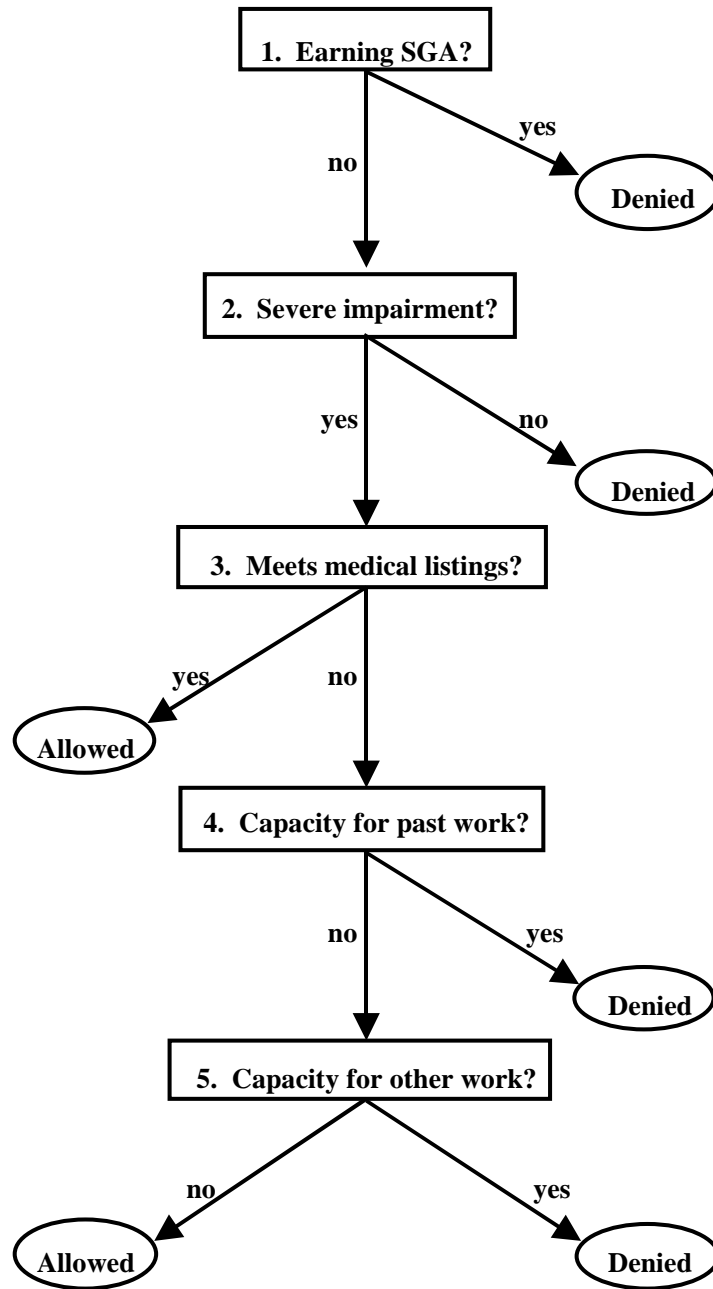


Chart 2.---The sequential disability determination model

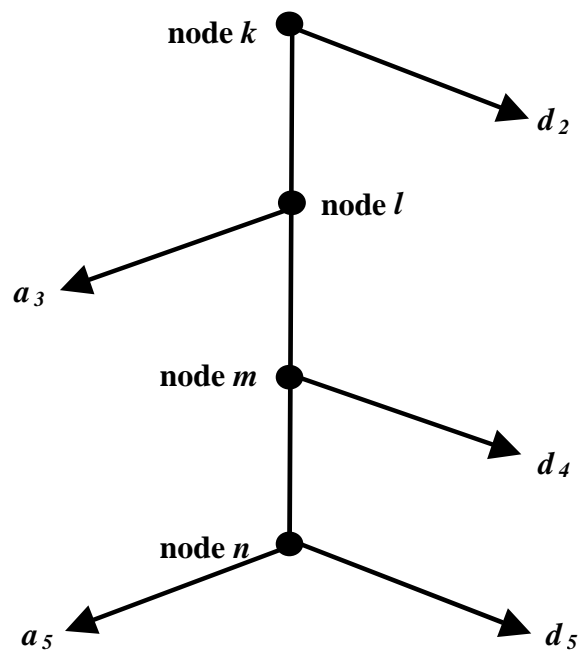


Chart 3.--Disability Allowance Probabilities, By Work Limitation Status, With Sample Selection, Full Sample

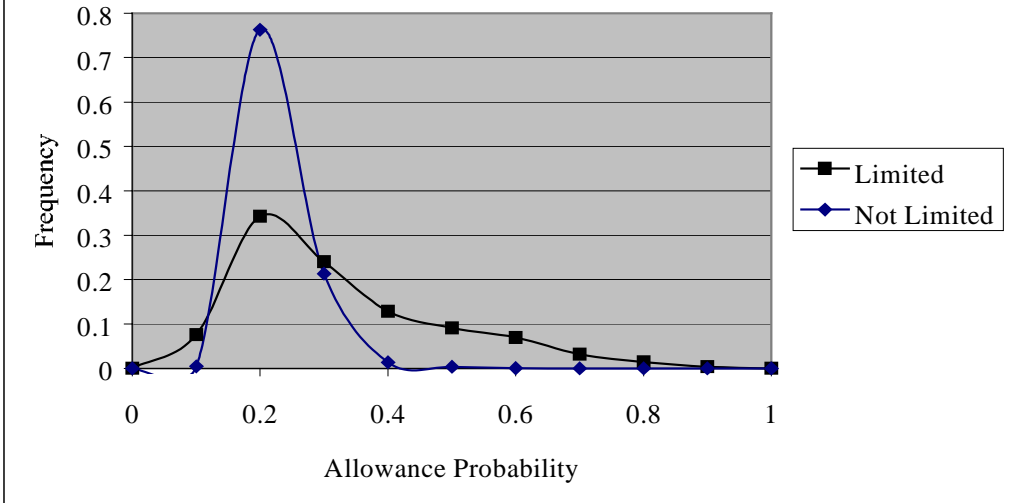
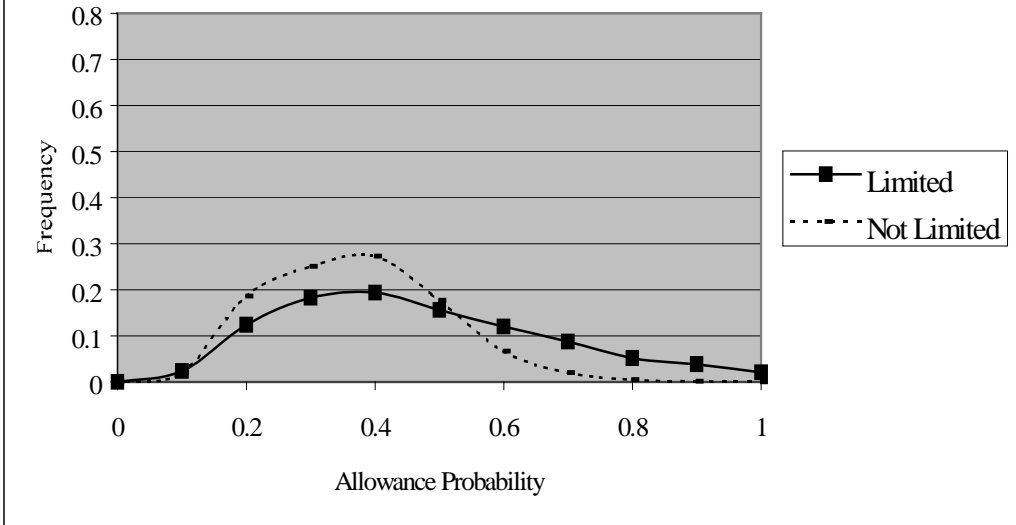


Chart 4.--Disability Allowance Probabilities, By Work Limitation Status, Without Sample Selection, Full Sample



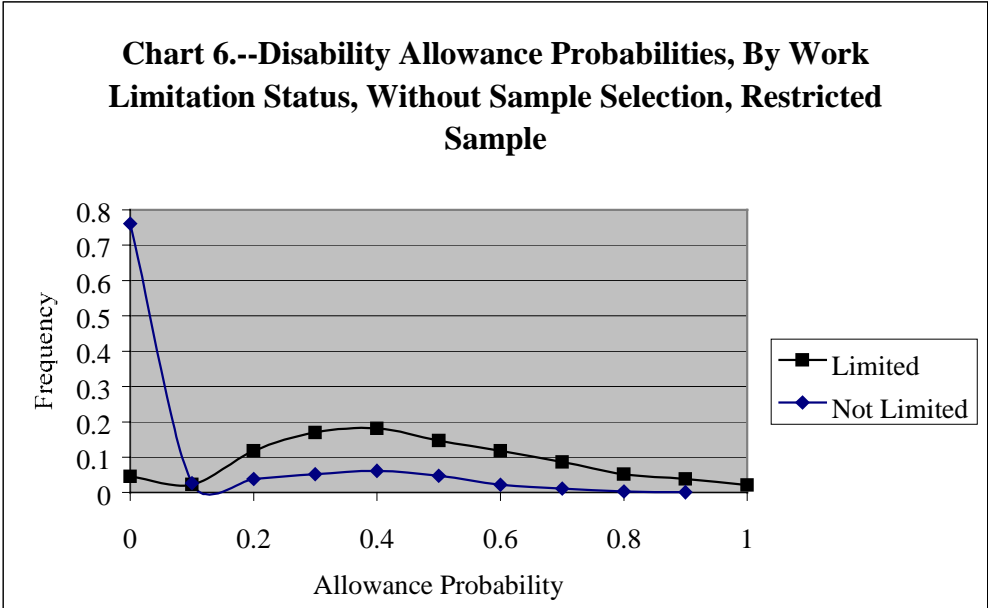
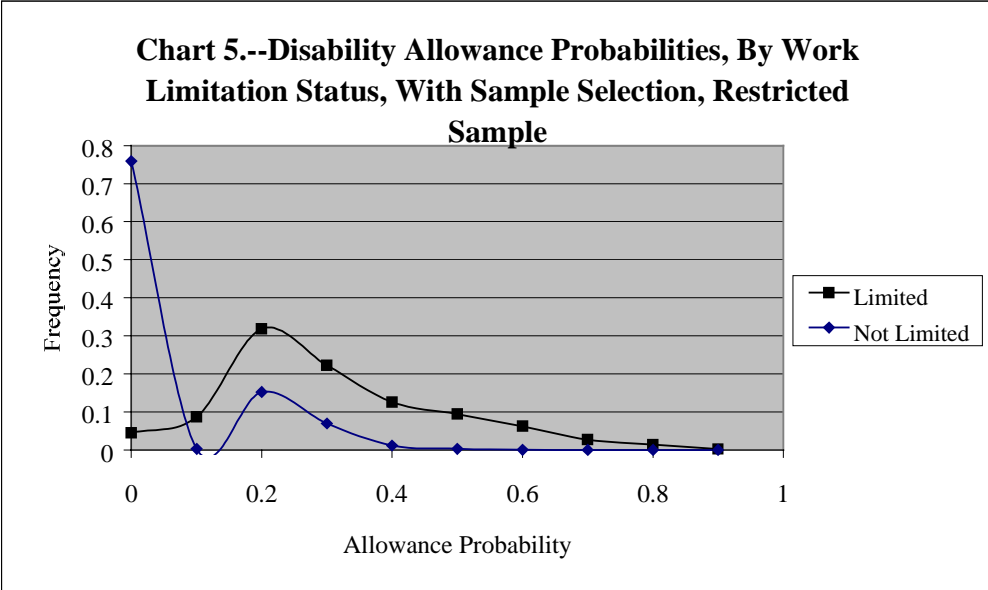
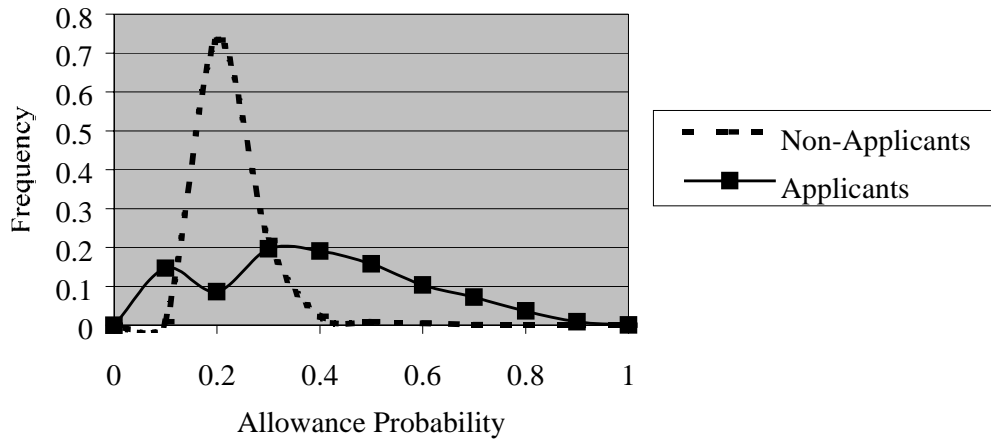


Chart 7.--Disability Allowance Probabilities, by Application Status, With Sample Selection, Full Sample



Appendix A

Dealing with Data Discrepancies: The SIPP and Administrative Records

The administrative records identify applicants, permitting us to match administrative variables into SIPP records for applicants in the SIPP sample. Data from both sources were used in Hu and others (1997) to estimate the disability determination model. But because program data are not available for nonapplicants, we need proxy variables from the survey. The left column of the table below lists the variables or variable categories used in Hu and others for the four steps we model. The right column accounts for availability in the SIPP or proxies used.

Variable	Availability/Proxy
Health Variables (includes all – ADL, IADL, mental, accidents, hospitalization, functional status...)	SIPP variables available
Duration - < 12 months	Available
Work attachment	Available
Job characteristics/DOT conditions	Available
Gender/Male Age Race Marital Status Region Education	All demographic variables are available in SIPP, although in some cases administrative data were used in the original disability determination model (see Hu and others).
Low Income Area	Low Income Area was used in the Hu and others estimation. We substitute Net Worth and Number of Vehicles as proxies for economic status.
Workload, Wait Time, DIProcess Time, Filing Date	We drop Workload, Wait Time, DIProcess Time, and use 1991 as the filing date.

In the Hu and others determination model the age control used is the age at the filing date from 831 disability records. Since we assume wave 7 as the date of filing for the general population, we substitute age in wave 7. This substitution is made in all steps in which the age variable is used interactively as well (i.e., Young/Skilled, Old/Low Education, and Young/No Mental).

If the limitation that prevents work is recent (less than one year), as shown in the 831 record, Hu and others find that this significantly reduces the probability of being passed on at step 2 of the determination process. The SIPP asks a comparable duration question ("Were you limited for less than one year?") that is used here as a proxy.

Among the applicants in our sample, 72 percent apply for Title II benefits. This variable significantly increases the probability of being passed on at step 2 (see Hu and others). There is no proxy for this variable among nonapplicants. It also does not make sense to assume values of 1 or 0. We run the determination model as a reduced form, without that variable, but rather including determinants of Title II instead. These variables are work experience and race (which have proven to be tied to application and work behavior).

Another variable that contributes only slightly, but significantly, to the disability determination process is the caseload of district offices relative to the national averages (Workload). This variable was not used because it was not available for the general population.²⁰

²⁰ In future work we will create a proxy that compares state averages to the national average.

The average waiting time for application processing within district offices (Wait Time) is not observable in the SIPP. Another variable unobserved in the SIPP measures the timing of the application filing date (Filing Date). Obviously this does not exist for nonapplicants. We use the date of wave 7 as the filing date. The SSA determination year (the year the decision is made) is 1991. This variable is interacted with the presence of any work experience. In the estimation for step 2, we use this variable for 1990 interacted with work experience as a control for the 1989 changes in mental health eligibility. We chose to assume that those in the general population have been exposed to such legislation and would therefore take a value of 1 for the 1990 legislation variable.

Appendix B

Logit and Probit Estimates of Eligibility on DI and SSI Applicants

Variables	Model With No Sample Selection	Sample Selection Model	
		Eligibility	Application
<i>Step 2</i>			
Intercept	1.5684 (0.25)**	-0.2570 (0.46)	-2.4153 (0.23)**
Hospital		0.1356 (0.11)	0.5110 (0.04)**
SevereADL	1.6145 (0.97)**		
SevereIADL	0.8658 (0.41)**	0.3360 (0.13)**	0.5868 (0.04)**
Functional Limit		0.4343 (0.14)**	0.1610 (0.06)**
Work Limitation		0.3424 (0.19)**	1.0885 (0.04)**
PoorHealth	0.6875 (0.23)**		
GoodHealth	-0.4977 (0.27)**		
Mental	1.1529 (0.37)**	0.4815 (0.15)**	0.2044 (0.06)**
Duration	-0.9047 (0.31)**		
Occas Work	-0.8651 (0.44)		
Never Worked	1.3546 (0.73)**		
Young	-0.1671 (0.23)		
WorkEnviron	-0.5662 (0.37)**		
Work/90	-0.5498 (0.26)**		
Gender	0.6564 (0.21)		0.1501 (0.04)**
Race	0.4052 (0.20)**		
YoungBlack	-1.6363 (0.59)**		
Net Worth	0.2850 (0.14)**		-0.0064 (0.00)**
Num Vehicles			-0.0484 (0.01)**
Boston	-0.5401 (0.49)**		
Chicago	-0.8840 (0.26)**		
Atlanta	-1.2290 (0.24)**		
Dallas	-1.1516 (0.28)**		
Age			0.0411 (0.01)**
Age Squared			-0.0004 (0.00)**
Education			-0.0516 (0.01)**
Some College			-0.1175 (0.05)**
White and South			-0.2046 (0.05)**
White and North			-0.2543 (0.04)**
λ			0.3517 (0.17)**

(To continue)

Logit and Probit Estimates of Eligibility on DI and SSI Applicants (Continued)

Variables	Model With No Sample Selection	Sample Selection Model
<i>Step 3</i>		
Intercept	-0.8018 (0.25)**	-0.5108 (0.13)**
Hospital	0.6209 (0.20)**	0.3733 (0.11)**
Severe ADL	1.1680 (0.61)**	0.7278 (0.34)**
2IADL	0.6981 (0.25)**	0.4286 (0.13)**
Mental/No Work	0.8993 (0.43)**	0.5428 (0.24)**
Unable to Work	-0.4377 (0.23)**	-0.2740 (0.13)**
Accident	-0.5170 (0.29)*	-0.2941 (0.15)**
Musculo	-0.6054 (0.27)**	-0.3560 (0.14)**
Neurolo	0.9131 (0.41)**	0.5678 (0.23)**
NeverMarried	0.4997 (0.24)**	0.3197 (0.13)**
Work/91	0.6247 (0.26)**	0.3784 (0.15)**
Work/92	0.6446 (0.24)**	0.3819 (0.14)**
Seattle	0.6172 (0.54)	0.3565 (0.29)
<i>Step 4</i>		
Intercept	-0.4596 (0.52)	0.1216 (0.27)
FL/Heavy Occu	0.2111 (0.35)	0.0256 (0.41)
Same Work	-0.7194 (0.36)**	-0.4524 (0.21)**
NoMental/GED	-0.4917 (0.32)	-0.4880 (0.17)**
White Collar/Edu	0.2452 (0.27)	0.0978 (0.15)
NeverWorked	1.8447 (0.47)**	0.9465 (0.24)**
WorkEnviron	0.7275 (0.29)**	0.0167 (0.21)
Boston	0.9787 (0.61)	0.4443 (0.32)
<i>Step 5</i>		
Intercept	-0.1843 (0.33)	-0.0964 (0.20)
Old/LowEdu	1.2754 (0.61)**	0.6762 (0.34)**
Young/NoMental	-1.5129 (0.33)**	-0.9233 (0.20)**
Old	1.0693 (0.43)**	0.6296 (0.26)**
Unskilled	1.0924 (0.54)**	0.6848 (0.32)**
Heavy Occu	-0.1427 (0.31)	-0.2777 (0.24)
Occas Work	-2.2058 (1.09)**	-1.0696 (0.55)**
Young/Skilled	-0.4798 (0.62)	-0.2838 (0.37)
Dummy91	0.7042 (0.38)**	0.3911 (0.23)
Dummy92	0.8401 (0.35)**	0.5106 (0.21)**
Dallas	-1.2049 (0.46)**	-0.7017 (0.27)**

* Significant at the 5% level. ** Significant at the 1% level.

Note: The first column of results reports the parameter estimates from estimating the disability determination model on applicants with no controls for sample selection. Because of slight changes to the model specification (variable substitutions using SIPP data) and sample restrictions, the results vary somewhat from the earlier work by Hu and others (1997) and Lahiri, Vaughan, and Wixon (1995), although not substantially. The numbers in parentheses are standard errors. Blanks indicate a variable was not included under a particular model specification.

Discussion of These Results

The effects of the health status variables are as expected across all models and steps of the process. Applicants with more health problems are more likely to be passed on or allowed. This is not surprising. In fact, introducing sample selection does not affect the health effects.

Comparing models with and without sample selection, we find some noteworthy differences. The model without sample selection includes socioeconomic factors in the model of step 2. Theoretically, socioeconomic factors should not affect the decision at step 2 because this step functions purely as a medical screen. However, without the sample selection control the model becomes a reduced form. Economic factors proxy for the decision to apply for benefits and are therefore significant factors in the decision process. Older, white males are much more likely to be passed on at step 2. Regional differences exist. Work environment and experience have a significant effect on the probability of being passed on at step 2. The model with sample selection is more structural in that it only includes indicators of health in the eligibility model. However, both health and socioeconomic factors significantly affect the decision to apply.

Older non-white men are more likely to apply.²¹ Those more highly educated are less likely to apply. Those with high assets (vehicles and net worth) are less likely to apply

Applicants who have been hospitalized are more likely to apply. This variable was not significant as an indicator of health in the model without sample selection at step 2.

Hospitalization is highly correlated with health insurance status. People without health insurance or resources may be hospitalized because they cannot get treatment elsewhere. Lack of health insurance implies a need for Medicare or Medicaid. It is also tied to work and disability insurance. Hence hospitalization may serve as a control for health insurance status in the application model, picking up some of the effect of applying for disability benefits to get on Medicare or Medicaid.

The effect of λ (the propensity to apply) is positive and significant, as expected. That is, the greater the propensity to apply, the higher the eligibility probability. The simulations do include each individual's λ so that the eligibility probability is conditional on the propensity to apply.

²¹ In part, this may reflect the fact that men are more likely to be insured for disability.

Appendix C

Variable Definitions and Data Sources

Health Variables

SevereADL	: Having three or more severe ADLs;
SevereIADL	: Having one or more severe IADLs;
PoorHealth	: Poor on a self-evaluative five-point health scale;
GoodHealth	: Excellent or very good health;
Mental	: Indication of ever having any mental conditions;
Hospital	: Hospital stay overnight or longer in the past 12 months;
2IADL	: Having two or more IADLs;
Accident	: Health conditions caused by an accident or injury;
Musculo	: Functional limitations caused by musculoskeletal conditions;
Neuro	: Functional limitations caused by special senses and neurological disorders;
Mental/NoWork	: Mental conditions and no work history;
Functional Limit	: Any functional limitation, such as difficulty seeing words and letters or difficulty walking three city blocks;
Work Limitation	: Health or condition limits the kind or amount of work;

Demographic Variables

Age	: Age in wave 7;
Age Squared	: Age squared in wave 7;
Education	: Highest grade level attained;
Some College	: Educational attainment greater than 12;
Gender	: Applicant is male;
Race	: Applicant is white;
Young	: Age less than 35 on the filing date;
YoungBlack	: Age less than 35 years and black;
NeverMarried	: Marital status is never married;
Young/NoMental	: Younger than 54 years and reporting no mental conditions;
Old	: Older than 55 years;
Old/LowEdu	: Older than 55 years, education less than 11 years, and previous occupation unskilled as defined in Dictionary of Occupational Titles (DOT) (i.e., specific vocational preparation, SVP, requiring short demonstration up to one month only);
WhiteCollar/Edu	: Education \geq 12 years and white collar occupation (sales and services);
White and South	: White and resides in Southern states;
White and North	: White and resides in Northern states;

Boston	: Federal Region I including Maine, New Hampshire, Vermont, Massachusetts, Connecticut, and Rhode Island;
Chicago	: Federal Region V including Minnesota, Michigan, Indiana, Ohio, Wisconsin, and Illinois;
Atlanta	: Federal Region IV including Kentucky, Tennessee, North Carolina, South Carolina, Alabama, Mississippi, Georgia, and Florida;
Dallas	: Federal Region VI including New Mexico, Texas, Oklahoma, Arkansas, and Louisiana;
Seattle	: Federal Region X including Washington, Oregon, Idaho, and Alaska;

Socioeconomic Variables

LowIncomeArea	: Percent of households in a district office area with income under \$15,000. These are U.S. Census data, and are available through the Profiling System Database of SSA's Office of Workforce Analysis;
NetWorth	: Net assets;
NumVehicles	: Number of vehicles possessed;

Occupational Variables

Duration	: Work limitation lasted less than 12 months;
Occas Work	: Able to work occasionally or irregularly;
NeverWorked	: Never been able to work at a job due to health conditions;
UnableToWork	: Reported unable to work;
FL/Heavy Occu	: (1) Having one of the three functional limitations (FL) – difficulty walking ¼ of a mile, walking upstairs, and lifting and carrying 10 lbs., and (2) previous occupation requiring 10 - 20 lbs. (Heavy work) or in excess of 20 lbs. (Very heavy work) of force constantly to move objects, as defined in DOT. First, the (0,1) dummy for the strength factor at the 9-digit level was aggregated to the 3-digit level using Crosswalk. We then defined Heavy Occupation as a dummy taking the value one if the aggregated value at the 3-digit level was nonzero and 0 otherwise;
SameWork	: Having work limitations, but able to do the same kind of work;
Unskilled	: Previous occupation requiring SVP between short demonstration and up to one month;
Heavy Occu	: Previous occupation requiring heavy or very heavy work;
WorkEnviron	: Previous occupation involving four or more hazardous work conditions. For each 9-digit occupation, DOT identifies the following work conditions: exposure to weather, extreme cold, extreme heat, wet and/or humid, noise, vibration, atmospheric conditions, proximity to moving mechanical parts, electrical shock,

high or exposed places, radiant energy, working with explosives, toxic or caustic chemicals, and other hazards. The (0,1) dummy indicating four or more hazardous work conditions at the 9-digit level was aggregated to the 3-digit level using the Crosswalk database;

- Work/90 : Having recent work experience and the decision year was 1990;
- Work/91 : Having recent work experience and the decision year was 1991;
- Work/92 : Having recent work experience and the decision year was 1992;

Other Variables

- Young/Skilled : Previous occupation was skilled (SVP more than 6 months), age less than 54 years, and had one or more severe functional limitations or ADLs;
- Workload : Difference in the dispositions per Full Time Equivalent staff year between state and the national levels collected from DDS staffing and workload analysis reports for years 1989 to 1993 (SSA Office of Systems, Office of Information Management);
- DIProcessTime : Mean overall processing time for all DI initial claims from the DO Profiling System Database, SSA's Office of Workforce Analysis;
- WaitTime : Average waiting time (in days) between the filing date and the date of decision at district offices;
- λ : Inverse Mills ratio;
- NoMental/GED : No mental conditions and occupation requiring general educational development (GED). The GED Scale ranging from 1 - 6 is composed of three divisions: Reasoning Development, Mathematical Development, and Language Development. Using the Crosswalk database, we used the average of these scores aggregated over all 9-digit occupations;
- Dummy91 : Decision year was 1991; and
- Dummy92 : Decision year was 1992.

References

- Benitez-Silva, Hugo, Hiu-Man Chan, John Rust, and Sofia Sheidvasser. 1999. "An Empirical Analysis of the Social Security Disability Application, Appeal, and Award Process." *Labour Economics* 6: 147-178.
- Blank, Rebecca M., and Patricia Ruggles. 1996. "When Do Women Use Aid to Families with Dependent Children and Food Stamps? The Dynamics of Eligibility versus Participation." *Journal of Human Resources* 31 (Winter): 57-89.
- Bound, John, Michael Schoenbaum, and Timothy Waidmann. 1995. "Race and Education Differences in Disability Status and Labor Force Attachment." *Journal of Human Resources* 30 (Supplement): S227-S267.
- Burkhauser, Richard V., Robert H. Haveman, and Barbara L. Wolfe. 1993. "How People with Disabilities Fare When Public Policies Change." *Journal of Policy Analysis and Management* 12 (Spring): 251-269.
- Cramer, J. S. 1997. "Predictive Performance of the Binary Logit Model in Unbalanced Samples." Unpublished manuscript, June.
- Giannarelli, Linda, and Nathan Young. 1992. "Estimation of an AFDC Participation Function for the TRIM2 Microsimulation Model." Urban Institute, Washington, D.C., June 1.
- Halpern, Janice, and Jerry A. Hausman. 1986. "Choice Under Uncertainty: A Model of Applications for the Social Security Disability Insurance Program." *Journal of Public Economics* 31(November): 131-161.
- Haveman, Robert, Barbara Wolfe, and Jennifer Warlick. 1988. "Labor Market Behavior of Older Men, Estimates from a Trichotomous Choice Model." *Journal of Public Economics* 36 (July): 153-175.
- Haveman, Robert H., Barbara L. Wolfe, Brent Kreider, and M. Stone. 1994. "Market, Work, Wages, and Men's Health." *Journal of Health Economics* 13:163-182.
- Hosmer, David H., and Stanley Lemeshow. 1989. *Applied Logistic Regression*. New York: Wiley.
- Hu, Jianting, Kajal Lahiri, Denton R. Vaughan, and Bernard Wixon. 1997. "A Structural Model of Social Security's Disability Determination Process." ORES Working Paper No. 72 (August). Also forthcoming in the May 2001 issue of *The Review of Economics and Statistics*.

- Kreider, Brent. 1998. "Workers' Applications to Social Insurance Programs When Earnings and Eligibility Are Uncertain." *Journal of Labor Economics* 16 (4): 848-877.
- Lahiri, Kajal, Denton R. Vaughan, and Bernard Wixon. 1995. "Modeling Social Security's Sequential Disability Determination Using Matched SIPP Data." *Social Security Bulletin* 58 (April): 3-42.
- Lemeshow, Stanley, Daniel Teres, Jill Spitz Avrunin, and Harris Pastides. 1988. "Predicting the Outcome of Intensive Care Unit Patients." *Journal of the American Statistical Association* 83: 348-356.
- McNeil, John M. *Americans with Disabilities 1991-92: Data from the Survey of Income and Program Participation*. U.S. Bureau of the Census, Current Population Reports, Series P-70, No. 33, December. Washington, D.C.: U.S. Government Printing Office.
- Rupp, Kalman, and David Stapleton. 1995. "Determinants of the Growth in the Social Security Administration's Disability Programs—An Overview." *Social Security Bulletin* 58 (April): 43-70.
- Social Security Administration. 1996. *Annual Statistical Supplement to the Social Security Bulletin*. Washington, D.C.: U.S. Government Printing Office.
- Stapleton, David, Burt Barnow, Kevin Coleman, Kimberly Dietrich, Jeff Furman, and Gilbert Lo. 1994. *Labor Market Conditions, Socioeconomic Factors and the Growth of Applications and Awards for SSDI and SSI Disability Benefits: Final Report*. Lewin-VHI, Inc., and Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation.
- Yelowitz, Aaron S. 1998. "Why Did the SSI-Disabled Program Grow So Much? Disentangling The Effect of Medicaid." *Journal of Health Economics* 17 (3): 321-349.