



CHICAGO JOURNAL

Articles from 2013 and after
are now only accessible on
the Chicago Journals website at
JOURNALS.UCHICAGO.EDU



The Society of Labor Economists

NORC at the University of Chicago

The Economic Implications of Public Disability Insurance in the United States

Author(s): Patricia M. Danzon

Source: *Journal of Labor Economics*, Vol. 11, No. 1, Part 2: U.S. and Canadian Income Maintenance Programs (Jan., 1993), pp. S170-S200

Published by: The University of Chicago Press on behalf of the Society of Labor Economists and the NORC at the University of Chicago

Stable URL: <http://www.jstor.org/stable/2535171>

Accessed: 06-06-2016 16:00 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at

<http://about.jstor.org/terms>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



NORC at the University of Chicago, The University of Chicago Press, Society of Labor Economists are collaborating with JSTOR to digitize, preserve and extend access to *Journal of Labor Economics*

The Economic Implications of Public Disability Insurance in the United States

Patricia M. Danzon, *University of Pennsylvania*

A review of previous analyses of labor supply effects of Social Security Disability Insurance (DI) concludes that estimates of labor supply effects and net social costs are upward biased because they ignore interactions between DI and other insurances. A model of optimal insurance, postinjury accommodations, and labor supply shows that reduction in labor supply and increase in consumption when disabled do not necessarily imply moral hazard. Optimal postinjury accommodations vary inversely with firm size. The Americans with Disabilities Act will reduce wages and labor supply of healthy workers, particularly in small firms. Effects on labor supply of the disabled are ambiguous.

I. Introduction

Estimates of the number of disabled persons in the United States vary because disability is not an objective medical condition. Whether or not persons with a given medical condition consider themselves disabled and withdraw from the labor force depends on psychological, economic, and social factors. According to the 1988 Current Population Survey, 8.6% of the population aged 16–64 years, or 13.4 million people, were “work disabled” in 1988.¹ The percent disabled increases from 3.8% of 16–24-year-

¹ Persons are classified as having a work disability if they (1) have a health problem or disability that prevents them from working or that limits the kind or amount of work they can do; (2) have a service-connected disability or ever retired or left a job for health reasons; (3) did not work in the survey reference week or previous year because of long-term illness or disability; or (4) are under 65 and are covered by Medicare or receive Supplemental Security Income. See U.S. Bureau of the Census (March 1989), table 592.

[*Journal of Labor Economics*, 1993, vol. 11, no. 1, pt. 2]
© 1993 by The University of Chicago. All rights reserved.
0734-306X/93/1101-0016\$01.50

Table 1
Persons with Work Disability, 1988

Age	Total	Male	Female	White	Black	Hispanic
16-24	1,285 (3.8)	674 (4.1)	610 (3.6)	963 (3.5)	291 (6.1)	125 (3.7)
25-34	2,414 (5.6)	1,249 (5.9)	1,165 (5.4)	1,874 (5.2)	464 (8.8)	189 (5.0)
35-44	2,455 (7.1)	1,308 (7.7)	1,147 (6.5)	1,957 (6.6)	433 (11.7)	217 (8.3)
45-54	2,443 (10.3)	1,190 (10.3)	1,252 (10.2)	1,852 (9.1)	502 (20.1)	190 (12.8)
55-64	4,825 (22.3)	2,285 (22.4)	2,540 (22.2)	3,898 (20.4)	822 (39.6)	290 (25.4)
Total	13,420 (8.6)	6,706 (8.7)	6,714 (8.4)	10,544 (7.9)	2,512 (13.7)	1,011 (8.2)
Percentage of work disabled receiving:						
Social Security income	29.5	31.9	27.2	29.7	30.0	23.4
Medicaid	21.6	18.1	25.2	17.8	37.0	32.8

SOURCE.—U.S. Bureau of the Census (1989), table 592.

NOTE.—000 omitted. Percent work disabled of total population in parentheses. Covers civilian non-institutional population. For definition of work disability, see text. Total includes other races not shown separately.

olds to 22.3% of 55-64-year-olds (table 1). Age-specific disability rates are almost twice as great for blacks as for whites.² Disability is inversely related to schooling (Berkowitz and Hill 1986).

Disability insurance in the United States is provided through a network of public and private programs (table 2).³ The Social Security Disability Insurance (DI) program has received the most attention. Disability Insurance expenditures increased from \$3.1 billion in 1970 to \$20.5 billion in 1987. However, as table 2 demonstrates, other programs have grown at least as rapidly as DI since 1970, and it now accounts for less than 20% of total expenditures under public or quasi-public programs. Supplemental Security Income (SSI), a means-tested income-support program for the aged, blind, and disabled that was established in 1974, paid out \$14.8 billion by 1987. Cash payments under workers' compensation (WC) have grown as rapidly as DI. Even more dramatic is the growth in liability insurance premiums, which can be viewed as a form of compulsory third party insurance. Auto liability insurance premiums were almost 2.5 times DI payments in 1987. Premiums for other liability, which includes product liability

² In 1988 almost 30% of the work disabled received Social Security income. There is very little difference between the percent of whites (29.7%) and blacks (30.0%) receiving Social Security income, in contrast to the almost twofold difference in percent reported work disabled.

³ The estimates for private insurance are downward biased because data on self-insured plans and payments under pension plans are not available.

Table 2
Trends in Expenditures on Disability Insurance Programs in the United States (Current \$ Billions)

	1970	1975	1980	1985	1987
SSDI*	3.09	8.51	15.52	18.83	20.52
Supplemental Security Income (SSI)	0	5.9	7.9	11.1	14.8
Workers' Compensation:†	2.0	4.6	9.7	15.1	16.7 [§]
WC total‡	NA	NA	13.56	22.47	25.02 [§]
WC premiums	3.49	6.19	14.24	17.05	23.43
Liability insurance:¶	11.10	17.30	31.01	50.40	74.09
Auto liability*#	8.96	13.32	23.32	36.09	49.21
Other liability**	2.14	3.98	7.69	14.31	24.88
Private insurance††	1.82	2.71	5.34	5.63	6.26

* U.S. Bureau of the Census (March 1989), tables 577, 568.

† Cash payments, U.S. Bureau of the Census (March 1989), table 568.

‡ Cash and medical payments. Includes payments under Longshoremen's and Harbor Workers' Act, Defense Bases Compensation Act, and Black Lung. U.S. Bureau of the Census (March 1989), table 590. § 1986.

¶ Premiums written.

Excludes auto physical damage.

** Includes medical malpractice.

†† Claim payments on short-term and long-term disability income loss policies by insurance companies. Group and individual HIAA (1989). For 1970 and 1975, data on individual policies not available.

and medical malpractice, grew from \$2.1 billion in 1970 to \$24.9 billion in 1987, or 20% more than payments under DI.⁴

Previous economic analyses of disability insurance in the United States have focused on the effects of DI on labor force participation. One purpose of this paper is to point out the limitations of such estimates for drawing welfare implications. A second purpose is to analyze the effects of the more rapidly growing form of social insurance for the disabled in the United States: antidiscrimination legislation that requires access to employment and public facilities and services for the disabled. The Rehabilitation Act of 1973 prohibits discrimination on the basis of handicap in any program or activity receiving federal assistance or by the federal government. Section 503 requires businesses with federal contracts of \$2,500 or more to take affirmative action to employ and advance qualified handicapped persons and to make "a reasonable accommodation to the physical and mental limitations of an employee or applicant." Enforcement is through the Office of Federal Contract Compliance Programs (OFCCP) of the Department of Labor, which can sue the company on behalf of the individual; section 503 cannot be enforced by private lawsuit, unlike other sections.

⁴ Liability insurance premiums are not strictly comparable to DI benefit payments. Liability premiums reflect the discounted present value of expected benefits payable on policies written in the calendar year, plus overhead expenses and return on capital. The DI benefit payments are on a pay-as-you-go rather than an accrual basis and omit overhead costs, including deadweight costs of raising tax revenues. Danzon (1992) discusses bias in comparing costs of public and private insurance.

The 1990 Americans with Disabilities Act vastly expanded the potential impact of antidiscrimination legislation. Hailed as the most sweeping civil rights legislation since 1964, the act extends the ban on employment discrimination on the basis of mental or physical disability to all employers with 15 or more employees by 1994 and requires access to public buildings, telephone services, mass transportation, government, and privately provided services such as restaurants, movie theaters, etc. Disability is defined to include persons with AIDS or who test HIV positive and those "who may be regarded as disabled by others." The act excludes disability related to use of illegal drugs. Employers are required to provide "reasonable accommodation" for the disabled, although "undue burden" is a defense against the requirement to retrofit existing buildings. In contrast to previous legislation, individuals may now bring suit on their own behalf.⁵

Mandating that employment and public facilities be available to the disabled is an extension of publicly mandated disability insurance. To the extent that the new legislation simply shifts the costs of accommodations from the disabled to others, the effects would be purely distributive. However, the analysis below indicates that changing the liability rule is likely to impose significant net costs and excess burdens. As a rough estimate, if the new legislation imposed a cost of \$10,000 on each of the 725,000 private establishments with 20 or more employees,⁶ the cost of accommodations to employers alone would be \$7.2 billion. Total costs could be much higher if the excess burden of distortions in labor supply, costs to providers of transportation and other services, litigation, and other overhead costs are included.⁷

In this paper Section II reviews the structure of the DI program and evidence of its effects on postdisability labor supply. Section III evaluates the full net economic impact of the DI program incorporating insurance and distributive effects. Section IV analyzes the effects of antidiscrimination legislation. A theoretical model of optimal levels of disability insurance, hours of work, and productivity-enhancing accommodations by employers is developed. The model shows that some reduction in labor supply is

⁵ In cases involving a pattern of discrimination, the attorney general can seek compensatory (but not punitive) damages and civil penalties of up to \$100,000 for repeated violations (U.S. Congress 1990, S933, sec. 308).

⁶ U.S. Bureau of the Census 1989, 1986 data, table 859.

⁷ Advocates of the disability community argue that "a tremendous percentage of people with disabilities really want to work and have the skills to work." Management attitude and misconceptions are cited as the major obstacles to employment. Several studies have reported minimal additional costs for special training or facilities for hiring the disabled (Collignon 1986). However, such studies based on voluntary hirings are subject to severe selectivity bias as a basis for estimating the potential costs of the new legislation, since voluntary hiring would be most likely to occur in precisely those cases where the productivity of the disabled worker is above average and costs of accommodation are below average.

consistent with optimal (full information) insurance and entails no dead weight loss, contrary to the implicit presumption of much of the literature on DI. The extent of market failure and likely effects of government-mandated accommodations are discussed. The paper concludes that the excess burden, per dollar of benefit to the disabled and in total, could be considerably greater under antidiscrimination statutes than under the Social Security Disability Insurance program.

II. The Social Security Disability Insurance Program

A. Program Structure

The Social Security Disability Insurance Program (DI) is similar in structure to the Canadian Disability Insurance program (see Maki, in this issue). The DI program was established in 1956 as an amendment to the Social Security Old Age and Survivors (OASI) Act and adopted some of the characteristics of OASI. Eligibility is not universal but is conditioned on having worked a minimum number of quarters in covered employment, including immediately prior to disability. The 1965 amendments significantly expanded eligibility under the program by broadening the requirement of “permanent” disability to “expected to last at least one year,” and by extending benefits to workers under 50.

Compensation under DI follows the OASI benefit structure. The monthly benefit is a piecewise linear function of the worker’s average indexed monthly earnings (AIME) over years in covered employment.⁸ Specific details of the formula, including adjustments for inflation, have changed over the years. Since 1983, the use of wage indexing for prior earnings, maximum taxable earnings, and for the bendpoints in the formula imply that real replacement rates (initial benefit relative to average indexed monthly earnings) have remained constant. Dependents of beneficiaries receive 50% of the insured’s benefit, subject to a maximum family benefit.⁹ Since 1978, replacement rates, inclusive of other federal, state, and local government transfers and workers’ compensation, are capped at 80% of pretax earnings. Prior to 1978, low wage workers with dependents may have exceeded this limit, and enforcement may remain incomplete. Disability Insurance beneficiaries become eligible for Medicare after 2 years on DI.

The DI program is financed by a payroll tax of 1.2% on wages up to the taxable maximum (1990 rates) for employees in covered employment. The number of covered workers has increased over time with the expansion

⁸ For 1990, the replacement rate is $.9(\$356) + .32(\$2,145 - \$356) + .15(\text{AIME} - \$2,145)$.

⁹ Disabled dependents of covered workers who are not disabled are not entitled to benefits unless they have established eligibility by their own work history.

of the OASI program. Although the federal government defines the structure and financing of the program, it is implemented by state agencies.

Table 3 shows trends in the number of covered workers, beneficiaries, and average benefit levels. The number of disabled workers receiving DI benefits increased sharply from 1.5 million in 1970 to 2.86 million in 1980. During the 1980s the number of DI worker beneficiaries fell initially, reflecting the tightening of eligibility requirements, but it has drifted up slightly since the low of 2.57 million in 1983. Average monthly benefits to disabled workers increased from \$380 in 1970 to \$508 in 1987; however, this average reflects the mix of recipients and is not a pure measure of changes in worker-specific expected benefits. If Medicare benefits paid to DI recipients are added to these cash payment figures, the total (income support and medical) is comparable to the \$25 billion paid in workers' compensation benefits in 1987, and the rate of growth has been similar. However, since the number of covered workers potentially eligible for DI has increased more rapidly than for WC, the rate of growth of beneficiaries per covered worker has been less under DI than under WC.

Since 1974, the means-tested SSI program has provided income support for low-income blind and disabled persons under 65. In principle, SSI benefits are coordinated with DI benefits, with a dollar-for-dollar offset of DI benefits. Supplemental Security Income beneficiaries are eligible for the means-tested public medical program, Medicaid, which would cover their medical expenses during the first 2 years on DI and (in most states) would cover Medicare's copayments and premium contributions.

B. Evidence on Labor-Supply Effects of DI

From 1959 to 1980, the labor force participation rate of males aged 45–59 fell from 96% to 88.5%. This parallels an expansion of number of covered

Table 3
Trends in Social Security Disability Insurance Recipients (Millions)

	Insured Workers*	Disabled Workers†	Disabled Workers and Dependents	Disabled/Covered Workers (%) (3 ÷ 1)	Average Monthly Benefit
1965	53.3	.99	1.74	1.95	...
1970	72.4	1.49	2.66	2.06	380
1975	83.3	2.49	4.35	2.99	470
1980	98.0	2.86	4.68	2.92	496
1985	109.2	2.66	3.91	2.43	511
1987	113.3	2.79	4.05	2.46	508

SOURCES.—U.S. Bureau of the Census (March 1989), tables 572, 575, 578, 579, 176; HIAA (1989), Source Book of Health Insurance Data.

* Insured for disability.

† Disabled workers under age 65.

Table 4
Labor Supply Effects of Social Security Disability on Males

Study	Data Set	Sample Size	Sample Analyzed	Results
Parsons (1980 <i>a</i>)	National Longitudinal Survey	3,219	Age 48–62 years in 1969	Elasticity of labor force nonparticipation .63
Haveman and Wolfe (1984 <i>b</i>)	Panel Study of Income Dynamics	741	Age 45–62 years in 1978	Elasticity of labor force nonparticipation .06–.21
Slade (1984)	Longitudinal Retirement History Survey	5,403	Age 58–63 years in 1969	Elasticity of labor force nonparticipation .81
Leonard (1979)	Social Security Survey of Health and Work Conditions	1,685	Age 45–54 years in 1972	Elasticity of beneficiary status .35

SOURCE.—Reproduced from Leonard (1986).

workers and increase in real benefit levels under DI and probably more lenient application of eligibility criteria. From 1968 to 1978 the number of DI recipients grew at a rate of 7% a year.¹⁰ Several studies have attempted to measure contribution of DI to this decline in labor force participation of prime-aged males (e.g., Parsons 1980*a*, 1980*b*, 1984*b*; Haveman and Wolfe 1984*a*, 1984*b*; Leonard 1979). Since reviews of this literature already exist (Danziger, Haveman, and Plotnick 1981; Leonard 1986), this review focuses on the main issues and findings, for purposes of comparison with Canada.

While the consensus is that DI benefit levels and screening criteria affect the number of DI recipients directly and labor force nonparticipation (LFNP) indirectly, the magnitude of the effect remains uncertain, despite numerous studies. Table 4 summarizes some of the main results. Estimates of elasticity of LFNP with respect to benefit levels range from 0.06 to 0.81 and higher.

1. Methodological Issues

The range in estimates of effects of DI reflects differences in sample (cohort effects), type of data (time series, cross-sectional, and longitudinal

¹⁰ From 1957 to 1978 the wage replacement rate under DI increased from 30% to 41% for the average nonsupervisory manufacturing worker with no dependents, and from 57% to 68% for the same worker with a wife and child (Haveman and Wolfe 1984*b*).

panel), and methodologies used. None of the estimates is conclusive, largely because of intrinsic data limitations.

The underlying model is usually a one-period model in which the worker's decision to withdraw from the labor force depends on expected utility or income under the two alternatives:

$$\text{LFNP} = f(W^d, B, X),$$

where W^d = expected income if in the labor force, B = expected income if out of the labor force, including DI benefits, and X = a vector of other individual characteristics.

Life-cycle effects.—This one-period model ignores the life-cycle implications of labor force withdrawal, including depreciation of human capital and signaling effects. Asset accumulation may also be affected because of pension accrual. If these life-cycle losses are higher for high wage workers, estimated effects of DI are likely to be upward biased since DI replacement rates are inversely related to income.¹¹

Cohort effects.—If DI were actuarially fair, its effect on labor supply would derive solely from the change in the relative price of leisure in the two states, healthy or disabled. There would be no life-cycle income effect. In fact, because DI benefits are financed from taxes on current workers, the first cohorts that became eligible for benefits received a massive net windfall. For later cohorts, the net income effect is positive only for low wage workers, because of proportional tax financing but declining marginal replacement rates. Thus, other things equal, larger elasticities of nonparticipation with respect to DI benefits are expected for early cohorts and low wage workers. The estimates in table 4 are consistent with a decline in the estimated elasticities of nonparticipation for more recent cohorts.

Unobservable wages and benefits.—A fundamental problem in estimating labor force participation (LFP) effects of DI is that W^d is not observed for labor force nonparticipants and B not observed for labor force participants and nonparticipants who either do not apply or are rejected. Different studies have used different techniques to address this problem, but even with the best possible methods applied to the available data, the resulting estimates may nevertheless be biased, for reasons discussed below.

Identification of wage and benefit effects.—The variable B is a piecewise linear function of average monthly earnings while working, with a declining marginal replacement rate,¹² and most workers are on the last two segments. Thus differences in the relationship between B and predisability wage W^d

¹¹ Sickles and Taubman (1986) include forgone OASI benefits in their analysis of health and retirement decisions of the elderly but do not distinguish effects of DI.

¹² See n. 8 above.

arise only as a result of differences in wage history over the working years, dependent benefits, and truncation by the minimum and maximum family benefit. Studies that use W^a as an instrument for W^d and impute B from the benefit structure and W^a , rather than the individual's lifetime work history and family status, cannot distinguish the effects of low W^d from the effect of a relatively high replacement rate B/W^d . If the analyst's proxy for W^d is more upward biased for low wage workers—for example, because disability reduces productivity more in jobs requiring physical strength—then the measurement error in the predicted wage W^d is positively correlated with B/W^d and the coefficient is likely to be upward biased.

Control for health status.—Imperfect control for the effects of disability on nonmarket productivity and real full income may lead to biased estimates of the effects of B .

In empirical work, measures of health status are implicitly proxies for nonmarket productivity. Results are sensitive to the health measures used. A widely used dummy variable for individuals who report that disability limits their ability to work is arguably endogenous. Lambrinos (1980) shows that including a more comprehensive vector of health status indicators or partitioning the sample into healthy and unhealthy individuals leads to higher estimates of the hours substitution elasticity of the disabled, but results are very sensitive to the specification. He attributes the higher hours elasticity of the disabled to the fact that they normally work shorter hours: in his sample, mean hours of work for the disabled is 894 hours, compared to 1,804 hours for the healthy.

If response to DI varies systematically with health status, then specifications that fail to control for attrition of the most disabled from the labor force may further bias estimates.¹³

Asset substitution.—The fact that private long-term disability (LTD) and private pension coverages are typically written as supplemental policies to Social Security and other public income transfers is direct evidence that private and public insurance are substitutes. Indeed, the low rates of LTD and private pension coverage for low wage workers are probably attributable to the fact that DI and OASI provide replacement rates for such workers at least equal to the maximum offered by private insurers (.60). To the extent that DI has simply replaced private insurance, long-run effects of DI on LFNP are overestimated. Most studies ignore private assets; those that have included a private asset measure (e.g., Haveman and Wolfe 1984b) do not model its endogeneity. Since the displacement of private assets is positively correlated with the DI replacement ratio B/W^d , this is likely to bias upward the estimated effect of DI.

Omitted fringe benefits.—None of the studies controls for the value of Medicare if the individual qualifies for DI or the expected value of private

¹³ Leonard (1979, 1986) demonstrates the effects of attrition bias.

fringe benefits if the individual remains in the labor force. Since low wage workers are less likely to receive employer-provided health insurance or other fringe benefits, the downward bias due to using B/W as a proxy for the full replacement rate, including noncash benefits, is greater for low wage workers. This will further upward bias the estimated effect of DI benefits on LFNP. The extrapolation from cross-section to time-series effects may also be biased by omitting fringe benefits, which are an increasing share of total private compensation in the 1970s.

Expansion of OASDI coverage.—Since 1957 the under-65 population potentially eligible for DI has increased, due to expansion of covered employment, lowering the age of eligibility and adding dependent benefits. Time-series analysis or projections of the effects of DI should distinguish an increase in the beneficiary rate from a given pool of eligibles from expansion of the pool. Similarly, cross-sectional studies that estimate B from the nominal benefit structure, without control for worker-specific eligibility, may yield upward-biased estimates of response if low wage workers were more likely to be in covered sectors.

Expansion of Workers' Compensation and SSI.—The low probability of being accepted for DI, at least on initial application, implies that labor force withdrawal in the hope of qualifying for DI is a risky strategy: DI alone has a large deductible of indefinite duration. Once on DI, there is an additional 2-year wait for Medicare eligibility. For workers who can show that disability is work related, this deductible is covered by workers' compensation which pays wage loss and medical expense, with a much shorter waiting period (typically 1 month or less) which is retroactively restored if the disability lasts more than a certain period. The expansion of WC coverage and increase in WC benefit levels in the 1960s and 1970s imply a reduction in the downside risk of labor force withdrawal and hence an increase in the expected benefits from LFNP. Similar effects are implied by the introduction of SSI in 1974, which provides means-tested benefits for the disabled and eligibility for Medicaid. In cross-sectional analysis, failure to control for WC and SSI would bias upward estimates of DI effects, since WC replacement rates and SSI benefits are positively correlated with the DI replacement rate.¹⁴

2. Empirical Estimates

Parsons (1980a) estimates a cross-sectional labor force participation equation for 1969, using a sample of 3,219 men from the National Longitudinal Survey who were 45–59 in 1966. Explanatory variables are proxies

¹⁴ In most states, the nominal WC replacement rate is truncated by a minimum and a maximum benefit; the maximum benefit implies sharply declining replacement rates at higher wage levels and in some states is binding for all workers above the average wage.

for DI benefits and other welfare benefits, both measured relative to an estimated wage; an index of mortality between 1969 and 1976, as a proxy for health status; age; and unemployment. Since 1969 wages are not observed for nonparticipants, Parsons uses 1966 wages as a proxy for 1969 potential wages and omits persons not in the labor force in 1966. The 1966 wage is also used to construct an estimate of B , using DI benefit tables.

Parsons estimates an elasticity of nonparticipation with respect to the replacement rate of 0.63. Applying the coefficients from the cross-section regression to changes in the dependent variables over time, Parsons concludes that DI can explain much of the decline in labor force participation of 45–54-year-old men between 1950 and 1980, and differences between blacks and whites.

These estimates of the effects of DI are likely to be upward biased for reasons already mentioned, several of which are recognized by Parsons. First, using W^a when working as a proxy for potential wage when disabled W^d is likely to introduce measurement error that is proportionately greater for low wage workers and is therefore correlated with B/W^d . Second, by the construction of B , all variation in the ratio B/W^d reflects variation in the potential wage. The moral hazard effect of high B/W^d cannot be distinguished from the effect of low W^d . Third, failure to control for private assets, WC, Medicare, and DI eligibility may all induce upward bias. As Parsons notes, the elimination of those not in the labor force in 1966, who might also be the most severely disabled, may also lead to upward-biased estimates of response.

Some evidence to support the hypothesis that Parsons's estimated elasticity is upward biased comes from John Bound's (1985) study of a sample of nonapplicants from the 1972 Social Security Survey of Disabled and Non-Disabled Adults. Using a similar specification applied to the nonapplicants, whose behavior is unlikely to reflect the influence of DI (unless they anticipated applying for DI later), Bound estimates an elasticity of 0.88, higher than Parsons's 0.63 estimate.

Haveman and Wolfe (1984*b*) use a sample of 741 men aged 48–62 from the 1978 Michigan Panel Study of Income Dynamics. They use a composite measure of disability income, which includes DI, SSI, and other public transfers and veterans and other disability pensions, which presumably includes private pensions. There is no attempt to distinguish net effects of DI. They use a three-step estimation procedure, following Lee (1979). The first step is a probit equation for labor force participation that does not include measures of expected income in and out of the labor force. This is used to impute a selectivity correction that is used in the potential earnings and disability income equations. At the third step the predicted incomes from these equations are included in a probit equation for predicting labor force participation. This technique requires the very strong specification

and identification assumptions, that some variable affects W and B but not labor force participation.¹⁵

The resulting elasticities for labor force nonparticipation range from .06 to .21; the lower estimates are obtained when additional controls are added for dependent benefits, selectivity, occupation, and other household income. The estimated response is extremely nonlinear, with higher elasticities for the more disabled and low wage earners. Although the elasticity of labor force participation with respect to disability income calculated at sample means is very low (between $-.0003$ and $-.0005$), their simulations (Have-
man and Wolfe 1984*b*, table 5) imply that a 25% increase in expected disability income would increase the disability reciprocity rate by 1.04 percentage points or an elasticity of 0.55.¹⁶

Slade (1984) analyzes a sample of men aged 58–63 from the 1969 Longitudinal Retirement History Survey. Wages for nonparticipants are imputed from a wage equation using a sample selection correction, and an earnings history is used to impute DI benefits. He finds an elasticity of nonparticipation with respect to DI benefits of .81, and .90 when the sample is restricted to those with self-reported health limitations, higher than Parsons's (1980*a*) estimate of .63. Note that this is consistent with the hypothesis of a larger income effect for earlier cohorts.¹⁷ Again, the estimates are suspect because of the difficulty of imputing a potential wage for the disabled and the intrinsic correlation between high DI replacement rates and low potential wages.¹⁸

Leonard (1979) uses a sample of 1,685 men aged 45–54 from the 1972 Social Security Survey of Health and Work Characteristics, merged with the Social Security earnings history, to estimate the effect of expected DI benefit levels on the probability of DI reciprocity. The structural model is

$$SSDK = F(SSD, W, X) + e,$$

$$W = Zb + u,$$

¹⁵ As Leonard (1986) points out, application of a Heckman selectivity correction is more complex in this context because it requires estimating the probability of observing a positive wage, which is closely related to estimating the probability of not being a DI recipient, which is what is ultimately to be measured.

¹⁶ See Bound (1985) and Leonard (1986) for discussion on this. Note that elasticities of LFNP with respect to benefit levels may differ from elasticities of reciprocity rates.

¹⁷ Optimal investment in human capital would also predict larger effects for older age groups. In order to distinguish pure age (human capital) from cohort (income) effects it would be necessary to compare age-specific elasticities for different cohorts.

¹⁸ Use of actual earnings history does not necessarily reduce the correlation because earnings history reflects both actual wages while working and intermittent pre-application labor force participation, which may indicate poor health and low potential wage once disabled.

and

$$p = Hg_1 + Xg_2 + m,$$

where

SSDK = probability of being a DI recipient,

SSD = expected DI benefits = $p \times B$,

W = expected labor market income,

p = probability of eligibility,

H = health status,

X, Z = vectors of individual background characteristics, and

e, u, m = error terms.

Leonard forgoes identification of the coefficient of W in the beneficiary equation, using indirect least squares:

$$\text{SSDK} = F(\text{SSD}, Zb + u, X) + e.$$

Expected DI benefits (SSD) are estimated as the product of de jure benefits if eligible (based on actual earnings record and number of dependents) and the probability of eligibility which is estimated from the third equation using a sample of recent applicants. As Leonard (1986) points out, the coefficient on this variable may be upward biased if errors in the eligibility and beneficiary equations are positively correlated. Health status is measured by a set of 27 dummy variables for specific health conditions.

Leonard estimates that a \$180 increase in annual benefits will increase the proportion of DI beneficiaries in the population by one percentage point, or an elasticity of 0.35. The probability of being a beneficiary is negatively related to education and to having predisability wages above the ceiling. He attributes this to greater job opportunities for partially disabled white-collar workers than for blue-collar workers. He concludes that the probability of eligibility is significantly lower for nonwhites. However, given the lower expected benefits corresponding to a given level of de jure benefits, race has no effect, after controlling for job opportunities and poor health. Applying these cross-section estimates to the time-series data, Leonard estimates that DI accounts for roughly half of the decline in male labor force participation between 1957 and 1975 or 1.8 of the 3.5 percentage points. Although Leonard's estimates are probably the most reliable, given the available data, the potential for upward bias remains, given the difficulty of distinguishing net effects of DI and other correlated factors.

If the cross-section evidence is discounted because of problems of distinguishing DI effects from those of low wages, what other factors might

explain the decrease in LFP, other than the simultaneous increase in number of DI beneficiaries? Leonard (1986) rejects arguments based on changes in marital status, structural unemployment, weak prior labor force attachment, veteran status, and increase in wives' labor force participation. Other possible contributing factors, which are not explicitly included in these studies, are the growth in WC benefit levels and duration of coverage for permanent total disabilities, the introduction of SSI and Medicaid, and private LTD and pensions benefits. In particular, the sharp increase in DI reciprocity rates between 1970 and 1975 may have been stimulated by introduction of SSI and the sharp 1972 increase in DI benefit levels, including a flaw in the indexing adjustment which created strong incentives for early retirement. Since this error was recognized but not corrected until 1977, there would have been an incentive to take advantage of a benefit structure that was temporarily higher than might be expected for the long run. The transitory effects of these events should be distinguished from the permanent effects of DI.

More recent experience suggests that screening stringency as well as benefit levels contribute to the DI applicancy rate and hence to the declining LFP of prime-aged males. Since 1980, the number of DI recipients has remained stable or even fallen, despite a modest increase in average monthly DI benefit levels (see table 2).¹⁹ The decline in DI reciprocity rates in the early 1980s partly reflects tightening of the screening process under the Reagan administration. Several studies have demonstrated the sensitivity of DI applicant rates to the acceptance rate. Marvel (1982) finds that cross-state differences in application rates are related to denial rates. Halpern and Hausman (1986) confirm that the probability of application is sensitive to the denial rate but conclude that "potential applicants seem more sensitive to the benefit level than to the probability of acceptance." None of these studies has attempted to measure the extent to which growth in reciprocity in the 1970s reflects a looser screening process as opposed to increased benefit levels.

Given the intrinsic difficulties in estimating directly the contribution of DI to the decline in labor force participation, other indirect evidence has been brought to bear. Several studies have looked at the work history of rejected applicants. Treitel (1976) found that 39.7% of male applicants who were initially denied benefits in 1967 did not work at all in the subsequent 4 years, and only 24.1% worked for 12 or more of the 16 quarters. Bound (1985) studied a sample of rejected applicants from the 1972 Survey of Disabled and Non-Disabled Adults and the 1978 Survey of Disability and Work. He finds that the majority of rejected applicants do not return to sustained work. Less than 50% were earning at any given time, with

¹⁹ As noted earlier, the DI benefit structure has remained constant in real terms since 1983.

yearly earnings for those who worked less than half of the earnings of other working men their age. Assuming that rejectees would on average be in better health than successful applicants, he concludes that DI accounts for at most 1.75% of the 4.25% of men aged 45–54 who were out of the labor force, much lower than estimates based on extrapolation from cross-sectional analyses. Halpern and Hausman (1986) note that, while 7.7 million adults between ages 20 and 64 report that they were unable to work or unable to work regularly, only 26% received DI benefits.²⁰ Of course, because the application process is uncertain and the outcome depends on demonstrated inability to work, the number of actual DI recipients may understate the number whose labor force withdrawal is influenced by the potential for DI.²¹

3. *Conclusions*

It seems safe to conclude that DI has contributed to the decline in LFP of males, but the magnitude of the effect remains uncertain, as does the relative importance of the dollar benefits and the stringency of the screening process. The failure to resolve this issue reflects inherent limitations of the data available, not defects in the analytic methods used. The evidence strongly suggests that DI is subject to type I and type II errors—acceptance of some who could work and rejection of others who cannot. Even if DI does have a significant effect on LFP, measurement of the productivity loss should not simply extrapolate from productivity of able-bodied workers. Both theory and empirical evidence confirm that those most affected have relatively low potential wages even prior to disability. Further, the disabled who do work report average annual hours and yearly earnings that are less than half the annual hours and earnings of the nondisabled. Assuming that those who continue to work have higher potential wages and hours than those who drop out, this would be an upward-biased estimate of the loss in output due to labor force withdrawal of those on DI.

III. Net Economic Effects of DI

Although most prior studies of DI have focused on its effect on labor force participation, a full evaluation of the economic impact of DI must also consider its effects on insurance, incentives for injury prevention, postinjury investments to raise productivity of the disabled, and overhead costs including labor supply distortions from the payroll tax levies.²² Mea-

²⁰ In 1987 30% of the work disabled received some form of Social Security; this overstates DI reciprocity rates since it includes those 62–64-year-olds who received OASI benefits (see table 1).

²¹ For example, in 1980 35.7% of successful DI recipients had been rejected on their initial application.

²² In addition, there may be a distortion from substituting pay-as-you-go financed DI insurance for funded private insurance. The effects of Social Security on capital

surement of the overall welfare effects of DI must also address the prior question, Relative to what?

It is not practical or relevant to compare DI to the first best full information optimum. One possible benchmark is the outcome with no government intervention and reliance on private insurance markets, unconstrained by government but subject to imperfect information which may result in adverse selection, predisability moral hazard (not taking care), and postdisability moral hazard (exaggerating disability status). For any individual, the net cost of DI (C_{di}) is then the difference between expected utility under DI and private insurance:²³

$$C_{di} = EU[B(DI), t(DI), r(DI), s(DI), h(DI)] - EU[B_p, \pi_p, r_p, s_p, h_p],$$

where t = DI tax rate, r = prevention expenditures, s = postinjury investments, h = hours of work, subscript p denotes optimal values under private insurance, conditional on imperfect information, and π denotes the private disability insurance premium.

As already noted, WC and private insurance and pension contracts are explicitly written as supplements to DI.²⁴ To the extent that DI simply substitutes for equivalent other insurance, there are no net welfare effects, and econometric estimates of the labor supply effects of DI that do not control for endogeneity of other insurance are upward biased. Although these may be unbiased estimates of the effects of disability insurance more generally, no welfare loss is necessarily implied, relative to either a full information or an information-constrained optimum.²⁵

In fact DI differs from private insurance in several ways that must be considered in a full evaluation. First, the social welfare function may place some value on DI's intragenerational redistribution from workers with high lifetime earnings to workers with low lifetime earnings.²⁶ Second,

formation remain an unresolved issue (see, e.g., Feldstein 1974, 1982; Leimer and Lesnoy 1982).

²³ A third possible benchmark is the best feasible mixed public-private program, given information and administrative cost constraints but ignoring political constraints.

²⁴ The most plausible explanation for upper limits on duration or dollar benefits for permanent total disability, under WC and private LTD policies, is the high probability that the individual will have qualified for DI by the time the limit is reached.

²⁵ Measuring relevant private insurance coverage is not easy. Many individuals would not know the details of their LTD and private pension coverages. For defined contribution plans, asset value is the market value of the individual's contribution and is unaffected by the timing of withdrawal from the labor force. But the present value of defined benefit plans may be greater, the earlier the withdrawal from the labor force.

²⁶ The redistributive effects of DI are greater than those of OASI, although they have identical financing and benefit structures, because high wage workers

casual evidence suggests that the risk of adverse selection constrains the coverage available in the nongroup market for LTD coverage and raises its cost. If private insurance markets are quantity constrained because of adverse selection, mandating a minimum level of universal coverage while permitting individuals to purchase supplementary coverage can be Pareto improving.²⁷

Another advantage of DI relative to most private LTD coverages is that DI benefits are indexed. As long as private insurance markets do not offer fair indexed annuities, the utility value of public insurance programs such as OASI and DI which do provide indexed annuities exceeds their actuarial value Bernheim (1987).²⁸

On the other hand, some characteristics of DI entail excess burdens compared to private insurance. Since DI tax levies do not experience rate industries or employers, in contrast to group LTD or WC, DI undermines employers' incentives for injury prevention and postinjury rehabilitation and distorts relative prices and industrial mix.²⁹ Second, if individuals for whom DI is unfair do not value the redistribution, the difference between their tax contributions and expected benefits should be viewed as an effective tax,³⁰ with associated tax-induced deadweight costs.

Third, for low wage workers $B(DI)$ may exceed the optimal level B_p , given moral hazard, and public insurance has no information advantage over private insurance to control moral hazard. Since 1978, DI benefits (including dependents' benefits) combined with workers' compensation and other governmental disability benefits cannot exceed 80% of "average current earnings," which for most workers is the average monthly earnings for the highest year in the 6 years covered employment immediately prior to disability.³¹ However, even if $B(DI)$ is less than full replacement of after-tax income plus fringe benefits prior to disability, it may considerably exceed potential after-tax earnings after disability, if disability reduces potential wage rates. This is not necessarily inconsistent with first best insurance under full information (see below), but is more likely to be excessive under imperfect information and moral hazard. Although optimal disability benefits under imperfect information cannot be determined a

are less likely to claim disability benefits, whereas they have higher expected duration of retiree benefits.

²⁷ The mandatory coverage could be privately or publicly provided.

²⁸ However, OASI and DI do not necessarily provide the optimal intergenerational allocation of nondiversifiable risk.

²⁹ For an analysis of the industry mix component of deadweight costs induced by flat rating in the Swedish social insurance scheme, see Hansson, Lyttkens, and Skogh (1984).

³⁰ See Browning (1985) for estimates of the effective OASI tax rate.

³¹ This does not include income from private insurance. Private insurers typically set a .60 cap on the replacement rate, including benefits from all sources.

priori, the fact that most private insurance coverage sets a cap of two-thirds or less suggests that $B(\text{DI})$ for low-income workers exceeds efficient levels, given moral hazard. The labor supply effects of this excess coverage above information-constrained efficient levels can be considered a net cost of DI or a trade-off between efficiency and equity.³²

Finally, applicant screening and periodic review may be less efficient under DI than with private insurance. Since applicant screening occurs at the state level whereas funding is fully from federal sources, most plausible models of bureaucratic decision making would predict a bias toward excessive acceptance rates, given the information available. Screeners are likely to attach a higher cost to rejection of valid applicants than to acceptance of invalid applicants.³³

To get some indication of the net welfare effects of DI, consider a simple case where r and s are unaffected by the insurance program. A worker faces an exogenous risk of disability p that entails a loss of income. Expected utility in the absence of insurance is

$$EU = (1 - p)U(Y^a) + pU(Y^d).$$

With the introduction of DI, the worker pays a proportional tax on labor income at rate t . If he becomes disabled, assume that the choice of continuing to work and earning income Y^d is dominated by withdrawal from the labor force, receiving benefit B and an increase in leisure. This gives a lower bound for the utility value derived from the option of labor force withdrawal offered by the DI program. The net change in utility from the introduction of DI is:

$$(1 - p)U(Y^a - tw^a h^a) + pU(Y^d + B - k) - (1 - p)U(Y^a) - pU(Y^d),$$

where

h^a = hours of work if healthy,

w^a = wage rate if healthy, and

k = the reduction in labor income due to withdrawal from the labor force, net of the dollar value of the increase in leisure.

With actuarial insurance, $pB = (1 - p)twb$ (ignoring discounting). If the individual were risk neutral, the net loss is $-pk$, where k is the earnings loss due to postinjury reduction in hours of work, net of the dollar value of increased leisure. Econometric estimates of the DI-induced earnings loss

³² Some studies, e.g., Leonard (1986), implicitly assume that all insurance-induced reduction in labor supply is a trade-off between efficiency and equity.

³³ See Mashaw (1988).

would overstate the net cost of DI by the value of increased leisure to the disabled.

But for risk-averse individuals, pk overstates the loss in money-equivalent expected utility because of the risk premium for insurance protection and because the optimal insurance policy may include some reduction in hours of work, even under full information (see below). With imperfect information and some moral hazard, optimal benefits B_p would equate, at the margin, the utility cost of excessive reduction in h with the utility gain from increased protection (Zeckhauser 1970). Thus for workers for whom $B_{di} \leq B_p$, we can assume that any deadweight costs in the postinjury labor market are at least offset by the utility value of risk reduction and increased leisure.

In summary, for workers for whom $B_{di} \leq B_p$, an upper bound on the cost of the DI program is the excess burden induced by the effective payroll tax,³⁴ for those for whom DI levies are less than fair, plus any distortions in injury prevention and rehabilitation due to lack of firm-specific experience rating. For workers for whom $B_{di} > B_p$, there is an additional excess burden due to reduction in postinjury labor supply. Against these gross costs must be offset any gains from indexed benefits; better protection for those who would have been suboptimally insured in a pure private insurance market because of adverse selection or myopia; elimination of free riding and costs associated with the Samaritan's dilemma;³⁵ and any value assigned to the redistributive effects of DI.

IV. Optimal Disability Insurance, Labor Supply, and the Effects of Antidiscrimination Laws

It is often implicitly assumed that the reduction in labor supply associated with disability insurance results from moral hazard and entails an excess burden. Contrary to this presumption, the model developed here shows that some postdisability reduction in labor supply is consistent with optimal (full information) disability insurance. The model also shows the interdependence between optimal insurance and optimal postinjury investment in "accommodations" of the type mandated by the 1990 Americans with Disabilities Act (ADA). The likely efficiency and distributive effects of the ADA are discussed.

Assume that a worker faces a probability p of suffering a partially disabling injury that reduces market productivity. Disability may also reduce nonmarket productivity, raising the time and goods required to produce consumption commodities, and change the relative marginal utilities of work and consumption.

³⁴ Nominal tax rate minus expected benefits.

³⁵ I am indebted to Steve Coate for this point.

The wage if healthy is w^a ; if disabled it is w^d , with $w^d < w^a$. Postdisability wage w_d can be raised by investments (s) with constant unit marginal cost, such as readers for the blind or other special equipment, up to a limit given by the predisability wage: $w_s^d \geq 0$ and $w^d(s) \leq w^a$. If insurance is available, assume initially that the insurer can costlessly observe the true state of the world, including w^d and s , such that the premium is actuarially fair. With full information, insurance is payable on the occurrence of disability and need not be conditioned on labor force withdrawal. Insurance coverage is expressed as a percentage g of the monetary disability-related loss B , which includes wage loss and expenses s . The worker chooses hours of work, h^a and h^d , postinjury expenditure on s , and insurance coverage g to maximize a Von Neumann Morgenstern concave utility function, which depends directly on consumption of composite commodities Z , and indirectly on money income Y , nonmarket time L , and the parameters of the production function of Z :

$$\begin{aligned} \max_{h^a, h^d, s, g} EU &= (1 - p)U^a\{Z^a[w^a h^a - gpB, L^a]\} \\ &+ pU^d Z^d\{[w^d(z)h^d - gpB + gB - s, L^d]\}, \end{aligned} \tag{1}$$

where

$$B = w^a h^a - w^d h^d + s.$$

The first-order conditions for an interior maximum imply

$$U_z^a Z_y^a w^a - gpw^a (U_z^a Z_y^a - U_z^d Z_y^d) = U_z^a Z_1^a, \tag{2}$$

$$U_z^d Z_y^d w^d + (1 - p)gw^d (U_z^a Z_y^a - U_z^d Z_y^d) = U_z^d Z_1^d, \tag{3}$$

$$(w_s h^d - 1)[U_z^d Z_y^d + (1 - p)g(U_z^a Z_y^a - U_z^d Z_y^d)] = 0, \tag{4}$$

and

$$U_z^a Z_y^a = U_z^d Z_y^d. \tag{5}$$

We are interested in the effect of disability on hours of work and s under various possible assumptions.

A. $W^d < W^a$, No Aids ($s = 0$), No Insurance ($g = 0$)

Consider first the case where disability lowers market productivity, and postinjury aids (s) and insurance are unavailable. The first-order conditions for h^a and h^d imply that the wage is equal to the marginal rate of substitution between income and nonmarket time in each state:

$$w_i = Z_1^i / Z_y^i \quad i = a, d.$$

The effect of disability on hours of work is theoretically indeterminate. Assume initially that nonmarket productivity is unaffected and utility is state independent. Without insurance, the reduction in market wage ($w^d < w^a$) has both an income and a substitution effect: $h^d < h^a$ if the substitution of time for goods in nonmarket production outweighs the income effect. But if disability also reduces the productivity of nonmarket time, the substitution effect is weakened and if $Z_1^d / Z_1^a < w^d / w^a$, the substitution effect operates to increase h^d . A decrease in h is more likely if disability adds fixed time costs of work (e.g., transportation becomes more difficult), raises the marginal disutility of work (work becomes more painful), or if discrimination reduces market opportunities.

B. $W_s \geq 0$, $W^d(s) \leq W^a$, No Insurance Available

Now consider the case where investment in s can raise w^d , with $w_s^d \geq 0$, $w^d(s) \leq w^a$. The first-order condition (eq. [3] but with $g = 0$) implies that the worker makes the income-maximizing postinjury investment in aids, even though he bears the full marginal cost and is uninsured:

$$w_s^d h^d = 1.$$

The optimal s^* depends not only on w_s^d but also on h^d . Comparative statics shows that $\delta s^* / \delta w^d$ and $\delta h^{d*} / \delta w^d > 0$: workers for whom w^d at $s = 0$ is relatively high, will optimally invest more in s , which in turn raises w^{d*} and h^{d*} . In this respect s is like any investment in human capital. The availability of s makes cross-sectional estimates of labor supply of the disabled more elastic.

C. $W^d(s) \leq W^a$, Fair Insurance Available

With fair insurance available, equation (5) implies the standard result, that optimal coverage equalizes the marginal utility of money income in both states: $U_z^a Z_y^a = U_z^d Z_y^d$. With state-independent utility and no effect on nonmarket productivity, this implies $g = 1$, that is, optimal insurance provides full replacement of wage loss and disability-related expenses:

$$B^* = w^a h^{a*} - w^d h^{d*} + s^*,$$

that is, optimal insurance compensates for earnings reduction due to loss of productivity per hour and reduction in hours of work. If disability lowers productivity of nonmarket time or goods, optimal insurance exceeds full coverage of wage loss and work-related expenses s , that is, $g > 1$. Of course, if insurance is not perfectly experience-rated because the insurer cannot observe ex ante the individual-specific probability of disability or

control moral hazard ex post, then optimal coverage subject to available information g^{**} is less than g^* . Note that Z_y^i is the inverse of the price per unit of Z , $w^i \hat{t}^i + \hat{x}^i$, where \hat{t}^i and \hat{x}^i are the endogenous input of time and goods per unit of Z , respectively. If disability reduces market productivity more than nonmarket productivity, $Z_y^a < Z_y^d$, that is, disability reduces the price per unit of Z . Then since $U_z^a Z_y^a = U_z^d Z_y^d$, it follows that $U_z^a > U_z^d$ and $Z^d > Z^a$, that is, total consumption of Z is higher when disabled.

Comparing the first-order conditions for h^d and h^a , with and without fair insurance, shows that h^d will be lower if insurance is available. Necessary conditions for optimal allocation of time, with or without insurance, imply that $Z_1^d = w^d Z_y^d$. Since fair insurance reduces Z_y^d , Z_1^d must also fall, that is, the availability of fair insurance reduces labor supply when disabled, relative to when insurance is not available.

With fair insurance, the worker makes the income-maximizing investment in s ($w^d h^d = 1$). This is the standard result, that fair insurance with perfect observability does not interfere with loss prevention.

Two important conclusions follow from this analysis. First, $h^{d*} < h^{a*}$, that is, the availability of insurance is likely to reduce optimal hours of work when disabled, even when the insurer has full information (no moral hazard) and insurance is actuarially fair. This has important implications for interpreting the effects of DI on labor force participation. Despite the reduction in h^d , expected utility is unambiguously higher when insurance is available. Estimates of the excess burden of DI are upward biased if based on the assumption that all insurance-induced reduction in hours of work by the disabled is nonoptimal.

Second, if the utility function is state independent and disability does not reduce nonmarket productivity, then with fair insurance the worker is actually better off when disabled: he has full insurance which equalizes monetary income, works shorter hours, and has higher consumption when disabled.³⁶ This implies that workers have strong incentives to fake disability or even incur minor disability. If disability reduces nonmarket productivity or the marginal utility of consumption, then the incentive to incur disability is reduced but the incentive to fake is not. Thus the correct measure of the excess burden of DI should include only this moral hazard effect, not the optimal insurance-induced reduction in labor supply.

D. The Firm as a Potential Insurer

Inability of private insurers to observe h^d , s , and w , does not undermine the feasibility of a first best optimum for insurance and investment in s if premiums are individually fair. In that case the cost of decisions with respect to h^d or s are borne by the worker, and incentives for first best

³⁶ I am indebted to Gary Becker for emphasizing this point to me.

choices are preserved. However, if insurance is not individually fair because the insurer lacks perfect information about p ex ante (adverse selection) or cannot observe the true state of disability ex post (moral hazard), then a first best is not achievable. Controlling ex post moral hazard may be particularly problematic because, as shown above, the optimal insurance contract provides for a higher level of consumption when disabled, which creates incentives to fake or even incur minor disability provided that disability does not seriously reduce the productivity of nonmarket time or the marginal utility of consumption. If employers can better observe the true state of disability, then employer-provided disability insurance through long-term contracts may be superior to individually purchased coverage. This section therefore analyzes the optimal form of employer-provided LTD coverage.³⁷

Assume that a firm employs N homogeneous workers. Each worker faces a probability p of suffering a partially disabling injury. Disability reduces market productivity, R , such that $R^d < R^a$. Assume initially that all workers work a fixed number of hours, that is, we abstract from the labor force participation and hours of work decisions and from effects of disability on nonmarket productivity. Productivity of disabled workers can be increased by investment in two types of aid: s is a variable input per disabled worker (such as a reader for blind workers) and x is fixed input costs that are incurred if any disabled workers are employed but are invariant to the number of disabled workers (e.g., providing ramps for wheel chairs). The cost per unit of s and x is 1; $R(x, s)$ is assumed to be quasi concave, with $R_{xs} \geq 0$.

In competitive labor markets, the long-term employment contract offered implies a choice of s , x , and wage rates for healthy and disabled workers that maximizes the expected utility of the typical worker, subject to an overall break-even constraint:

$$\begin{aligned} \max_{w^a, w^d, s, x} EU &= (1 - p)U(w^a) + pU(w^d) \\ &+ \mu \{ N^a(R^a - w^a) + N^d[R^d(z, x) - w^d - s] - x \}, \end{aligned}$$

where N^a is the number of healthy workers, N^d is the number of disabled workers, and μ is a Lagrange multiplier. With a long-term contract, the relative number of healthy and disabled workers N^a/N^d will correspond to the underlying odds ratio of becoming disabled, $(1 - p)/p$. First-order conditions imply:

$$U_y^a = U_y^d = m, \quad (6)$$

³⁷ Control of adverse selection and economies of scale in administration are additional reasons why most LTD coverage is in fact obtained through employment.

$$R_s + R_{xs} = 1, \quad (7)$$

and

$$R_x + R_{sx} = 1/N^d. \quad (8)$$

Equation (6) implies that the firm acts as a mutual insurer and provides optimal risk sharing among workers. If disability does not affect the marginal utility of income, wages are equal regardless of disability status: $w^a = w^d = \bar{w}$. The common wage rate depends on average productivity, and each worker bears a prorata share of the cost of aids, x and s , each set at revenue- (and utility-) maximizing levels:

$$\bar{w} = (N^a/N)R^a + N^d/N(R^d - s) - x/N.$$

Not surprisingly, the utility-maximizing investment in the fixed component x varies inversely with the number of disabled workers. The common wage rate therefore also varies inversely with N^d . One implication is that if firms offer long-term contracts, small firms that have disabled workers would not be able to compete with larger firms, in the absence of other offsetting advantages.

E. Short-Term Contracts

If the labor contract does not provide for any guarantee of continued employment if the worker becomes disabled, then competitive spot labor markets will yield contracts that maximize the utility for each worker type, subject to separate break-even constraints:

$$w^a = R^a$$

$$w^d = R^d - s - x/N^d,$$

$$R_s + R_{xs} = 1,$$

and

$$R_x + R_{sx} = 1/N^d.$$

Investment in x and s is still at revenue-maximizing levels, but the full incidence of the cost is on disabled workers. Thus spot labor markets provide suboptimal insurance against the risk of becoming partially disabled. A further implication is again that, to the extent that optimal investment in productivity-enhancing equipment entails fixed cost, com-

petition will lead to specialization, with some firms hiring a relatively large number of disabled workers and other firms hiring none. This is an efficient distribution of employment of the disabled, in that it maximizes their net income, although it may appear to involve "discrimination."³⁸

F. Hours Variable and Insurance Available

Since long-term contracts provide better insurance for workers, consider now the optimal long-term contract when hours are variable and insurance is also available. The firm now chooses w , s , and x to maximize the worker's expected utility, subject to the additional constraint that a worker selects h to maximize a utility function over income and nonmarket time in each state and an overall break-even constraint for the firm. Define R , s , and w to represent hourly rates:

$$\max_{w^a, w^d, B, s, x} EU = (1 - p)U[Z(w^a h^a, L^a)] + pU[Z(w^d h^d + B, L^d)],$$

subject to

$$N^a(R^a - w^a)h^a + N^d[(R^d(s, x) - w^d)h^d - s - B] - x = 0,$$

and

$$Z_y^i w^i = Z_i^i, \quad i = a, d.$$

First-order conditions for w^i are

$$(UZ_y^i - \lambda)h^i + \lambda(R^i - w^i)h_w^i = 0, \quad i = a, d,$$

where λ is the Lagrange multiplier of the break-even constraint. The optimal level of B implies

$$U_z^d Z_y^d = \lambda,$$

which in turn implies that the optimal wage of the disabled is equal to their marginal revenue product, $R^d(s^*, x^*) = w^{d*}$. Substituting in the break-even constraint implies

$$w^a = R^a - [p(s + B) + x]/(1 - p)h^a.$$

³⁸ In this simple model, efficiency is equivalent to income maximization. If the disabled derive disutility from working in firms with a large number of other disabled workers, this would constrain the optimal degree of specialization below the wealth-maximizing level. However equal distribution of the disabled across all firms would be optimal only if the marginal rate of substitution between income and "equal employment opportunity" were zero for disabled workers.

Thus healthy workers are paid less than their marginal revenue product in order to cover the expected costs of insurance and accommodations if disabled. But in contrast to the case where hours of work are fixed, the optimal contract does not fully insure money income of the disabled if hours supplied by the healthy increase with w^a :

$$U_z^a Z_y^a = U_z^d Z_y^d \{1 - E_{h^a w^a} [p(s + B) + x] / (1 - p) w^a h^a \}.$$

Assuming utility is state independent, $Y^a > Y^d$, and the difference increases with p , B^* , s^* , x^* , and $E_{h^a w^a}$, the labor supply elasticity of the healthy. Thus employer-provided insurance, where the cost is borne by a wage offset when healthy, may be less than first best because of the wage offset distorts labor supply. In principle, the employer could mitigate this by specifying the number of hours required of healthy workers.

Optimal levels of accommodations, s^* and x^* , are inversely related to h^{d*} ; x^* is also inversely related to the number of disabled workers (N^d):

$$R_s^d = 1/h^d \quad \text{and} \quad R_x^d = 1/N^d h^d.$$

G. Effects of Laws Requiring Equal Wages and “Reasonable Accommodations”

It has been shown that, when labor supply elasticity is nonzero and insurance is available, the optimal long-term contract implies that wage rates of the healthy and disabled should differ in order to minimize the excess burden of distortions in labor supply, which is borne by workers. The disabled are paid their marginal product, conditional on revenue-maximizing investments in accommodations which the firm has every incentive to make. The healthy receive a wage that is less than their marginal product in order to finance the expected nonwage costs of insurance and accommodations in the event of disability.³⁹ When long-term contracts are not available, investments in accommodations are still at revenue-maximizing levels, but the incidence of the cost is on disabled workers.

If antidiscrimination statutes simply require $w^d = w^a$, clearly firms would be unwilling to hire disabled workers as long as $R^d(0, 0) < R^a$. Even if $R^d(s^*, x^*) = R^a$ and disabled workers were willing to bear the cost of accommodations, they would be precluded from doing so by the equal wage requirement, unless they could pay for the accommodations in cash.

³⁹ The results of the analysis are similar if there are three states of the world: healthy; totally disabled, with insurance B and no work; and partially disabled, in which case the worker may be rehired, and nonzero investments in accommodations may be optimal.

Thus simply mandating equal wages and that employers provide reasonable accommodations would reduce employment opportunities for the disabled.

Not surprisingly, the 1990 ADA also bans discriminatory hiring. The effect of the joint requirements of equal wages, reasonable accommodations, and nondiscriminatory hiring is that w^a must fall and w^d increases. This distorts labor supply of both groups. If the government mandates $w^d > w^{d*}$, the optimal level of B is lower: $dB^*/dw^d < 0$. If B cannot be reduced—for example, because B is set by statute (either DI or workers' compensation)—then the reduction in w^a and resulting excess burden on workers is greater than if B can be reduced.

The reduction in w^a required to meet the zero profit constraint depends also on the interpretation of "reasonable accommodations" by the courts. It has been shown that the optimal level of accommodations depends inversely on the average hours of work of the disabled and, for fixed costs, on the number of disabled workers. Thus total social costs are minimized if large firms specialize in hiring the disabled. Anecdotal evidence is consistent with this assumption. The firms cited as being exemplary in their hiring of disabled workers are all large (e.g., McDonalds). Conversely, the National Federation of Independent Business, which represents small firms, was more active than representatives of large firms in opposing the 1990 legislation. If small and large firms are held to similar definitions of "reasonable accommodations," this is equivalent to a tax on labor in small firms. The incidence of this tax depends on the mobility of workers and other factors in small firms. Further, requiring that all firms accommodate the disabled clearly entails higher total social costs than if specialization is permitted.

One of the claimed benefits for the 1990 act is that it will increase employment of the disabled and thereby reduce the burden on social insurance and welfare programs. If the law simply imposed requirements on employers that increase potential wage rates of the disabled, some increase in labor supply and employment of the disabled would be expected, assuming positive supply elasticities, but with some reduction in labor supply of the healthy. But the law also requires that all services and facilities be accessible to the disabled. This will increase the productivity of non-market time and the real income of the disabled. Unless there is an offsetting reduction in benefits paid under social insurance programs, income and substitution effects away from market time could reduce labor supply and hence increase the burden on social insurance programs.

Because many of the costs of accommodating the disabled in providing services are fixed costs, regardless of the number of disabled actually using the services, placing an accommodations requirement on all service providers entails an excessively high social cost of achieving a given level of real income for the disabled. As in the case of employment, there are large potential gains from specializing and large excess burdens from requiring

universal and uniform accommodations. Offsetting these costs may be the utility value to the disabled of equal access in all activities.⁴⁰

It seems likely that many of the disabled would prefer the cash equivalent of the social cost of providing universal access. However, the political process appears to preclude this potentially more efficient solution. The costs of mandating accommodations are hard to measure, and their incidence is very diffuse; politically, it is hard to favor “discrimination” against the disabled. Conversely, raising benefit levels would require explicit appropriations in federal or state budgets. Providing political favors by mandating costs on the private sector is becoming a familiar pattern in this era of tight government budget constraints. But the social costs, in terms of excess burden per dollar of benefit to the disabled, may well be much higher than under the explicit social insurance programs such as DI and SSI.

V. Conclusions

Previous studies of disability insurance in the United States have focused on the labor supply effects of the Social Security Disability Insurance (DI) program. But other forms of social insurance—in particular, tort liability and antidiscrimination legislation—are expanding more rapidly and may have larger economic effects. Evaluation of the net welfare effect of DI should include not only its labor supply effects but also its effects on insurance and life-cycle redistribution of income. When all these factors are considered, its net welfare cost, relative to the realistic alternative of private insurance with imperfect information, may be quite low.

A model of optimal disability insurance, postinjury accommodations, and labor supply shows that reduction in labor supply when disabled is consistent with full information insurance and does not necessarily imply moral hazard. However because utility may be higher when disabled, disability insurance creates incentives for fraudulent claims and excessive reduction in labor supply. If employers have better information than insurers about true disability status, then employer-provided disability insurance through long-term contracts may be superior to individually purchased insurance.

If long-term employment contracts are enforceable, equal wages for the healthy and the disabled are optimal only if labor supply of the healthy is totally inelastic. When labor supply elasticity is nonzero, the optimal long-term contract pays the disabled their marginal product, conditional on

⁴⁰ There may also be external effects; e.g., the value of wheelchair ramps inside buildings is greater if the access roads also have ramps. Although these external effects are not explicitly modeled, the general point applies that optimal investment in accommodations that are public goods or have external effects is lower, the lower the frequency of disabled in the population. Universal access requirements are unlikely to be optimal.

revenue-maximizing investments in accommodations. The healthy are paid less than their marginal product in order to finance insurance benefits and accommodations for the disabled. The wage differential between healthy and disabled workers is increasing in the labor supply elasticity of healthy workers, the probability of being disabled, insurance benefits for the disabled, and the optimal investment in accommodations. If accommodations entail fixed costs, regardless of the number of disabled (e.g., wheelchair ramps), the optimal investment in accommodations is positively related to the number of disabled workers. Some specialization is therefore optimal, with employment of the disabled concentrated in larger firms.

When long-term contracts are not available, investments in accommodations are still at revenue-maximizing levels, but the incidence of the cost is on the disabled. Insurance is suboptimal if insurers cannot perfectly monitor disability status, hours of work, and investment in accommodations by the disabled.

Antidiscrimination statutes that require employers to provide accommodations and pay equal wages would reduce employment opportunities for the disabled. If discriminatory hiring is also banned, as it is by the ADA, the wages of the healthy must fall, which will reduce their labor supply. Wages for the disabled increase, but effects on their labor supply are ambiguous because real income and the productivity of nonmarket time are increased by the mandate that all services and facilities be accessible to the disabled.

Because some accommodations are fixed costs, the ADA's universal access requirements imply a tax rate that varies inversely with size of firm or service establishment. The incidence of this tax will be on immobile factors in these firms and establishments. If the total social costs and the incidence of the costs of this method of expanding insurance for the disabled were explicit, it seems unlikely that it would be widely supported. However, this conclusion is tentative because the analysis here has not considered all possible effects of the ADA. Specific issues that should be addressed in future analysis include its effects on discrimination, costs of litigating valid and frivolous claims, and consumption and production externalities from providing universal access for the disabled.

References

- Berkowitz, Monroe, and Hill, M. Anne. "Disability and the Labor Market: An Overview." In *Disability and the Labor Market*, edited by Monroe Berkowitz and M. Anne Hill, pp. 1–28. Ithaca, N.Y.: Industrial and Labor Relations Press, 1986.
- Bernheim, B. Douglas. "The Economic Effects of Social Security." *Journal of Public Economics* 33 (1987): 273–304.
- Bound, John. "The Health and Earnings of Rejected Disability Insurance Applicants." Mimeograph. Cambridge, Mass.: Harvard University, 1985.

- Browning, Edgar. "The Marginal Social Security Tax on Labor." *Public Finance Quarterly* 13 (1985): 227–51.
- Collignon, Frederick C. "The Role of Reasonable Accommodation in Employing Disabled Persons in Private Industry." In *Disability and the Labor Market*, edited by Monroe Berkowitz and M. Anne Hill, pp. 196–214. Ithaca, N.Y.: Industrial and Labor Relations Press, 1986.
- Danziger, Sheldon; Haveman, Robert; and Plotnick, Robert. "How Income Transfers Affect Work, Savings and Income Distribution." *Journal of Economic Literature* 19 (September 1981): 975–1028.
- Danzon, Patricia. "Hidden Overhead Costs: Is Canada's System Really Less Expensive?" *Health Affairs* 11 (Spring 1992): 21–43.
- Feldstein, Martin. "Social Security, Induced Retirement, and Aggregate Capital Accumulation." *Journal of Political Economy* 82 (September/October 1974): 905–26.
- . "Social Security and Private Saving: Reply." *Journal of Political Economy* 90 (June 1982): 630–42.
- Halpern, Janice, and Hausman, Jerry. "Choice under Uncertainty: A Model of Applications for the Social Security Disability Insurance Program." *Journal of Public Economics* 31 (1986): 131–61.
- Hansson, Ingemar; Lyttkens, Carl H.; and Skogh, Goran. "The Excess Burden of Public Insurance: Some Results from Swedish Data." *International Review of Law and Economics* 4 (1984): 23–28.
- Haveman, Robert H., and Wolfe, Barbara L. "The Decline in Male Labor Force Participation: Comment." *Journal of Political Economy* 92, no. 3 (1984): 532–41. (a).
- . "Disability Transfers and Early Retirement: A Causal Relationship?" *Journal of Public Economics* 24 (1984): 47–66. (b)
- Health Insurance Association of America (HIAA). *Source Book of Health Insurance Data*. Washington, D.C., 1989.
- Lambrinos, James. "Health: A Source of Bias in Labor Supply Models." *Review of Economics and Statistics* (1980), pp. 206–12.
- Lee, Lung-Fei. "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables." *Econometrica* 47 (1979): 977–96.
- Leimer, Dean, and Lesnoy, Selig. "Social Security and Private Savings: New Time-Series Evidence." *Journal of Political Economy* 90 (June 1982): 609–29.
- Leonard, Jonathan S. "The Social Security Disability Program and Labor Force Participation." Working Paper no. 392. Cambridge, Mass.: National Bureau of Economic Research, 1979.
- . "Labor Supply Incentives and Disincentives for Disabled Persons." In *Disability and the Labor Market*, edited by Monroe Berkowitz and M. Anne Hill, pp. 64–94. Ithaca, N.Y.: Industrial and Labor Relations Press, 1986.
- Maki, Dennis R. "The Economic Implications of Disability Insurance in Canada." In this issue.
- Marvel, Howard P. "An Economic Analysis of the Operation of Social

- Security Disability Insurance." *Journal of Human Resources* 17 (1982): 393-412.
- Mashaw, Jerry L. "Lessons for the Administration of Workers' Compensation from the Social Security Disability Insurance program." In *New Perspectives in Workers' Compensation*, edited by John E. Burton, Jr. Ithaca, N.Y.: Industrial and Labor Relations Press, 1988.
- Parsons, Donald O. "The Decline in Male Labor Force Participation." *Journal of Political Economy* 88, no. 1 (1980): 117-34. (a)
- . "Racial Trends in Male Labor Force Participation." *American Economic Review* 70 (December 1980): 911-20. (b)
- . "Social Insurance with Imperfect State Verification: Income Insurance for the Disabled." Mimeograph. Columbus: Ohio State University, 1984. (a)
- . "Disability Insurance and Male Labor Force Participation: A Response to Haveman and Wolfe." *Journal of Political Economy* 92, no. 3 (1984): 542-49. (b)
- Sickles, Robin, and Taubman, Paul. "An Analysis of the Health and Retirement Status of the Elderly." *Econometrica* 54, no. 6 (1986): 1339-56.
- Slade, Frederick P. "Older Men, Disability Insurance and Incentive to Work." *Industrial Relations* 23, no. 2 (1984): 260-77.
- Treitel, Ralph. "Appeal by Denied Disability Claimants." Staff Paper no. 23. Washington, D.C.: U.S. Department of Health, Education and Welfare, Social Security Administration, Office of Research and Statistics, 1976.
- U.S. Bureau of the Census. *Statistical Abstract*, Current Population Survey. Washington, D.C.: U.S. Government Printing Office, March 1989.
- U.S. Congress. House. American with Disabilities Act of 1990. Conference Report (to accompany S. 9333). 101st Cong. 2d sess., July 12, 1990.
- Zeckhauser, Richard. "Medical Insurance: A Case Study of the Trade-off between Risk Spreading and Appropriate Incentives." *Journal of Economic Theory* 2 (1970): 10-26.