



ELSEVIER

Journal of Public Economics 57 (1995) 1–23



Economic aspects of optimal disability benefits

Peter Diamond^{a,*,\dagger}, Eytan Sheshinski^b

^a*Department of Economics, Massachusetts Institute of Technology, E52-344,
50 Memorial Drive, Cambridge, MA 02139-4307, USA*

^b*Eliezer Kaplan School of Economics and Social Sciences, The Hebrew University,
Jerusalem, Israel*

Received October 1992, revised version received July 1993

Abstract

This paper analyzes optimal disability and retirement (or welfare) benefits with imperfect disability evaluation (with some able workers judged disabled and some disabled workers judged able). Thus the levels of both disability and retirement benefits affect labor supply. With anyone not working eligible for retirement benefits, we analyze the optimal structure of benefits for a given disability screening mechanism and briefly consider the problem of optimal evaluations of disability evidence. In the United States, there is an overlap in eligibility for disability and retirement portions of Social Security. More generally, welfare is sometimes available to people denied disability benefits.

Key words: Social security; Disability; Retirement

JEL classification: H53; H21

1. Introduction

The U.S. Social Security system began with retirement benefits but no disability benefits. When disability insurance was added to the system (in 1956), the formula for retirement benefits was adapted to determine disability benefits. With the introduction of early retirement benefits at age

* Research supported by the National Science Foundation and the Binational Science Foundation. Research assistance done by Douglas Galbi.

^{\dagger} Corresponding author.

62 (in 1956 for women, 1961 for men), there has been a three-year overlap when workers could apply for either retirement or disability benefits. Economists have been studying the relationship between labor supply of older workers and the availability of retirement benefits.¹ Economists have also studied the optimal design of retirement benefits in recognition of the effect of benefits on labor supply.² While it has been recognized that disability benefits also affect labor supply³, we are unaware of any studies of the optimal design of disability and retirement benefits in recognition of the effect of benefits on labor supply.⁴ This paper begins such a study, using a simple static model to analyze optimal levels of disability and welfare (or retirement) benefits with recognition of the imperfect nature of disability evaluation.

We begin with a discussion of the problems of modeling disability and disability screening mechanisms. We next analyze the optimal structure of benefits for given disability levels. We then specialize the model to the case of only two disability levels to simplify the analysis of comparative statics and to examine optimal disability standards and benefits levels. We return to the continuous model for additional analysis of optimal standards and benefits. We conclude with a logarithmic example and suggestions for further research.

2. Modelling disability

At any time there is a vast array of actual and potential marginal products among employed and not-employed individuals in the economy. Similarly, there is a vast spread in the difficulties or disutilities associated with working. For some, the alternative source of income should they not work is disability income rather than retirement income or welfare benefits. In principle, social security disability benefits are available for those unable to find any remunerative employment as a consequence of their disabilities. Yet, some people with considerable handicaps are employed, while others with seemingly similar handicaps are receiving benefits. While eligibility for benefits is straightforward for some handicaps, with others it is difficult to

¹ Diamond and Hausman (1984), Burtless and Moffitt (1984), and Gustman and Steinmeier (1985).

² Diamond and Mirrlees (1978, 1982, 1986).

³ Halpern and Hausman (1986), Leonard (1986).

⁴ The paper by N. Stern (1982) can be interpreted as contrasting welfare benefit and disability programs. For another analysis of disability insurance that recognizes imperfect measurement of disability, see Parsons (1990). See also Akerlof (1978) for optimal benefits with one type of classification error, but not both types. For a general discussion of disability programs in eight countries, see Haveman et al. (1984).

determine. For some handicaps it is hard for any outsider to judge the ability to work or its difficulty.

Our problem is to model this complex picture in a way that is helpful for the analysis of economic variables such as benefit levels. Since we will analyze a static model, we will not explicitly model individual histories (including, for example, industrial accidents) which may shed light on abilities to work. We have identified two separate aspects of working—the ability to carry on economically useful activities and the difficulty (including physical pain) in carrying them out. For convenience, we will work with a single variable by assuming that all individuals have the same marginal products, thus ignoring correlations between productivity and disability incidence. In contrast, we will assume that each individual has an additive disutility of work, with disutilities distributed over all non-negative values. For modeling purposes here, it makes little difference whether someone is totally unable to work or has an infinite disutility of work. We further simplify the model by assuming that all workers have the same (concave, increasing) utility of consumption, $u(c)$, while all non-workers have the same (concave, increasing) utility of consumption $v(c)$.⁵ Thus the distribution of disutilities plays a central role in determining willingness to work. We take work to be a 0–1 variable and do not model varying hours or work intensity.

Recognizing the problem of determining disability, we assume that disutility is not perfectly observable. For a start, we assume that there is a disability evaluation mechanism in place.⁶ When someone with disutility θ is evaluated for disability, there is a probability, $p(\theta)$, that the individual will be judged disabled.⁷ Naturally, we assume that $p(\theta)$ is increasing in θ and, for notational convenience, that $p(0) = 0$. In Section 6 we look into the public decisions that affect the shape of $p(\theta)$. These decisions relate the sizes of type I and type II errors to the strictness of the disability standard.

3. Continuous model

We consider three different consumption levels: c_a for active workers; c_b for welfare (or early retirement) beneficiaries; and c_d for disabled

⁵ We also assume that both u' and v' go from ∞ to 0 as c goes from 0 to $+\infty$.

⁶ We assume that the evaluation process is costless for both the applicant and the evaluator. It would be interesting to introduce costs.

⁷ We assume that $p(\theta)$ is known to the government decision-makers. Obviously, it is not precisely known, but one can make reasonable hypotheses about its basic form. This paper can be interpreted as a mapping from beliefs about the accuracy of disability evaluation into desired benefit levels. For some evidence on disability evaluation, see Nagi (1969). For evidence that some people denied disability benefits work while others do not, see Bound (1989).

beneficiaries. The utility function for a worker with labor disutility level θ is written $u(c_a) - \theta$. We assume that θ is non-negative and distributed in the population with distribution $F(\theta)$ [and density $f(\theta)$]. For convenience, we assume that $f(\theta)$ is continuous and positive for all non-negative values of θ . The utility function for non-workers is $v(c)$, where c will take on the values c_b or c_d depending on whether the non-worker is receiving welfare or disability benefits. We assume that disability benefits are at least as large as welfare benefits (otherwise, there is no reason to have a separate disability program with a selective screening mechanism):

$$c_d \geq c_b. \quad (1)$$

Given the utility functions and the consumption levels for workers and non-workers, we can analyze which individuals would prefer working to receiving retirement benefits and which would prefer working to receiving disability benefits (if eligible). Those with disutilities of labor below the threshold values which equate utilities will choose to work. Thus the threshold values satisfy

$$\begin{aligned} \theta_b &= \max[0, u(c_a) - v(c_b)], \\ \theta_d &= \max[0, u(c_a) - v(c_d)]. \end{aligned} \quad (2)$$

Since disability benefits are greater than retirement benefits (1), the threshold labor disutility for persons who apply for disability benefits is lower than the threshold for those who apply for retirement benefits. Thus we have

$$\theta_b \geq \theta_d. \quad (3)$$

With the fraction of the population with disutility θ who are eligible for disability benefits written as $p(\theta)$, individuals' labor supply choices and consumption levels are as shown in Fig. 1.

We normalize by setting the marginal product of a worker equal to one. We can now state the problem of the selection of consumption levels to maximize social welfare as

$$\begin{aligned} \max \int_0^{\theta_d} [u(c_a) - \theta] dF(\theta) &+ \int_{\theta_d}^{\theta_b} [p(\theta)v(c_d) + (1 - p(\theta))(u(c_a) - \theta)] dF(\theta) \\ &+ \int_{\theta_b}^{\infty} [p(\theta)v(c_d) + (1 - p(\theta))v(c_b)] dF(\theta), \end{aligned} \quad (4a)$$

subject to

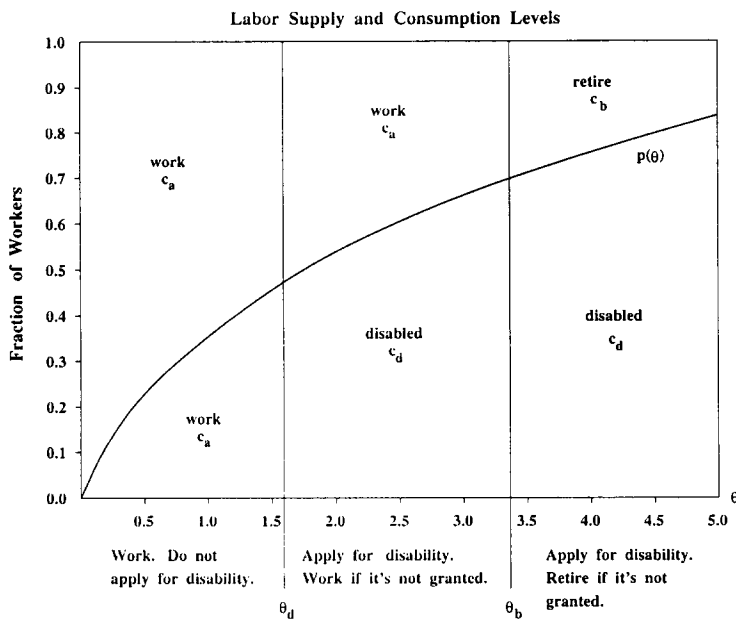


Fig. 1.

$$\begin{aligned}
 & \int_0^{\theta_d} (c_a - 1) dF + \int_{\theta_d}^{\theta_b} [p(\theta)c_d + (1 - p(\theta))(c_a - 1)] dF \\
 & + \int_{\theta_b}^{\infty} [p(\theta)c_d + (1 - p(\theta))c_b] dF = R, \tag{4b}
 \end{aligned}$$

where R are the resources available to the economy. We first discuss the first-order conditions under the assumptions that $\theta_d > 0$ and $\theta_b > \theta_d$. We then state sufficient conditions for these strict inequalities to hold at the optimum.

Using the Lagrange multiplier λ and assuming also that $\theta_d > 0$, we can state the first-order conditions as the resource constraint (4b) and the following three equations:

$$\begin{aligned}
 & [u'(c_a) - \lambda] \int_0^{\theta_d} dF + [u'(c_a) - \lambda] \int_{\theta_d}^{\theta_b} (1 - p(\theta)) dF \\
 & = \lambda(c_a - 1 - c_d)p(\theta_d)f(\theta_d) \frac{\partial \theta_d}{\partial c_a}
 \end{aligned}$$

$$\begin{aligned}
& + \lambda(c_a - 1 - c_b)(1 - p(\theta_b))f(\theta_b) \frac{\partial \theta_b}{\partial c_a} \\
= & \lambda[(c_a - 1 - c_d)p(\theta_d)f(\theta_d) + (c_a - 1 - c_b)(1 - p(\theta_b))f(\theta_b)]u'(c_a), \tag{5a}
\end{aligned}$$

$$[v'(c_d) - \lambda] \int_{\theta_d}^{\infty} p(\theta) dF = -\lambda(c_a - 1 - c_d)p(\theta_d)f(\theta_d)v'(c_d), \tag{5b}$$

$$[v'(c_b) - \lambda] \int_{\theta_b}^{\infty} (1 - p(\theta)) dF = -\lambda(c_a - 1 - c_b)(1 - p(\theta_b))f(\theta_b)v'(c_b), \tag{5c}$$

where we have used (2) to obtain the derivatives of θ_b and θ_d . If, at the optimum, θ_d is zero, we have the same conditions since $p(0) = 0$. The left-hand sides of all three expressions are the social values of the differences between giving consumption to the agents in the three different positions rather than holding the resources. The right-hand sides are the social values of the resource savings from the induced changes in labor supply as a consequence of altered benefits. The private return to working for someone ineligible for disability benefits is $c_a - c_b$. Comparing this with the marginal product, we see that there is an implicit tax on work when $c_a - c_b < 1$. From (5c), we will have an implicit tax on work if, at the optimum, $v'(c_b) > \lambda$. Below we will note plausible sufficient conditions for this conclusion.

Dividing the expressions in (5) by the marginal utilities on the right-hand side and adding, we see that the inverse of the Lagrangian equals the average of the inverses of the marginal utilities of consumption:

$$\begin{aligned}
& \int_0^{\theta_d} (u'(c_a))^{-1} dF + \int_{\theta_d}^{\theta_b} [(1 - p(\theta))(u'(c_a))^{-1} + p(\theta)(v'(c_d))^{-1}] dF \\
& + \int_{\theta_b}^{\infty} [(1 - p(\theta))(v'(c_b))^{-1} + p(\theta)(v'(c_d))^{-1}] dF = \lambda^{-1}. \tag{6}
\end{aligned}$$

We turn next to sufficient conditions for the solution to be internal. The question of when it is desirable to have some work in the economy ($\theta_b > 0$) is a question of the wealth of the economy, i.e. the size of R . If no one works, it is optimal for everyone to have the same consumption ($c_b = c_d = R$). This

allocation will not be optimal if those with the least disutility of labor were to choose to work for the additional consumption equal to their marginal products:

$$u(R + 1) > v(R). \tag{7}$$

We assume that (7) is satisfied (implying $\theta_b > 0$) and refer to it as the poverty condition.

We turn next to sufficient conditions for disability benefits to exceed welfare benefits. Consider the economy with a single benefit program. The level of benefits is chosen relative to the consumption of workers to balance the effect on labor supply with the desire to allocate consumption efficiently over workers and non-workers. Introducing a small disability program permits higher benefits for a fraction of non-workers, $\int_{\theta_b}^{\infty} p(\theta) dF$, at the cost of a decrease in labor supply of size $p(\theta_b)f(\theta_b)$. Without a separate disability program, the attempt to pay more to all non-workers raises benefits for the $1 - F(\theta_b)$ non-workers and decreases the labor supply by $f(\theta_b)$. Thus, having $p(\theta)$ increasing in θ is sufficient to make a disability program desirable, provided the marginal utility of consumption of non-workers exceeds that of workers at the optimum without a disability program.⁸

A sufficient condition for this latter relation can be expressed in terms of the Diamond and Mirrlees (1978) moral hazard condition that equating utilities between non-workers and workers who like work the most leaves marginal utility higher for non-workers:

$$u(x) = v(y) \text{ implies } u'(x) < v'(y). \tag{8}$$

By (7) there is some work at the optimal allocation. Thus, at least the most able worker ($\theta = 0$) must work, implying $u(c_a) > v(c_b)$. And so, by (8), $u'(c_a) < v'(c_b)$. While it is not so plausible that (8) holds for workers with $\theta = 0$, it is plausible that the same condition does hold for a sufficient range of values of θ to yield the condition needed, that, at the optimum, $u'(c_a) < v'(c_b)$. We assume that (8) is satisfied and conclude that $c_d > c_b$.

From the equation for λ , Eq. (6), we then have $\lambda < v'(c_b)$. From the first-order condition, (5c), we then have an implicit tax on work, $1 > c_a - c_b$. It follows that $1 > c_a - c_d$ and so, from (5b), $\lambda \leq v'(c_d)$, with equality if $\theta_d = 0$. We can then conclude that $\lambda > u'(c_d)$. Thus both welfare and disability programs have benefit levels below the level that would equate the marginal utilities of beneficiaries with that of workers.

⁸ To see this formally, add (5b) and (5c) under the condition $c_b = c_d$ and so $\theta_b = \theta_d$. The fact that $p(\theta)$ is increasing will then imply that the left-hand side of (5b) exceeds the right-hand side, provided that $v'(c_d) > \lambda$.

4. Labor disutility observable

We have used the realistic assumption that labor disutility (or disability) is not perfectly observable. It may assist in the interpretation of the first-order conditions to analyze the full optimum where disability is observable. The maximization problem in (4) is changed by making θ_b and θ_d control variables rather than endogenously determined by benefit levels in (2). In this case, there is no reason for a distinction between welfare and disability benefits ($c_b^* = c_d^*$). Consumption is optimally allocated to equate marginal utilities of consumption:

$$u'(c_a^*) = v'(c_d^*) . \quad (9)$$

All individuals with disutility levels below some cutoff θ^* should work. The cutoff is determined by comparing the utility gain from extra work, $u(c_a^*) - \theta^* - v(c_d^*)$, with the value of extra net consumption as a consequence of work, which is the sum of marginal product and the change in consumption which results from the change in status:

$$u(c_a^*) - \theta^* - v(c_d^*) = -u'(c_a^*)(1 - c_a^* + c_d^*) . \quad (10)$$

The allocation determined by (9), (10), and the resource constraint (4b) differs from that analyzed in Section 3 in that θ^* is determined by social needs while θ_b and θ_a are determined by workers to equate utilities with and without work. From the moral hazard condition we know that an attempt to implement an allocation satisfying (9) results in no one choosing to work when disutility is not observable. Thus, without observability, it is optimal to adjust consumption levels away from the equal marginal utility condition to induce a more appropriate labor supply.

5. Two-class model

To get some additional feeling for the nature of the first-order conditions, we consider a case where there are only two types in the economy, θ_1 and θ_2 ($\theta_1 < \theta_2$), with population weights f_1 and $f_2 = 1 - f_1$. The probability of a type- i person being judged disabled is p_i , with $p_2 > p_1$. We assume that the optimum has the form that the θ_1 types accept disability benefits if judged disabled but otherwise work; while the θ_2 types do not work whatever their eligibility for disability benefits. For a logarithmic example, sufficient conditions for the optimum to have this form will come out of analysis of that special case.

The social welfare optimization now takes the form

$$\max[(u(c_a) - \theta_1)(1 - p_1) + v(c_d)p_1]f_1 + [v(c_b)(1 - p_2) + v(c_d)p_2]f_2, \quad (11a)$$

subject to

$$(c_a - 1)(1 - p_1)f_1 + c_d p_1 f_1 + c_b(1 - p_2)f_2 + c_d p_2 f_2 = R, \quad (11b)$$

$$v(c_d) \geq u(c_a) - \theta_1 \geq v(c_b) \geq u(c_a) - \theta_2, \quad (11c)$$

$$c_d \geq c_b. \quad (11d)$$

The constraints in (11c) ensure that individual behavior coincides with that described in the objective function and the resource constraint. One can check that the condition $u(c_a) - \theta_1 \geq v(c_b)$ and the moral hazard condition (8) imply that at the optimum

$$u(c_a) - \theta_1 = v(c_b). \quad (12)$$

Thus the optimum is described by the resource constraint (11b), the equal utility condition (12) and a single first-order condition which can be obtained by differentiating social welfare in (11a) with respect to c_d , with the c_a and c_b functions of c_d given by (11b) and (12). Upon differentiation we have

$$N \equiv \{(1 - p_1)f_1 v'(c_b)[v'(c_d) - u'(c_a)] + (1 - p_2)f_2 u'(c_a)[v'(c_d) - v'(c_b)]\},$$

$$\frac{dW}{dc_d} = N[p_1 f_1 + p_2 f_2][v'(c_b)(1 - p_1)f_1 + u'(c_a)(1 - p_2)f_2]^{-1}. \quad (13)$$

The moral hazard condition implies that dW/dc_d is positive at $c_d = c_b$. Thus the optimum has a separate disability program. Since c_d does not affect behavior at the margin, $v'(c_d)$ is equal to the Lagrangian on the resource constraint.

Of the various comparative static exercises one could do, we restrict ourselves to one simple one. Consider a set of different economies which happen to have chosen the same levels of c_a and c_b . Assume that the different economies have different levels of resources R and different levels of one of the variables p_1 , p_2 , or f_1 . Then, comparative statics is done by examining the equation, $N = 0$. Differentiating N , and evaluating where $N = 0$, we have

$$\frac{\partial N}{\partial c_d} < 0, \quad \frac{\partial N}{\partial p_1} < 0, \quad \frac{\partial N}{\partial p_2} > 0, \quad \frac{\partial N}{\partial f_1} > 0. \quad (14)$$

Thus, across these different economies, those with a more discriminating disability system (p_1 smaller or p_2 larger) or a smaller population with high disability have larger benefit levels. While the former proposition is intui-

tive, the latter rests on the fact that for a program increase, there must be decreases in both wages and welfare benefits to finance the change, and this is more attractive the smaller is the welfare population.

6. Disability standard

Relative to the set of people one would ideally like to see receiving disability benefits, any attempt to evaluate abilities to work will be subject to two types of error—admission of people ideally omitted and exclusion of people ideally included.⁹ In this section, we generalize the model of Section 5 to relate the probabilities of a disability finding, p_1 and p_2 , to a disability standard, θ^* . We also consider the first-order condition for optimization of θ^* . In the next section we will extend the analysis to the continuous case.

Individuals are now characterized by two variables, θ and θ_e , being true and estimated disutilities of labor, respectively. We assume that the distributions of θ_e for the two types are $G_1(\theta_e)$ and $G_2(\theta_e)$.¹⁰ Since observation of θ_e is merely a device for making inferences about θ , it is possible to define measured disability so that two conditions are met (see Milgrom, 1979). First (which is implied by the second), the assumption that $p_1 < p_2$ for any disability standard θ^* :

$$p_1(\theta^*) = 1 - G_1(\theta^*) < 1 - G_2(\theta^*) = p_2(\theta^*) . \quad (15)$$

Second, the appropriate definition of measured disability gives us the monotone likelihood ratio condition for all θ^* :

$$\frac{g'_1(\theta^*)}{g_1(\theta^*)} < \frac{g'_2(\theta^*)}{g_2(\theta^*)} . \quad (16)$$

In keeping with the analysis of the previous section, we would like to exclude all type one's from disability benefits while including all type two's. However, some type one's are included for any finite standard and some type two's are excluded for any positive standard. Thus it is inevitable that some errors of classification are made. The problem is to strike the appropriate balance between the two types of error.

Continuing to assume that the optimum occurs with type one's accepting disability but not welfare benefits while type two's do not work, we have the same maximization problem as in (11) with p_i now functions of the control variable θ^* . Recalling that the Lagrangian is equal to $v'(c_d)$, the first-order condition for θ^* is

⁹ We draw on the analysis of auto insurance underwriting by Smallwood (1975).

¹⁰ We assume that G_1 and G_2 have positive support on the entire positive halfline.

$$[v(c_d) - v(c_b)](f_1 g_1 + f_2 g_2(\theta^*)) = v'(c_d)[(c_d + 1 - c_a)f_1 g_1(\theta^*) + (c_d - c_b)f_2 g_2(\theta^*)]. \tag{17}$$

Rearranging terms, it is instructive to write the first-order condition as

$$[v(c_d) - v(c_b) - v'(c_d)(c_d - c_b)](f_2 g_2(\theta^*)) = [v'(c_d)(c_d + 1 - c_a) - (v(c_a) - v(c_b))]f_1 g_1(\theta^*). \tag{18}$$

By the concavity of v , the left-hand side is positive at the optimum. Thus the right-hand side is positive, implying that $1 - c_a + c_d$ is positive; this also follows from the presence of an implicit tax on work as part of the optimal structure which was argued in Section 3.

Using the monotone likelihood ratio condition, (16), one can draw comparative statics from (18). Let us consider a θ^* and a set of economies that have the same levels of benefits and wages, the same g_i functions, but different levels of R and f . Considering (13) and (18), this can only happen where $g_2 G_1 / g_1 G_2$ is independent of θ for some range. Thus, we can differentiate (18) alone, knowing that (13) continues to be satisfied. By implicit differentiation of (18), we see that θ^* is larger the greater the size of the type one disability eligible population.

7. Disability standard: Continuous case

In a similar fashion, we can introduce a choice of disability standard into the continuous case. We write the distribution of observed disutility, θ_e , conditional on true disability being θ as $G(\theta_e; \theta)$, with density $g(\theta_e; \theta)$.¹¹ Then, for disability standard θ^* , we have

$$p(\theta) = 1 - G(\theta^*; \theta). \tag{19}$$

Returning to the welfare maximization problem, Eqs. (4), we can now add a further first-order condition by differentiating with respect to θ^* (assuming $\theta^* > 0$):

$$\int_{\theta_d}^{\theta_b} [v(c_d) - u(c_a) + \theta]g(\theta^*; \theta) dF(\theta) + \int_{\theta_b}^{\infty} [v(c_d) - v(c_b)]g(\theta^*; \theta) dF(\theta) = \lambda \left[\int_{\theta_d}^{\theta_b} [c_d - c_a + 1]g(\theta^*; \theta) dF(\theta) + \int_{\theta_b}^{\infty} [c_d - c_b]g(\theta^*; \theta) dF(\theta) \right]. \tag{20}$$

¹¹ For convenience, we assume that g is positive for all positive values of θ_e .

Thus a loosening of the standard attracts some people to the disability program from the labor force while shifting others from retirement. The former shift has a net social value $v(c_d) - \lambda c_d - (u(c_a) - \theta - \lambda(c_a - 1))$, while the latter has a social value $v(c_d) - \lambda c_d - (v(c_b) - \lambda c_b)$. We write these differences as $\Delta_a + \theta$ and Δ_b , respectively. Across the set of people shifted, these net gains must balance out for an optimum. That is, the first-order condition (20) can be written as

$$\Delta_b \int_{\theta_b}^{\infty} g(\theta^*; \theta) dF(\theta) + \Delta_a \int_{\theta_d}^{\theta_b} g(\theta^*; \theta) dF(\theta) + \int_{\theta_d}^{\theta_b} \theta g(\theta^*; \theta) dF(\theta) = 0. \tag{21}$$

From the concavity of v and the fact that $v'(c_d) \geq \lambda$, we have

$$\begin{aligned} v(c_d) - v(c_b) &> v'(c_d)(c_d - c_b) \\ &\geq \lambda(c_d - c_b). \end{aligned} \tag{22}$$

Thus, at the optimum, with a decrease in disability standards, social welfare goes up from additional non-workers added to the disability rolls ($\Delta_b > 0$), while it goes down from at least some of the workers switching to disability ($\Delta_a < 0$).

8. Multiple disability standards

If individuals with different health problems (e.g. physical vs. mental, or sight, lower back, etc.) have different distributions of actual and observed disutilities, we can think of the population as made up of the sum of distributions $F_i(\theta)$ with conditional distributions of observed disutilities $G_i(\theta_e; \theta)$. Similarly, grouping workers by age results in groups with different distributions of actual and observed disutilities. With the same benefits for all recipients (as with the U.S. disability program) and the same Lagrangian for the single government budget constraint, the first-order condition (20) tell us how the standard θ_i^* should vary with distribution $G_i(\theta_e; \theta)$, while the other first-order conditions are modified by adding over different health problems. In (21), θ_i^* , F_i , and g_i are the only elements to vary with respect to i .

To proceed, we now assume the analog of (18), which Milgrom (1979) calls the local monotone ratio property:

$$\frac{\partial}{\partial \theta} \log \frac{\partial}{\partial \theta_e} g_i(\theta_e; \theta) > 0. \tag{23}$$

Unlike the two-class case, this is an assumption on the joint distribution of θ_e and θ rather than following from a suitable definition of θ_e . Given (23), the left-hand side of (21) is increasing in θ^* . It is natural to ask how the

standard should vary with the distribution. That is, across different health problems, how does θ_i^* vary with F_i and g_i in Eq. (21). We shall examine the special case where the distributions of disutilities are the same, $F_1(\theta) = F_2(\theta)$, but the conditional distributions G_i are different. A simple rightward shift for the difference between distributions is obviously offset by a comparable shift in θ^* . The interesting case to analyze would be for a difference in the shape of the distributions.

9. Logarithmic example

This section explores in more detail the two-class model of Section 5 with the specification of logarithmic utility for workers and non-workers. There are six possible configurations of the economy. Workers of each type might either work (case I), apply for disability and work if it is not granted (case II), or apply for disability and retire if it is not granted (case III). Since disutility of work is greater for type two workers than for type one workers ($\theta_1 < \theta_2$), there are six possible configurations as follows.

Configuration 1 – (I,I). Both types work and do not apply for disability. The resource constraint implies $c_a = R + 1$, social welfare:

$$W_1 = \log(R + 1) - \theta_1 f_1 - \theta_2 f_2.$$

Configuration 2 – (I,II). Type one works, type two applies for disability and works if it is not granted. The incentive compatibility constraints are

$$\log(c_b) \leq \log(c_a) - \theta_2 \leq \log(c_d) \leq \log(c_a) - \theta_1. \quad (24)$$

From (24), the marginal utility for persons on disability is higher than for those working. The social planner thus wants to make c_d as high as possible, consistent with (24). Hence

$$\log(c_d) = \log(c_a) - \theta_1 \quad \text{or} \quad c_d = c_a e^{-\theta_1}. \quad (25)$$

The resource constraint is

$$(c_a - 1)f_1 + c_d p_2 f_2 + (c_a - 1)(1 - p_2)f_2 = R. \quad (26)$$

Solving (25) and (26) gives

$$c_a = \frac{R + 1 - p_2 f_2}{e^{-\theta_1} p_2 f_2 + 1 - p_2 f_2}. \quad (27)$$

Social welfare is

$$W_2 = (\log(c_a) - \theta_1)f_1 + \log(c_d)p_2 f_2 + (\log(c_a) - \theta_2)(1 - p_2)f_2.$$

Using (25) and (27) gives

$$W_2 = \log\left(\frac{R + 1 - p_2 f_2}{e^{-\theta_1} p_2 f_2 + 1 - p_2 f_2}\right) - \theta_1(f_1 + p_2 f_2) - \theta_2(1 - p_2)f_2.$$

Configuration 3 – (I,III). Type one works, type two applies for disability and retires if it is not granted. The incentive compatibility constraints are

$$\log(c_a) - \theta_2 \leq \log(c_b) \leq \log(c_d) \leq \log(c_a) - \theta_1. \quad (28)$$

From (28), the marginal utility of persons on retirement or disability is greater than that of workers. Thus the social planner sets c_d and c_b as high as possible consistent with (28):

$$\log(c_b) = \log(c_d) = \log(c_a) - \theta_1 \quad \text{or} \quad c_b = c_d = c_a e^{-\theta_1}. \quad (29)$$

The resource constraint is

$$(c_a - 1)f_1 + c_d p_2 f_2 + c_b(1 - p_2)f_2 = R. \quad (30)$$

Solving (29) and (30) gives

$$c_a = \frac{R + f_1}{f_2 e^{-\theta_1} + f_1}. \quad (31)$$

Social welfare is

$$W_3 = (\log(c_a) - \theta_1)f_1 + \log(c_d)p_2 f_2 + \log(c_b)(1 - p_2)f_2.$$

Using (29) and (31) gives

$$W_3 = \log\left(\frac{R + f_1}{f_2 e^{-\theta_1} + f_1}\right) - \theta_1.$$

Configuration 4 – (II,II). Both types apply for disability and work if it is not granted. The incentive compatibility constraints are

$$\log(c_b) \leq \log(c_a) - \theta_2 \leq \log(c_a) - \theta_1 \leq \log(c_d).$$

The social planner equalizes marginal utility for workers and disabled persons by setting

$$c_a = c_d. \quad (32)$$

The resource constraint is

$$(c_a - 1)((1 - p_1)f_1 + (1 - p_2)f_2) + c_d(p_1 f_1 + p_2 f_2) = R. \quad (33)$$

Solving (32) and (33) gives

$$c_a = R + 1 - p_1 f_1 - p_2 f_2. \quad (34)$$

Social welfare is

$$W_4 = (\log(c_a) - \theta_1)(1 - p_1)f_1 + \log(c_d)p_1 f_1 + (\log(c_a) - \theta_2)(1 - p_2)f_2 + \log(c_d)p_2 f_2. \quad (35)$$

Combining (32), (34), and (35) gives

$$W_4 = \log(R + 1 - p_1 f_1 - p_2 f_2) - \theta_1(1 - p_1)f_1 - \theta_2(1 - p_2)f_2.$$

Configuration 5 – (II,III). Both types apply for disability. If disability is not granted, type one’s work while type two’s retire. This is the configuration analyzed in Section 5. With $u(x)$ and $v(x)$ logarithmic, solving (11b), (12), and (13) gives

$$c_a = (R + (1 - p_1)f_1) \frac{(1 - p_1)f_1 + (1 - p_2)f_2}{(1 - p_1)f_1 + e^{-\theta_1}(1 - p_2)f_2}, \tag{36}$$

$$c_b = c_a e^{-\theta_1}, \tag{37}$$

$$c_d = R + (1 - p_1)f_1, \tag{38}$$

$$W_5 = \log(R + (1 - p_1)f_1) + \left[\log \frac{(1 - p_1)f_1 + (1 - p_2)f_2}{(1 - p_1)f_1 + e^{-\theta_1}(1 - p_2)f_2} - \theta_1 \right] \cdot [(1 - p_1)f_1 + (1 - p_2)f_2].$$

Configuration 6 – (III,III). Both types apply for disability and retire if it is not granted. The incentive compatibility constraints are

$$\log(c_a) - \theta_2 \leq \log(c_a) - \theta_1 \leq \log(c_b) \leq \log(c_d).$$

Setting $c_b = c_d$ equalizes marginal utility for persons on retirement and persons on disability. From the resource constraint $c_d = c_b = R$; social welfare is $W_6 = \log(R)$.

Figs. 2–6 present social welfare and benefit levels under the optimal program for parameter variations around a base setting of

$$\theta_1 = 1, \quad \theta_2 = 4, \quad f_1 = f_2 = 0.5, \quad p_1 = 0.25, \quad p_2 = 0.75 \quad \text{and} \\ R = 0.45.^{12}$$

Fig. 2 shows how different configurations are optimal for different levels of R . Note that Configuration 4, in which both types apply for disability and work if it is not granted, is not optimal for any value of R , given the above base settings for the other parameters. The difference between p_1 and p_2 , i.e. the effectiveness of disability screening, needs to be increased significantly to generate Configuration 4.¹³ That is not surprising, since in

¹² If workers were free to choose their hours h given wage $1/8$ per hour and disutility of work $\theta_i/8$ per hour, they would choose h so as to maximize $\log(h/8) - h_i(\theta_i/8)$. The θ_i 's were chosen so that type one would want to work 8 hours while type two would want to work 2 hours. The base setting for R was chosen so as to be in the middle of the range of R for which Configuration 5 is optimal (see Fig. 2).

¹³ For example, there is configuration 4 with $p_1 < 0.08$ for $p_2 = 0.75$ at the base setting of the other parameters.

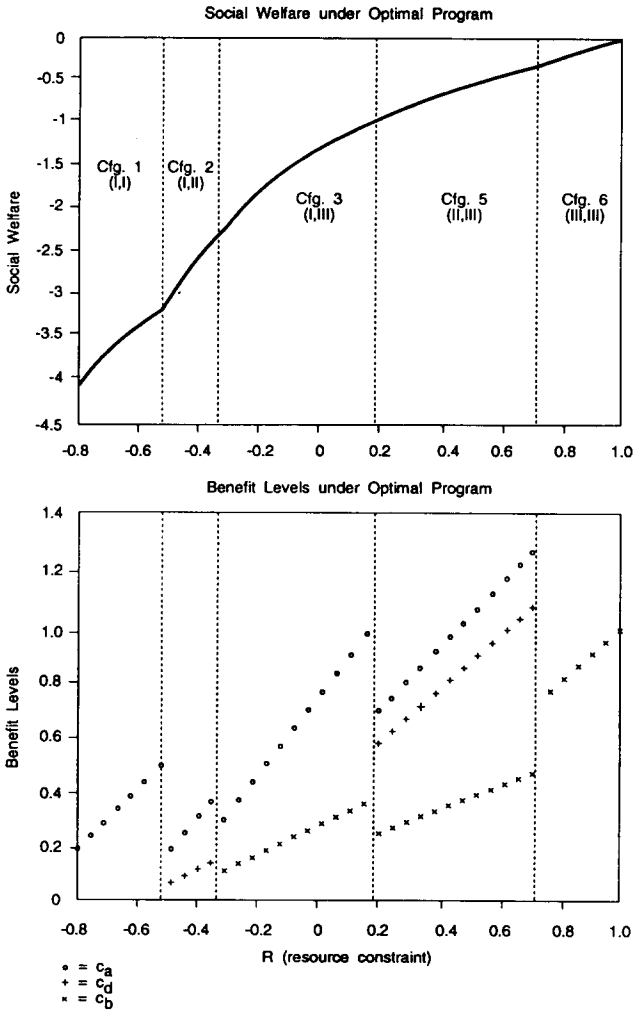


Fig. 2.

Configuration 4 both types make the same labor supply choice at the optimal consumption levels. Disability screening has to compensate for differences in disutility of work in order to have this pattern of benefits optimal.¹⁴ Social welfare is continuous in resources, although consumption levels are not, with discontinuous jumps when the configuration changes. As

¹⁴ Configurations 1 and 6, which also involve symmetric labor supply choices, can always be generated by sufficiently extreme values of R .

resources increase, avoiding the disutility of work becomes more attractive. Each configuration change is an increase in the number of people not working financed in part from a fall in wages. Presumably the discontinuity would not be present with a continuous distribution of types.

Fig. 3 shows the effects of varying f_1 given the base settings for the other parameters. If f_1 is less than 0.05, Configuration 3 is optimal; otherwise, Configuration 5 is optimal, as is the case for the range shown. From (38) one

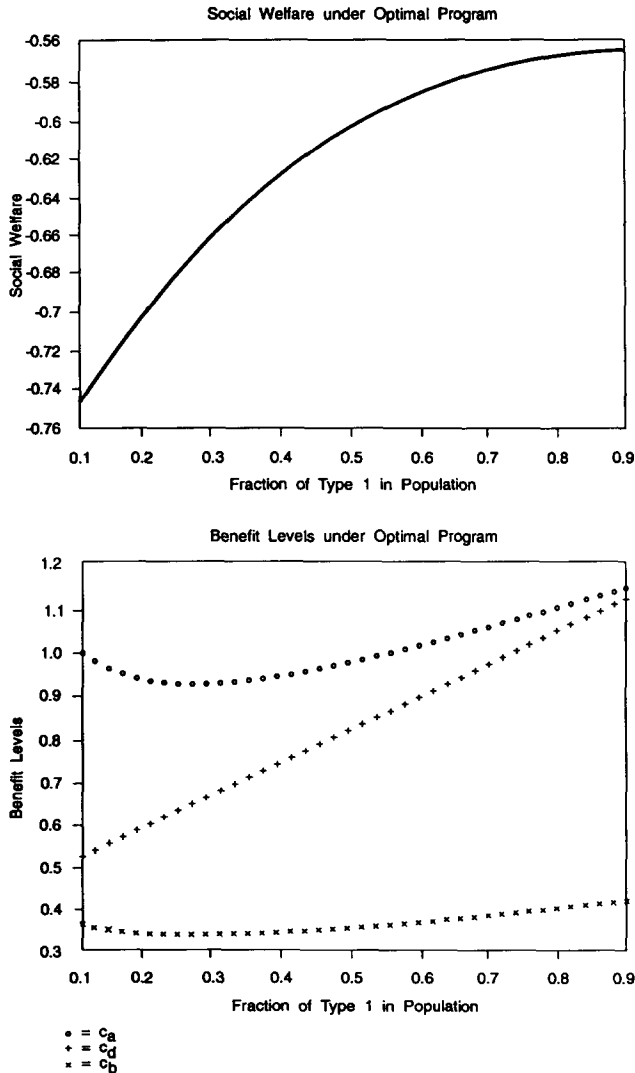


Fig. 3.

can verify that c_d is linear in f_1 in Configuration 5. From (37) one sees that c_b is a constant fraction of c_a . It is interesting that c_a and c_b are not monotone in f_1 , although social welfare rises as f_1 increases, since a rise in f_1 means that the share of workers with low disutility has increased.

Figs. 4-6 show the effects of variations in the disability screening parameters p_1 and p_2 . By definition, $p_2 > p_1$. Increases in p_1 , holding p_2

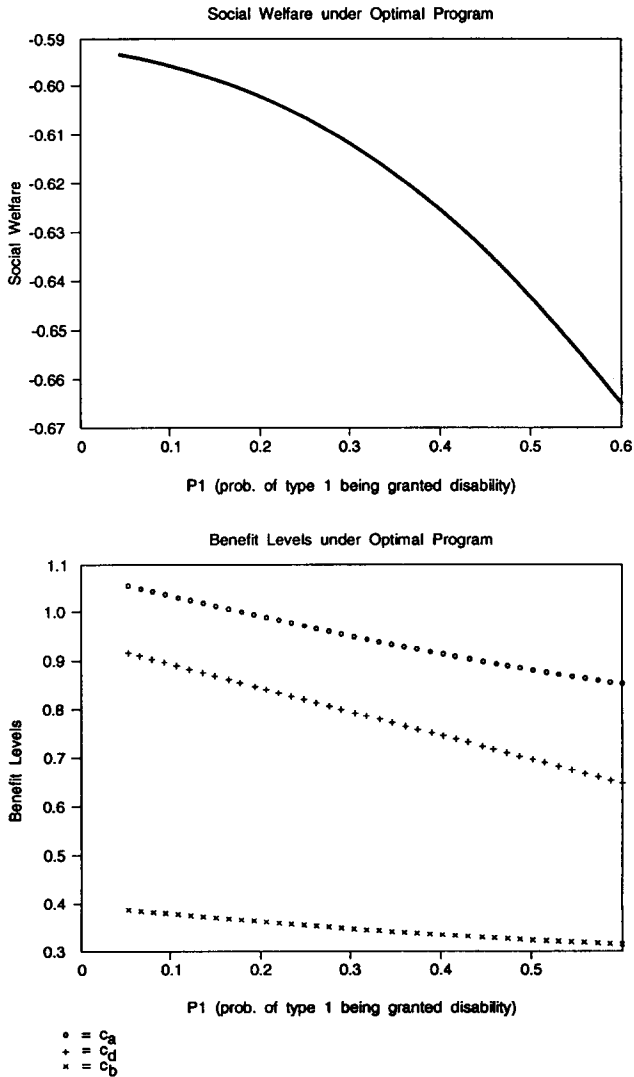


Fig. 4.

constant at the base setting of 0.75, cause a fall in social welfare and benefit levels (see Fig. 4). This happens because a larger fraction of type one's are granted disability, hence a smaller fraction of work, and there is less income generated in the economy. If p_1 is pushed into the range 0.62–0.75, Configuration 5 is no longer optimal; instead, Configuration 3 is.

An increase in p_2 , holding p_1 constant at the base setting of 0.25, causes

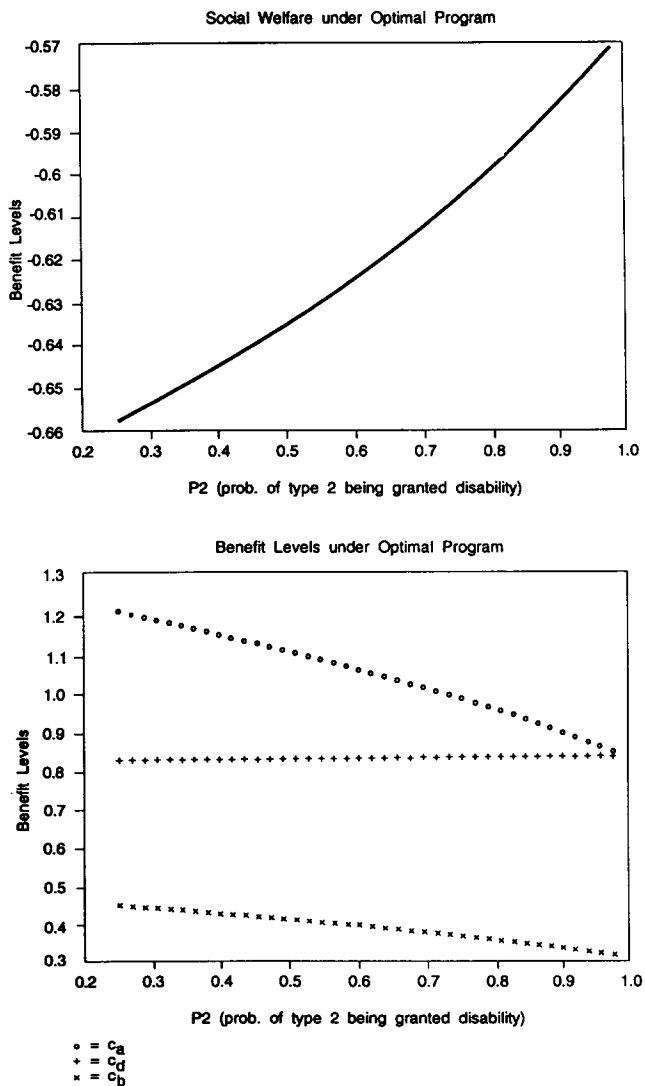


Fig. 5.

social welfare to rise (see Fig. 5). An increase in the probability of type two being awarded disability means that a larger share of income is directed toward type two's. Since the marginal utility of income is greater for type two's, social welfare rises. Note from (38) that c_d is independent of p_2 . Thus c_a and c_b fall to finance the increased size of the disability program.

Fig. 6 shows the effect of varying the ratio p_2/p_1 , holding the fraction

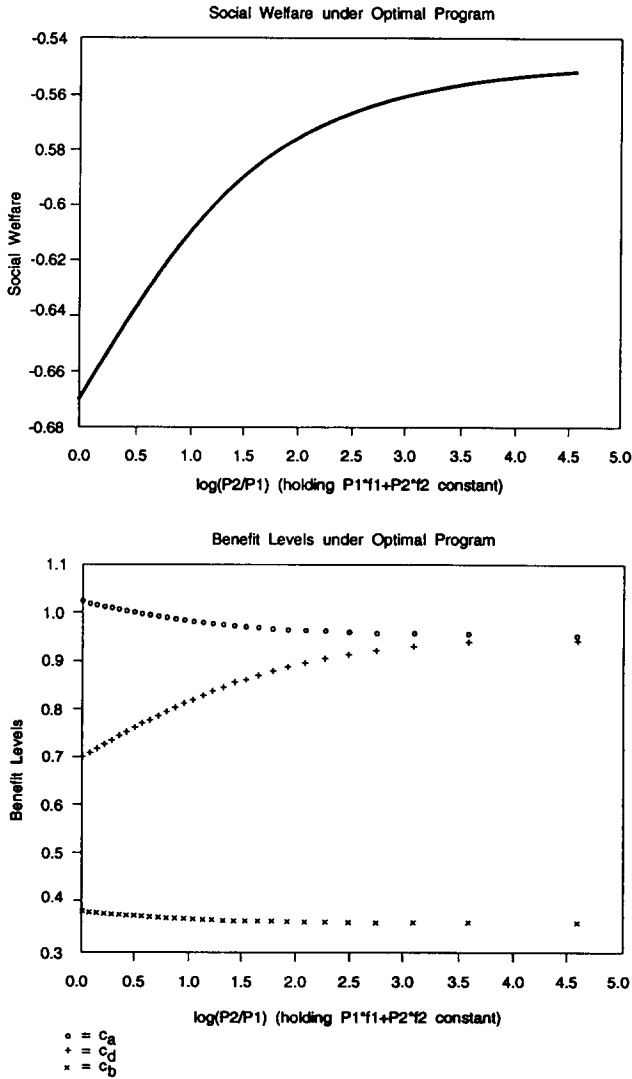


Fig. 6.

awarded disability, $p_1 f_1 + p_2 f_2$, constant at the base level of 0.5. This amounts to varying the efficiency of disability screening while holding awards constant. As the screening technology improves (p_2/p_1 increases), the incentive compatibility constraint weakens and hence social welfare rises. In the limit as $p_2/p_1 \rightarrow \infty$, the same benefits can be given to workers of type one and to non-workers of type two, i.e. $c_d = c_a$. Note that c_b is fixed at a constant fraction of c_a by (37), but c_b becomes irrelevant as the fraction of type two's awarded disability approaches 1.

10. Concluding remarks

Analysts of welfare and retirement programs have long recognized the central role of labor supply in the design of a good benefit program. Benefit design requires a balancing of income redistribution goals and labor supply disincentives. The same issue arises in the design of a disability benefit program, but it arises in a more complicated way because of the role of a (necessarily imperfect) screening device for disability. For example, completely blind adults are eligible for disability benefits. Yet many blind people choose to work instead. The severity of the medical problem is not the source of measurement difficulty, but rather the medical problem, alone, is not a sufficient guide to labor supply. In contrast, some problems, such as back problems, are notoriously difficult to evaluate as medical problems. Thus it is important to model the interaction between the screening device and labor supply.

There are a number of different ways that disability programs can be structured. In the United States, all individuals judged disabled by Social Security are eligible for benefits determined by the same benefit formula (that relates benefits to past earnings). In contrast, some countries have separate partial disability and total disability programs. The focus in this paper was not on the design of an optimal mechanism, but on the optimal use of a benefit system that resembles that in the United States. In the United States, the benefit formula used for retirement benefits was also used for disability benefits (adjusted for the length of earnings records, and between 62 and 65, for the actuarial adjustment for early retirement). The delay in the normal retirement age, which will be phased in early in the next century, increases the length of this overlap. Moreover, there has been concern about the way the benefit formula worked for young workers. The history of the program is a history of adaptation of a retirement system for disability, with recognition of anomalies, not design of a disability system. The main thrust of this paper is to put in focus the information that one needs in order to design a good system. A data set that links health problems with labor supply is a critical part of such information. Moreover,

following the employment experience of people denied disability benefits is also an important source of information. This paper has focused on how benefits should vary with labor supply effects, making clear the important role of such information. In addition, the structure lends itself to raising alternative design issues as well as issues of the right choice of parameters within the present design.

We have examined how benefits and disability standards would be set by a very knowledgeable social welfare maximizing government. We have assumed that it is costless to evaluate disability applicants. That is obviously wrong. If there were a social cost but no private cost to applying, the social cost would be paid by all the $1 - F(\theta_d)$ individuals who would apply for benefits. By introducing a fee for applying, the government could cut down the number of applicants. The value of such a fee depends on the distribution of disabilities among individuals dissuaded from applying. Particularly attractive would be a fee negatively related to the disutility of work. The absence of work for a prior period would be such a fee, although it represents a social cost for those who are rejected and for whom θ lies between θ_d and θ_b . Purely financial application fees avoid some of the social cost, but possibly not all since they may involve undesirable income redistributions (as would be apparent with wealth as well as disutility differences in the population). More important is the distribution of disincentives to apply. The optimal application fee structure would be an interesting extension of the model.

Acknowledgements

Research supported by the National Science Foundation and the Binational Science Foundation. Research assistance was done by Douglas Galbi.

References

- Akerlof, G., 1978, The economics of 'tagging' as applied to the optimal income tax, welfare programs, and man power planning, *American Economic Review* 68, 8–19.
- Bound, J., 1989, The health and earnings of rejected disability insurance applicants, *American Economic Review* 79, 482–503.
- Burtless, G. and R. Moffitt, 1984, The effect of Social Security benefits on the labor supply of the aged, in: H.J. Aaron and G. Burtless, eds., *Retirement and economic behavior* (Brookings Institution, Washington, DC) 135–174.
- Diamond, P. and J. Hausman, 1984, The retirement and unemployment behavior of older men, in: H.J. Aaron and G. Burtless, eds., *Retirement and economic behavior* (Brookings Institution, Washington, DC) 97–134.
- Diamond, P. and J. Mirrlees, 1978, A Model of social insurance with variable retirement, *Journal of Public Economics* 10, 295–336.

- Diamond, P. and J. Mirrlees, 1982, Social insurance with variable retirement and private savings, MIT Working Paper No. 296.
- Diamond, P. and J. Mirrlees, 1986, Payroll-tax financed social insurance with variable retirement, *Scandinavian Journal of Economics* 88, 25–50.
- Gustman, A. and T. Steinmeier, 1985, The 1983 Social Security reforms and labor supply adjustments of older individuals in the long run, *Journal of Labor Economics* 3, 237–253.
- Halpern, J. and J. Hausman, 1986, Choice under uncertainty: A model of applications for the Social Security disability insurance program, *Journal of Public Economics* 31, 131–161.
- Haveman, R., V. Halberstadt and R. Burkhauser, 1984, Public policy toward disabled workers: Cross-national analyses of economic impacts (Cornell University Press, Ithaca, NY).
- Leonard, J., 1986, Labor supply incentives and disincentives for disabled persons, in: M. Berkowitz and M.A. Hill, eds., *Disability and the labor market* (ILR Press, Ithaca, NY).
- Milgrom, P., 1979, Good news and bad news: Representation theorems and applications, *Northwestern DP* 407.
- Nagi, S., 1969, *Disability and rehabilitation: Legal, clinical, and self-concepts and measurement* (Ohio State University Press, Columbus, OH).
- Parsons, D., 1990, Social insurance with imperfect state verification, unpublished, Ohio State University.
- Smallwood, D., 1975, Competition, regulation, and product quality in the automobile insurance industry, in: A. Phillips, ed., *Promoting competition in regulated markets* (Brookings Institution, Washington, DC) 241–300.
- Stern, N., 1982, Optimum taxation with errors in administration, *Journal of Public Economics* 17, 181–211.