WORKING PAPER SERIES

The Timing of Disability Insurance Application: A Choice-Based Semiparametric Hazard Model

Yang-Woo Kim
J.S. Butler
Richard V. Burkhauser

Working Paper 1996-005A
http://reseach.stlouisfed.org/wp/1996/96-005.pdf

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
411 Locust Street
St. Louis, MO 63102

THE TIMING OF DISABILITY INSURANCE APPLICATION: A CHOICE-BASED SEMIPARAMETRIC HAZARD MODEL

ABSTRACT

We use a choice-based subsample of Social Security Disability Insurance applicants from the 1978 Social Security Survey of Disability and Work to test the importance of policy variables on the timing of application for disability insurance benefits following the onset of a work limiting health condition. We correct for choice-based sampling by extending the Manski-Lerman (1977) correction to the likelihood function of our continuous time hazard model defined with semiparametric unmeasured heterogeneity and find that this correction significantly affects the results. We find that economic variables—the size of the disability benefit, expected wage earnings and accommodation—matter.

JEL CLASSIFICATION: D6, C4

KEYWORDS: Disability Insurance Application, Semiparametric Hazard Model, Choice-based Sample

Richard V. Burkhauser
Center for Policy Research
Syracuse University

J.S. Butler
Department of Economics
Vanderbilt University

Yang-Woo Kim
Senior Economist
The Bank of Korea
Visiting Scholar
Federal Reserve Bank of St. Louis
411 Locust Street
St. Louis, MO 63102

This study is funded by the National Institute on Aging, Program Project #1-PO1-AG09743-01, "The Well Being of the Elderly in a Comparative Context". Kim completed this work with a research award from the Arthritis Foundation. The authors thank George A. Slotsve, Kathryn H. Anderson and Theodore P. Pincus for careful readings on earlier drafts of this paper.

Abstract

We use a choice-based subsample of Social Security Disability Insurance applicants from the 1978 Social Security Survey of Disability and Work to test the importance of policy variables on the timing of application for disability insurance benefits following the onset of a work limiting health condition. We correct for choice-based sampling by extending the Manski-Lerman (1977) correction to the likelihood function of our continuous time hazard model defined with semiparametric unmeasured heterogeneity and find that this correction significantly affects the results. We find that economic variables—the size of the disability benefit, expected wage earnings and accommodation—matter.

Introduction

Too often disability is seen as a medical problem with a medical solution, but the major changes in the size and composition of the disability transfer population in the United States—those receiving disability transfer benefits from either Social Security Disability Insurance or Supplemental Security Income—over the last two decades makes it clear that the transition onto the disability rolls is also influenced by the social environment faced by those with disabilities.

While the road to disability benefit status begins with a health condition, the transition onto the disability rolls can be influenced by the personal and economic characteristics of individuals, as well as by the government policies and labor market conditions they face. In this paper we recognize the dynamic nature of the transition onto the disability rolls by those who have a health condition that affects their ability to work by using a continuous time hazard model to measure the speed at which such workers apply for disability insurance benefits following the onset of their health condition.

To estimate our model we use the 1978 Social Security Survey of Disability and Work. This dataset includes a choice-based subsample of Social Security Disability Insurance applicants. We correct for choice-based sampling by extending the Manski-Lerman (1977) correction to the likelihood function defined with semiparametric unmeasured heterogeneity (Butler, Anderson, and Burkhauser 1989). Our results show important effects of the correction for choice-based sampling.

The importance of disability insurance on the decision by workers with serious health conditions to leave the labor force has been in dispute for over a decade. Parsons (1980, 1991), for instance, argues that older workers with health conditions are highly sensitive to the reward structure of the disability insurance system relative to labor earnings. Other researchers have found that the expected replacement rate of disability insurance transfers influences labor supply but to a much smaller degree. (See, for instance, Haveman, de Jong, and Wolfe 1991.) Perhaps most damaging to the view that disability insurance plays an important role in the retirement decision of health impaired workers is the finding by Bound (1989) that less than 50 percent of unsuccessful disability insurance applicants were subsequently employed and that only about two-fifths of the 50 percent worked full-time. (For further discussion of this point, see Bound 1991; Parsons 1991.)

The fact that the majority of unsuccessful candidates never return to work does not mean, however, that disability insurance policy does not affect the decision to leave the labor force and apply for disability benefits. Substantial time may elapse between the onset of a health condition, its first impact on work performance, and its subsequent influence on job exit and application for disability benefits. Moreover, application for disability insurance is itself a risky gamble in which the ultimate outcome can be delayed for months or, in some cases, years. Applicants for disability benefits may "invest" in not being able to work to maximize their chances in what can be a long drawn out review process. What Bound has found is that, for most workers, the decision to apply for benefits is tantamount to a decision to withdraw permanently from the labor market. But the size of disability benefits may still be a policy lever which importantly affects the point at which health-impaired workers take that gamble.

It is important to look at the timing of the disability application in order to see not only whether workers ever apply for disability insurance but also the speed at which they do so following the onset of a health condition. This dimension of the empirical facts uncovered by Bound suggests that the disability application process can be described by a hazard model.

Data

The 1978 Survey of Disability and Work is a nationally representative economically based dataset containing information on disabled workers. (For technical details, see Bye and Scheckter

1982.) This survey of the prevalence of work disabilities in the working age population was conducted by the Social Security Administration and contains two sampling frames. The first is a subsample of the Health Interview Survey (HIS) and is representative of the general population of noninstitutionalized persons aged 18 to 64. It contains data on 5,652 persons. Unfortunately, because official disability status is a relatively rare event, the actual number of respondents in this population who applied for disability insurance benefits is small. For this reason, a second frame was developed to increase the number of respondents who had applied for disability benefits. This second frame is drawn from administrative records and consists of 4,207 persons who applied for disability insurance.

Respondents in both frames were asked identical questions, including the identity of any health conditions they had and when their main health condition began. In our study, the time of onset is defined as the year when either the main health condition began to bother the respondent or when the respondent first became aware of its presence. Additional retrospective information on labor market activity, including occupation, industry, and household characteristics at the time each respondent's health began to limit his ability to work, is available. Information regarding application for disability insurance is also reported.

These survey data were matched with Social Security earnings records for each respondent. These data contain the yearly earnings of a worker since 1951 and the number of quarters of Social Security-covered employment since 1938. Combining information from the 1978 Survey of Disability and Work with Social Security earnings records allows us to trace an individual's economic behavior from the time his health condition first starts until the date of application for disability insurance.

In our empirical model, we analyze the behavior of the male working age population, i.e., those under age 60 at the survey date and older than age 20 at the onset of their main health condition. Our sample is further confined to those employed at the onset of their health condition. Those with missing information on either time of onset or date of application are excluded.[1]

As a result of this selection process, our sample consists of 1,430 observations—348 observations from the Health Interview Survey and 1,082 observations from the administrative sample.

Choice-Based Sampling

Due to the nature of a two-frame sampling approach, special attention must be paid in selecting the study sample. Since the administrative frame is composed of disability insurance applicants only, application duration can be calculated directly without censoring. However, using this sample to look at duration until application for disability benefits would lead to biased results, since the administrative frame oversampled early applicants. Therefore, the proportions of our mixed total sample will inconsistently estimate the corresponding population proportions. Manski and Lerman (1977) show that treating choice-based samples as if they were random and calculating estimators appropriate to random samples yield inconsistent estimates. They introduced a weighted likelihood function that can generate consistent estimates.

To obtain consistent estimators, we use the total sample with an estimator corrected for choice-based sampling. By merging the two frames, a larger and more heterogeneous population is obtained. Each observation's contribution to the log likelihood is weighted by $P(j)/S(j)$, where $P(j)$ is the probability density of a person with the characteristics of person j and $S(j)$ is the sample probability of a person with the characteristics of person j. This is defined based on the sample design (oversampling or undersampling certain people) and the population of applicants. These weights are available from the 1978 Survey. However, 75 persons, 22 percent of the Health

Interview Survey, appeared in the administrative sample. These respondents are not identified in the administrative sample. For these respondents, we use adjusted weights W_b where:

$$W_b = \frac{1}{\frac{1}{W_s} + \frac{1}{W_h}} - W_s, \tag{1}$$

where $W_s = pr(x \mid \text{population}) / pr(x \mid \text{administrative sample})$,

$W_h = pr(x \mid \text{population}) / pr(x \mid \text{Health Interview Survey})$.

The first term in equation (1) is the correct total weight for respondents in both samples, and since they appear twice in the sample, we subtract their assigned weight, W_s. Because W_s cannot be identified in the administrative survey, we assign them the mean value of the administrative weights.

The Hazard Model

The empirical model adopted here is a variant of the hazard model used by Diamond and Hausman (1984). Our hazard rate is the probability of applying for disability insurance following the onset of a health condition, conditional on an initial state of working. The application decision is modeled as a single transition process.

Since information is available on the time of onset of a health condition, the measure of duration does not suffer from left censoring. However, some spells are censored on the right. Our empirical model is estimated with a univariate interval hazard model. Since the time of application for disability insurance is measured in yearly intervals, a point hazard is not appropriate. We control for unobserved individual heterogeneity. This heterogeneity may exist because of omitted variables, uncertainty, or differences in the distribution function across individuals. If these differences are not controlled, then there may be spurious duration dependence. This unobserved factor is integrated out of the likelihood function by assuming a semiparametric form.[2]

The form of our hazard is

$$h(t) = \exp(X'\beta + t\gamma_1 + t^2\gamma_2 + \epsilon)$$

where t is time,

$$\epsilon \sim \lambda(0, \sigma^2) \text{ or } \epsilon \sim f(\epsilon),$$

and $f(\epsilon)$ is estimated as a discrete probability mass function defined at points chosen from numerical integration theory, with the mass estimated. Both Heckman and Singer (1984) and Butler, Anderson, and Burkhauser (1989) estimate a hazard model with an unspecified distribution of the unmeasured heterogeneity by substituting a discrete distribution. The former estimates both the points of support and the probability mass; the latter, which we use here, fixes the points of support and estimates the probability mass. The resulting likelihood function can be made arbitrarily accurate in principle by allowing more points of support.

Variables Affecting Duration Until Disability Insurance Application

The explanatory variables in our hazard models are defined in Table 1. They include economic and health status variables as well as control variables.

Primary Insurance Amount (PIA)

To receive disability insurance benefits a worker must have sufficient quarters of coverage to be eligible for the program, and he or she must be unable to perform any substantial gainful activity. Applicants must have a physical or mental impairment that has prohibited them from working for five months and will make it unlikely that they can work for at least one year. If accepted onto the rolls, the worker's benefits will be determined by program rules related to age and their Social Security earnings history. Based on these factors, a Primary Insurance Amount (PIA) is calculated.

Previous researchers (see, for instance, Leonard 1979; Halpern and Hausman 1986; Bound 1989; Burkhauser, Butler, and Kim 1995) have used individual data on disability insurance applicants to estimate the expected replacement rate of all individuals in this sample, that is, their probability of receiving benefits multiplied by their PIA and divided by their previous wage earnings. Here we use the value of the person's PIA based on his earning record to the time of analysis. PIA then becomes a time-varying variable in our hazard model representing the potential benefit of disability insurance benefits. The larger the benefit, the larger should be the probability or hazard of application for disability benefits.

Expected Earnings

We estimated expected earnings, following the style of macroeconomics, with a vector autoregression. To do this we took advantage of the fact that we have the Social Security earnings history of all respondents in the 1978 Survey. We started at the beginning of each person's earning record, adjusted for the fact that earnings above the Social Security taxable maximum are not captured by these data.[3] Then we used each person with at least five years of earnings to define a vector autoregression of earnings at time t, e_t, on four lagged values of earnings, e_{t-1}, e_{t-2}, e_{t-3}, and e_{t-4}, on a dummy variable, d_t, for whether the person had a health condition at time t that limited his or her ability to work , and interactions between lagged earnings and lagged dummy variables, d_{t-1}, d_{t-2}, d_{t-3}, and d_{t-4}. The number of observations per person varied according to how long the person had earnings; a few people had fewer than five years of earnings and contributed no observations to the regression.

The results of our expected earnings regressions are reported in Table 2. There were 5,745 person years in the HIS sample and 18,463 person years in the administrative records sample. In each case, the autocorrelation possibly present was ignored, because our objective was to obtain an equation to predict wages following the onset of a health condition that limited work, rather than a structural model with theoretically correct standard errors. The variables are all available in the individual's information set and produce a very high R^2 in both samples. Given the presence of recent past earnings, it is doubtful that additional variables such as education could contribute any noticeable additional explanatory power. The equations are broadly similar, with the largest effect coming from one-period lagged earnings, but earnings in the administrative records sample are higher, and the effects of a health condition are lower for the intercept and higher for the slopes in the administrative records sample. Expected earnings represent the opportunity cost of applying for disability insurance benefits, and we expect that those with higher expected benefits will have a lower probability or hazard of application for benefits.

Accommodation

Most economic models of disability insurance application ignore the importance of an employer's behavior on this outcome.[4] Yet the willingness of an employer to adjust the workplace to compensate for an employee's work limitation may play an important role in allowing the worker to continue on his job. This is certainly the belief of those who supported the Americans with Disabilities Act of 1990, which requires employers to provide reasonable accommodations to disabled workers as long as these accommodations do not create an undue hardship on the operation of business. Our data, which predate the passage of this act, allow us to estimate the effect of accommodation on disability insurance application. A variable reports whether an employer did anything to accommodate the worker when his health condition first began to affect his ability to do his job. We expect accommodation to reduce the risk of applying for disability insurance benefits.

Socioeconomic Variables

Age at Onset, Marital Status at Onset, Race, Education, and Experience are also included in our empirical model. Married men, men with more education and experience, young men, and nonblacks are generally found to have stronger ties to the labor force and, hence, we expect these characteristics to delay application.

Job Characteristics

Several researchers have looked at the importance of job attributes on the decision of workers to retire. We report our results using a binary variable to distinguish white collar workers. In estimations not shown we used more elaborate job attribute measures developed by Roos and Treiman (1980) from the Dictionary of Occupational Titles. These measures were used by the Social Security Administration in their 1986 report. We expect that a worker with a disability is less at risk to apply for disability insurance benefits if he is working on a white collar job.

Health Measures

Finally, we attempt to account for variations with our sample by use of a comorbidity measure. Because all of the respondents in our sample have a health condition that affects their ability to work, we are not estimating the unconditional impact of health on the speed of applications. Naturally, those with no health problem are less likely to apply for benefits. Rather, we are attempting to account for variations in health within the population with a health condition. Specifically, we want to capture the speed of application following the onset of such a condition. To see if different types of health conditions influence the speed of application, we choose the two most common physical conditions among the disability insurance population, cardiovascular and musculoskeletal conditions, as well as a measure of multiple conditions. We expect that those with multiple conditions are more likely to apply for DI benefits.

Results

In addition to providing a definition of all variables used in this analysis, Table 1 also reports the unweighted and weighted means of the combined sample of 1,430 persons used in our estimations. The corrected means reflect the Manski-Lerman weights. As will be seen, the weights make a profound difference in the mean characteristics of the total sample because the characteristics of the respondents in the two subsamples are quite different. The choice-based disability insurance applicants are on average older, less educated, and more experienced, and were in worse health at the time their health condition first began to affect their work than the random sample of people with health conditions. The choice-based applicants were also less likely to have been accommodated than the random sample, and they had lower PIA and expected earnings.

Weighting the sample also has a substantial effect on the distribution of spell lengths. Table 1 also shows average spell length increases in the weighted sample, since our subsample of applicants oversampled those who applied for benefits after relatively short spells with a health condition. When we weight the sample, average spell duration increases. This can be seen more clearly in Table 3, which shows the distribution of spell lengths between onset of a health condition and application for benefits. Table 3 reports Kaplan-Meier estimates of the length of spell based on the unweighted and weighted samples. In the unweighted sample, 25.5 percent of the sample apply for benefits in the first year, 63.3 percent after five years, and 77.8 percent after ten years. These probabilities are overstated, because of the oversampling of short spells in the administrative record sample. Using the Manski-Lerman corrected sample, we find only 9.5 percent of respondents applied in the first year, 30.2 percent after five years, and 40.4 percent after ten years. These results suggest that weighting the sample may importantly affect the results.

Table 3 traces the speed of application unadjusted for heterogeneity. In Table 4 we control for both observed and unobserved heterogeneity in our unweighted and weighted samples using a hazard model with log normal and semiparametric unmeasured heterogeneity. The semiparametric model is preferred, because it imposes less restrictive assumptions on the estimation. As can be seen in Table 4, however, using the log normal or semiparametric model makes little difference in the results. The use of Manski-Lerman weights, however, makes a much greater difference. Several coefficients reverse sign, and several coefficients become significant. The unweighted estimates are biased, and are shown here simply to illustrate that using weights in hazard model estimation with choice-based samples can make a difference.

In both samples our principal policy variables—PIA, expected earnings, and accommodation—significantly affect duration until application. A greater PIA, a smaller expected earnings, or the absence of accommodation will all increase the speed at which men will apply for disability benefits following the onset of a health condition.

Other variables have the following effects. Older age at onset, not being married, being white, and having less education increase the speed of application. More experience discourages application for benefits. White collar status is the only insignificant variable in the model. [5] Comorbidity is unexpectedly associated with slower disability insurance application, and a cardiovascular condition increases the speed of application, while a musculoskeletal condition decreases the speed of application.

Table 5 uses the results from the corrected semiparametric estimation of the weighted data to calculate two intuitively appealing measures of the effects of the exogenous variables: 1) their effect on the probability of applying for disability insurance benefits within ten years of the onset of a health condition and 2) their marginal effect on expected duration until application. Both of these are nonlinear functions of the coefficients, and their standard errors are derived using the delta method. A year of education increases duration until application by 0.48 years and decreases the probability of applying for benefits within ten years by 1.6 percentage points. The effect of accommodation is to postpone application by 5.4 more years and to decrease the probability of applying for benefits within the first ten years by 19 percentage points. A $100 monthly increase in PIA increases the probability of application in ten years by 1.8 percent, while a $100 per month increase in expected earnings reduces the probability of application by 0.6 percent. The effect of a $100 increase in the PIA on the expected duration is a reduction of 0.53 years, and the effect of a $100 increase in expected earnings is an increase of 0.19 years. So a dollar of potential benefit from disability insurance has the effect of $3.00 of expected earnings. All of these effects are significantly different from zero.

Conclusion

The 1978 Social Security Survey of Disability and Work contains both a random sample of workers with health conditions and a choice-based sample of such workers. While combining these data provides a larger sample of applicants for disability insurance benefits, the sample is not representative of the population of all people with health conditions. In this paper, a hazard model is estimated to measure the speed at which workers who experience a health condition apply for disability insurance benefits. In so doing, we apply a Manski-Lerman choice-based sampling correction to our semiparametric unmeasured heterogeneity specification of the hazard model. We find that this correction importantly affects our results by altering the estimates and levels of significance of the effects of several explanatory variables.

The estimates show that increasing the Primary Insurance Amount increases the speed of application for disability insurance, while increasing expected earnings and experience decreases the speed of application. A number of demographic, economic, and physical factors affect application. Accommodation by an employer in particular reduces the speed of application, adding as much time to the work life as 70 months of experience or a reduction in PIA of $1,000 per month.

Appendix

Estimation of Yearly Labor Earnings for Persons Whose Wages Exceed the Social Security Taxable Maximum

The mean of yearly labor earnings of persons whose yearly labor earnings exceed the Social Security taxable maximum is calculated in Table A-1. A constant flow of labor earnings is assumed. Y designates labor earnings; M designates the maximum. The task is to assign a value of labor earnings when Y is recorded as M, and the quarter in which M was reached is known.

In the first set of columns, the Fox (1984) technique is illustrated. When M is reached in the fourth quarter, M is assigned; when M is reach in the first, second, or third quarter, 4M, 2M, or 4M/3 is assigned. This technique assumes the maximum is reached at the end of the quarter, creating a bias toward zero.

In the second set of columns, log normal distributions of labor earnings in the population are assumed and estimated using the sample of persons with labor earnings in that year. Then the expected value of the log normal labor earnings, given that it exceeds the statutory tax maximum in a given quarter, is calculated. In the fourth quarter, $4M/3 > Y > M$. In the third quarter, $2M > Y > 4M/3$. In the second quarter, $4M > Y > 2M$. In the first quarter $Y > 4M$. These values exceed the standard Fox technique values, because the truncation is explicitly modeled.

Endnotes

1. Some respondents made multiple applications for the program. The questionnaire asked only the date of the last application. If we use only the first applicant sample, there may be a selectivity problem. Fortunately, our interval hazard estimation technique allows us to use all applications. The time of the first application, which is the one we analyze, must be before the last application; thus, an interval from zero to the time of the last application includes the time of the first application.
2. As shown below, we also do the analysis using a log normal distribution. It does not alter the main findings.
3. In the Social Security earnings records file, when a person's labor income exceeds the Social Security taxable maximum in any year, the maximum is recorded along with the quarter in which the maximum was reached. Those interested in assigning actual earnings rather than maximum taxable earnings must estimate earnings. The Fox technique (Fox 1984) uses the maximum scaled according to the quarter, i.e., the maximum 4M or 2M, 4M/3 or M, if the limit is reached in the first, second, third, or fourth quarter. A more sophisticated approximation method is to assume a distribution of income and to calculate the expected value of the appropriate truncated distribution. We assume a log normal distribution of income. See the Appendix for the calculations and values used for each year from 1951 to 1978.
4. An exception in the economics literature is Baldwin and Johnson (1994), which attempts to measure the part of the wage difference between those with and without disabilities that is caused by discrimination.
5. We found similar results in alternative equations using Roos and Treiman's (1980)

inputed scores to measure the degree of strength require on the job and the job's physical demands. The use of these alternative measures also had no effect on the significance of other variables.

References

Baldwin, Marjorie and William G. Johnson. 1994. "Labor Market Discrimination Against Persons with Disabilities," *Journal of Human Resources*, 29 (Winter): 1-19.

Bound, John. 1989. "The Health and Earnings of Rejected Disability Insurance Application," *American Economic Review*, 79 (June): 482-503.

Bound, John. 1991. "The Health and Earnings of Rejected Disability Insurance Applicants Reply," *American Economic Review* (December): 1427-1434.

Butler, J.S., Kathryn, H. Anderson, and Richard V. Burkhauser. 1989. "Work and Health After Retirement: A Competing Risk Model with Semiparametric Heterogeneity," *Review of Economics and Statistics*, 71(1) (February): 49-53.

Bye, B. and E. Scheckter. 1982. "The 1978 Survey of Disability and Work: Technical Introduction," *Social Security Administration*, January.

Diamond, Peter and Jerry Hausman. 1984. "The Retirement and Unemployment Behavior of Older Men." In Henry Aaron and Gary Burtless (eds.), *Retirement and Economic Behavior*. Washington, DC: Brookings Institution, pp. 97-134.

Fox, Alan. 1984. "Income Changes At and After Social Security Benefit Receipt: Evidence from the Retirement History Study," *Social Security Bulletin*, 47(9) (September): 3-23.

Halpern, Janice and Jerry Hausman. 1986. "Choice Under Uncertainty: Labor Supply and Application for the SSDI Program." In R. Blundell and I. Walker (eds.), *Unemployment, Search and Labor Supply*. Cambridge: Harvard University Press, pp. 294-302.

Hausman, Jerry and David A. Wise. 1981. "Stratification on Endogenous Variables and Estimation: The Gary Income Maintenance Experiment." In Charles Manski and Daniel McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge: MIT Press, pp. 365-391.

Heckman, James and R. Robb. 1985. "Alternative Methods for Evaluating the Impact of Interventions." In James Heckman and Burton Singer (eds.), *Longitudinal Analysis of Labor Market Data*. Cambridge: University Press, pp. 156-246.

Leonard, Jonathan. 1979. "The Social Security Program and Labor Force Participation," NBER Working Paper No. 392. Washington, DC: NBER.

Leonard, Jonathan. 1986. "Labor Supply Incentives and Disincentives for Disabled Persons." In Monroe Berkowitz and M. Anne Hill (eds.), *Disability and the Labor Market*. Ithaca, NY: Cornell University Press, pp. 64-94.

Manski, Charles and S. Lerman. 1977. "The Estimation of Choice Probabilities from Choice Based Samples," *Econometrica*, 45(8): 77-88.

Manski, Charles and Daniel McFadden (eds.). 1981. "Alternative Estimators and Sample Designs for Discrete Choice Analysis." In *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge: MIT Press, pp. 2-50.

Parsons, Donald. 1980. "The Decline in Male Labor Force Participation," *Journal of Political Economy*, 88: 117-134.

Parsons, Donald. 1991. "The Health and Earnings of Rejected Disability Insurance Applicants: Comments," *American Economic Review* (December): 1419-1426.

Roos, P.A. and D.J. Treiman. 1980. "DOT Scales for the 1970 Census Classification." In *Work, Jobs, and Occupations*, A.R. Miller, D.J. Treiman, P.S. Cain, and P.A. Roos (eds.). Washington, DC: National Academy Press, pp. 336-389.

Social Security Administration. 1986. "Increasing the Social Security Retirement Age: Effect on Older Workers in Physically Demanding Occupations or Ill Health," report pursuant to Section 201(d) of P.L.: 98-21.

Table 1. Definitions of Variables and Mean Values for the Sample With and Without the Manski-Lerman Weights

Variables	Definitions	Mean Unweighted	Mean Weighted
Primary Insurance Amount (first year)	Primary Insurance Amount in the first year following onset of a health condition that limits work.	2,289.16 (1,338.74)	2,604.41 (1,789.58)
Expected Earnings[a] (first year)	Expected yearly earnings in the first year following onset of a health condition that limits work.	5,911.37 (3,526.79)	7,453.93 (5,093.82)
Primary Insurance Amount[b] (tenth year)	Primary Insurance Amount in the tenth year following onset of a health condition that limits work.	1,837.42 (1,248.35)	2,114.43 (1,395.39)
Expected Earnings[a] (tenth year)	Expected yearly earnings in the tenth year following onset of a health condition that limits work.	4,985.31 (3,564.65)	6,368.66 (5,085.70)
Experience (100 quarters)	Quarters of coverage in all covered employment at time of onset	0.75 (0.36)	0.64 (0.38)
Accommodation	Equals 1 if at onset the employer provided help to respondent to remain on the job, otherwise 0	0.26 (0.44)	0.34 (0.47)
Age at Onset	Age at onset, years.	41.9 (10.2)	38.5 (10.7)
Married at Onset	Equals 1 if married, otherwise 0.	0.78 (0.41)	0.82 (0.39)
Nonwhite	Equals 1 if nonwhite, otherwise 0.	0.15 (0.36)	0.14 (0.35)
Education	Years of formal education.	9.9 (3.3)	10.9 (3.3)
White Collar	Equals 1 if the occupation at onset is professional or managerial, otherwise 0.	0.17 (0.37)	0.17 (0.37)
Comorbidity	Equals 1 if the respondent had multiple health conditions at onset, otherwise 0.	0.76 (0.43)	0.66 (0.47)
Cardiovascular[c]	Equals 1 if the main health condition is in the cardiovascular disease group, otherwise 0.	0.28 (0.45)	0.17 (0.38)
Musculoskeletal[d]	Equals 1 if the main health condition is in the musculoskeletal disease group, otherwise 0.	0.33 (0.47)	0.47 (0.50)
Length of Spell	Number of years from onset to application or censoring	5.95 (6.77)	7.18 (7.02)

[a]Based on estimates in Table 2.

[b]While individual PIAs increase over time, the mean PIA of those who have still not applied for benefits falls over time since those with higher PIAs apply for benefits sooner.

[c]Includes heart attack, arteriosclerosis, and other heart troubles..

[d]Includes chronic stiffness in arm, hand, foot, leg, or back; other deformities of the back or spine, muscular atrophy, and lupus.

Table 2. Regressions Used to Estimate Labor Earnings

	HIS Sample		Administrative Records Sample	
Variable	Coefficient	Standard Error[a]	Coefficient	Standard Error[a]
Constant	686.777	61.016	1,083.482	38.339
Disabled	-450.627	96.686	-409.078	71.951
Earn -4	0.0933	0.0295	0.1511	0.0180
Earn -3	0.0968	0.0379	-0.0122	0.0216
Earn -2	0.0666	0.0369	-0.0449	0.0221
Earn -1	0.7158	0.0297	0.7730	0.0173
Interactions				
Disabled*E-4	0.0687	0.0358	0.0246	0.0317
Disabled*E-3	-0.0096	0.0449	0.0145	0.0346
Disabled*E-2	0.0848	0.0422	0.1107	0.0285
Disabled*E-1	-0.0480	0.0318	-0.1832	0.0240
R^2	0.9210		0.8883	

[a]Standard errors are adjusted for heteroscedasticity but not for autocorrelation; these regressions are used to predict wages and have no structural interpretation.

Table 3. Kaplan-Meier Estimates of the Time To Application for Disability Insurance With and Without the Manski-Lerman Correction[a]

	Unweighted Sample					Weighted Sample				
	Number of Persons			Probabilities		Number of Persons			Probabilities	
Year	Censored	Applying	In Sample	Applying	Surviving	Censored	Applying	In Sample	Applying	Surviving
1	8	363	1,430	0.255	0.745	78.0	132.2	1,430.0	0.095	0.905
2	14	247	1,059	0.235	0.570	161.0	79.6	1,219.8	0.070	0.842
3	20	107	798	0.136	0.493	97.6	38.1	979.2	0.041	0.807
4	12	93	671	0.140	0.424	55.4	71.3	843.5	0.087	0.737
5	14	75	566	0.134	0.367	71.9	35.9	716.8	0.053	0.698
6	11	49	477	0.104	0.329	60.8	23.9	609.0	0.041	0.669
7	9	44	417	0.107	0.294	32.9	10.5	524.3	0.021	0.655
8	2	36	364	0.099	0.265	17.5	10.5	480.9	0.022	0.641
9	9	28	326	0.087	0.242	70.1	21.5	452.9	0.051	0.608
10	9	21	289	0.074	0.224	31.1	6.8	361.3	0.020	0.596
11	4	27	259	0.105	0.200	10.7	9.3	323.4	0.029	0.578
12	4	20	228	0.088	0.183	21.8	4.1	303.4	0.014	0.570
13	3	15	204	0.074	0.169	25.5	5.3	277.5	0.020	0.559
14	7	21	186	0.115	0.150	19.2	22.6	246.7	0.095	0.506
15	5	15	158	0.096	0.135	22.6	5.0	204.9	0.026	0.492
16	3	10	138	0.073	0.125	6.8	3.7	177.3	0.021	0.482
17	4	9	125	0.073	0.116	9.5	1.9	166.8	0.012	0.476
18	3	12	112	0.109	0.104	7.5	5.0	155.4	0.033	0.461
19	3	11	97	0.115	0.092	18.9	3.1	142.9	0.023	0.450
20	5	5	83	0.062	0.086	33.4	2.2	120.9	0.021	0.440
21	4	4	73	0.056	0.081	24.9	0.6	85.3	0.008	0.437
22	0	9	65	0.138	0.070	2.2	0.8	59.8	0.014	0.431
23	1	7	56	0.126	0.061	2.5	4.0	56.8	0.072	0.400
24	1	5	48	0.105	0.055	2.2	0.8	50.3	0.016	0.393
25	0	4	42	0.095	0.049	0.0	0.5	47.3	0.011	0.389
Total[b]	167	1263				921.7	508.4			

[a]These estimates are based on allocating half of the number censored in each year to the sample available for application in each year. That assumes that censoring is distributed uniformly within a year. Survival means surviving into the next period on the job, i.e., not yet applying for Disability Insurance.

[b]Years where the unweighted sample size is 40 or less are not shown but are included in the total.

Table 4. Time to Apply for Disability Insurance Using Unweighted and Weighted Samples and Assuming Log Normal or Semiparametric Unmeasured Heterogeneity

Explanatory Variables	Unweighted				Weighted			
	Log Normal		Semi-Parametric[a]		Log Normal		Semi-Parametric[a]	
	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
Constant	-0.35	0.36	0.04	0.35	0.06	0.25	-0.63	0.45
Age of onset	0.33	0.08	0.32	0.08	0.38	0.06	0.75	0.08
Married at onset	-0.12	0.11	-0.10	0.10	-0.49	0.13	-0.85	0.16
Nonwhite	0.20	0.13	0.14	0.12	-0.27	0.11	-0.49	0.17
Education (ten years)	-0.57	0.16	-0.49	0.14	-0.68	0.11	-0.51	0.14
Accommodation	-0.65	0.12	-0.47	0.10	-0.44	0.08	-0.62	0.10
Experience (100 months)	-0.66	0.24	-0.48	0.22	-0.10	0.14	-0.91	0.22
White collar	0.17	0.13	0.19	0.12	-0.09	0.10	-0.20	0.12
Comorbidity	0.18	0.10	0.16	0.10	-0.14	0.07	-0.37	0.10
Cardiovascular	0.14	0.12	0.09	0.11	0.28	0.09	0.36	0.13
Musculoskeletal	-0.45	0.11	-0.37	0.10	-0.96	0.11	-1.41	0.12
PIA (annual; $1,000)	1.12	0.06	1.09	0.06	0.52	0.05	0.60	0.05
Expected earnings (annual; $1,000)	-0.22	0.02	-0.21	0.02	-0.17	0.02	-0.21	0.02
Time	-1.21	0.25	-1.29	0.23	-2.41	0.31	-1.63	0.27
Time squared	0.37	0.09	0.37	0.08	0.79	0.11	0.65	0.11
Variance of the Unmeasured Heterogeneity	0.93	0.21			1.33	0.56		

[a]A discrete distribution of the unmeasured heterogeneity is also estimated as a part of this model. This is based on four possible points, -2.33 (probability 0.15 uncorrected, 0.18 corrected), 0.74 (probability 0.00 and 0.39), 0.74 (probability 0.75 and 0.00), and 2.33 (probability 0.10 and 0.43). (See Butler, Anderson, and Burkhauser 1989).

Table 5. Marginal Effects of Variables in the Weighted Sample Hazard Model on Time Until Application for Disability Insurance[a]

	Probability		Expected Duration	
Explanatory Variables	Coefficient	Standard Error	Coefficient	Standard Error
Age of onset	0.23	0.08	-6.59	2.34
Married at onset	-0.26	0.10	7.44	2.94
Nonwhite	-0.15	0.07	4.27	1.89
Education (ten years)	-0.16	0.06	4.48	1.79
Accommodation	-0.19	0.07	5.40	1.89
Experience (100 months)	-0.27	0.11	8.03	3.25
White collar	-0.06	0.04	1.73	1.27
Comorbidity	-0.11	0.05	-3.12	1.41
Cardiovascular	0.11	0.05	-3.12	1.41
Musculoskeletal	-0.43	0.15	12.28	4.34
PIA (annual; $1,000)	0.18	0.06	-5.28	1.87
Expected earnings (annual; $1,000)	-0.06	0.02	1.85	0.68

[a]These calculations are based on the semiparametric estimator using the weighted data set (i.e., using the Manski-Lerman correction). The expected duration is 16.8 years, and 31.7 percent of the distribution is estimated not to apply in 30 years.

Appendix Table A-1. Predicted Labor Earnings of Persons with Truncated Social Security Records

	HIS Sample							
	Adjusting the Maximum Alone				Mean of the Truncated Log Normal			
Year	1st	2nd	3rd	4th	1st	2nd	3rd	4th
1951	14400	7200	4800	3600	21337	9634	5812	4144
1952	14400	7200	4800	3600	26455	9913	5856	4156
1953	14400	7200	4800	3600	28773	10015	5876	4163
1954	14400	7200	4800	3600	30074	10030	5876	4162
1955	16800	8400	5600	4200	25809	11364	6808	4844
1956	16800	8400	5600	4200	26774	11455	6826	4850
1957	16800	8400	5600	4200	32222	11655	6852	4856
1958	16800	8400	5600	4200	32783	11664	6852	4856
1959	19200	9600	6400	4800	28552	12919	7772	5535
1960	19200	9600	6400	4800	31241	13092	7795	5540
1961	19200	9600	6400	4800	30522	13059	7792	5539
1962	19200	9600	6400	4800	41224	13402	7840	5551
1963	19200	9600	6400	4800	51973	13533	7858	5556
1964	19200	9600	6400	4800	47074	13516	7860	5558
1965	19200	9600	6400	4800	46126	13504	7858	5557
1966	26400	13200	8800	6600	43871	18052	10727	7620
1967	26400	13200	8800	6600	44467	18107	10739	7624
1968	31200	15600	10400	7800	45767	20929	12618	8990
1969	31200	15600	10400	7800	55547	21512	12706	9013
1970	31200	15600	10400	7800	49288	21225	12667	9004
1971	31200	15600	10400	7800	64290	21722	12731	9018
1972	36000	18000	12000	9000	74218	25049	14686	10404
1973	43200	21600	14400	10800	72668	29517	17539	12461
1974	52800	26400	17600	13200	73347	34803	21228	15171
1975	56400	28200	18800	14100	83553	37670	22743	16220
1976	61200	30600	20400	15300	90150	40923	24702	17612
1977	66000	33000	22000	16500	100781	44342	26658	18994
1978	69200	34600	23067	17300	101875	46267	27931	19914

Appendix Table A-1. Continued

	Administrative Records Sample							
	Adjusting the Maximum Alone				Mean of the Truncated Log Normal			
Year	1st	2nd	3rd	4th	1st	2nd	3rd	4th
1951	14400	7200	4800	3600	23901	9811	5840	4152
1952	14400	7200	4800	3600	24801	9873	5853	4156
1953	14400	7200	4800	3600	27112	9977	5871	4162
1954	14400	7200	4800	3600	34284	10093	5885	4164
1955	16800	8400	5600	4200	30147	11572	6837	4851
1956	16800	8400	5600	4200	30604	11608	6845	4854
1957	16800	8400	5600	4200	31052	11639	6852	4857
1958	16800	8400	5600	4200	33767	11688	6855	4857
1959	19200	9600	6400	4800	32354	13166	7809	5544
1960	19200	9600	6400	4800	36053	13294	7826	5548
1961	19200	9600	6400	4800	37772	13334	7831	5549
1962	19200	9600	6400	4800	41753	13425	7845	5553
1963	19200	9600	6400	4800	42763	13448	7849	5555
1964	19200	9600	6400	4800	39873	13427	7850	5556
1965	19200	9600	6400	4800	40779	13455	7856	5558
1966	26400	13200	8800	6600	39798	17854	10708	7618
1967	26400	13200	8800	6600	38222	17738	10696	7617
1968	31200	15600	10400	7800	42797	20662	12592	8989
1969	31200	15600	10400	7800	45098	20948	12637	9000
1970	31200	15600	10400	7800	50612	21330	12686	9010
1971	31200	15600	10400	7800	55476	21558	12720	9019
1972	36000	18000	12000	9000	55595	24411	14605	10387
1973	43200	21600	14400	10800	68875	29297	17503	12451
1974	52800	26400	17600	13200	77651	35162	21266	15175
1975	56400	28200	18800	14100	80391	37174	22630	16178
1976	61200	30600	20400	15300	92667	40620	24568	17547
1977	66000	33000	22000	16500	112920	44424	26582	18944
1978	69200	34600	23067	17300	126153	47045	27971	19900