2000

2000-18 Evaluating Profiling as a Means of Allocating Government Services

Mark C. Berger

Dan Black

Jeffrey A. Smith

ISSN:0318-725X
ISBN:0-7714-2282-2

RESEARCH REPORT 2000-18

Evaluating Profiling as a Means of Allocating Government Services

by

Mark C. Berger Dan Black Jeffrey Smith

September 2000

Department of Economics
Social Science Centre
University of Western Ontario
London, Ontario, Canada
N6A 5C2
econref@julian.uwo.ca

Evaluating Profiling as a Means of Allocating Government Services

Mark C. Berger
University of Kentucky

Dan Black
Syracuse University

Jeffrey Smith
University of Western Ontario and NBER

Version of September 11, 2000

Berger is the Sturgill Professor of Economics at the University of Kentucky. Black is Professor of Economics and Senior Research Associate, Center for Policy Research, at Syracuse University. Smith is Associate Professor of Economics at the University of Western Ontario and a Faculty Research Fellow of the NBER. The authors thank Amitabh Chandra, Steve Allen and Roy Sigafus for research assistance and Markus Frölich and Chris O'Leary for helpful comments. Berger and Black thank the Center for Business and Economic Research at the University of Kentucky for research support. Smith thanks the Social Science and Humanities Research Council of Canada for financial assistance.

1. Introduction

Several alternative mechanisms exist for allocating government benefits and obligations. Among them, statistical profiling, which assigns individuals to different programs based on *predicted* outcomes or *predicted* program impacts, has lately attracted substantial interest in the policy and research communities.[1] This paper considers the question of how to evaluate profiling as an allocation mechanism, within the context of the choice between alternative allocation mechanisms for specific government programs. We argue that in the midst of the recent enthusiasm for profiling in North America, too little attention has been paid to the design and evaluation of these systems, with the result that they may not accomplish all that they could.

We place our discussion within the specific context of the use of profiling in the United States to allocate mandatory reemployment services to unemployment insurance (UI) recipients.2 In this program, data are collected on all persons starting a new spell of unemployment. These data are then used to predict each person's probability of

[1] Recent work in economics related to statistical treatment rules includes Black, Smith, Berger and Noel (2000), Dehejia (1999), Manski (1999,2000), and O'Leary, Decker and Wandner (1998).

[2] In the U.S., eligibility for UI benefits depends on having earned a specific amount in the first four of the five quarters prior to the claim. This minimum earnings level varies by state, with some states also requiring a certain number of weeks of employment in that period. The weekly benefit amount depends on previous earnings. In general, workers are eligible for at most 26 weeks of benefits. Workers who quit their job or who are fired for cause do not qualify for UI. In general, the alternative social assistance benefits (including Temporary Aid to Needy Families, food stamps, and general assistance) available in place of UI or when UI has been exhausted are not very generous, particularly for persons without children. See Storey and Neisner (1997) for more details and a comparison of the U.S. UI system with those of the other G-7 countries.

exhausting his or her unemployment insurance benefits. This prediction comes from a model estimated using data on the observable characteristics and UI benefit receipt durations of earlier cohorts of recipients. Those new claimants whose observable characteristics indicate a high probability of benefit exhaustion must then either participate in the mandatory reemployment services or give up their benefits. The services themselves can be interpreted either as a valuable opportunity to learn new employment-related skills, or as an in-kind tax on the leisure of the UI claimant.

Although we will consider the issue of evaluating a statistical profiling mechanism within the context of the UI profiling program, the issues raised here are completely general. Any government program that does not apply to everyone must have some allocation mechanism to determine who participates in it and who does not. In practice, these mechanisms vary widely across programs. In some cases, the allocation depends on deterministic rules, as in means-tested transfer programs or affirmative action programs that depend on membership in specific demographic groups. In other cases, service allocation depends on the discretion of government employees, perhaps within limits set by program eligibility rules. Many government job training programs around the world employ this allocation method, with government caseworkers, usually in consultation with the potential trainee, deciding on what, if any, program services to provide. The literature expresses great concern about this allocation method due to the possibility of cream-skimming, which occurs when caseworkers choose to serve those most likely to do well even without the program's help, usually with the goal of improving the program's performance relative to a performance management system.[3]

Statistical profiling represents something of an intermediate case between deterministic rules and caseworker discretion. It attempts to finely differentiate among potential service recipients but in a deterministic way. Making the right decisions about how to allocate services in various program contexts requires an understanding of how profiling works and the ability to evaluate its effectiveness *as an allocation mechanism*.

The two key issues that determine the effectiveness of a profiling system are the choice of the profiling variable – that is, the variable whose predicted value will determine the allocation of services – and the choice of variables to use in predicting the profiling variable. The optimal profiling variable depends critically on the social goal or goals underlying the profiling. A program whose primary goal in the allocation of services is equity – say, serving those most in need – will likely require a very different profiling variable than one whose goal in the allocation of services is efficiency – serving those with the largest net benefits from participation. In particular, there should be a strong relationship between the profiling variable and the social goal that the profiling mechanism exists to serve. In the case of UI profiling, for example, if the social goal is efficiency, then the expected duration of UI benefit receipt is a good choice for a profiling variable only if it varies positively with the net impact of the program. More generally, if achievement of the social goal of the program does not depend much on who is served, as

---

[3] See Heckman, Heinrich and Smith (1997) and Heckman, Smith and Taber (1996) for further discussion and evidence on the empirical importance of cream-skimming.

in a program that has the same (or no) impact on everyone, then profiling will likely not be the preferred allocation mechanism.

The choice of variables to profile on also depends on the availability of good data to use in constructing predicted values. For example, if the available data do a poor job of predicting individual impacts from service receipt, then attempting to profile on the basis of expected impacts will do little to enhance the efficiency of the service allocation mechanism. Profiling makes little sense if the available data lack the ability to sort persons based on the profiling variable.

One important theme of this paper is that the problem of evaluating profiling as an allocation mechanism and the problem of evaluating the impacts of the service being allocated by profiling are conceptually and practically distinct. A profiling system (or any other system of assignment to services) might do a good job of allocating an ineffective service or it might do a bad job of allocating an effective service. To make this distinction concrete, assume that efficiency is the goal of the allocation. Consider first an ineffective job training program that reduces the earnings of its participants by consuming time that would otherwise have been spent on useful job search.[4] If the extent of the earnings reduction for each participant depends on his or her observable characteristics, and if the profiling mechanism uses those characteristics to assign the program to those with the smallest earnings reductions, then it has done a good job of allocating a bad program. Second, consider a program that increases the earnings of women by more than the

---

[4] This example is not as unlikely as it might seem. Bloom, Orr, Cave, Bell and Doolittle (1993) report negative experimental impact estimates for young men in the U.S. Job Training Partnership Act (JTPA) program.

earnings of men, but costs the same amount for all participants. A profiling mechanism that does not assign more women than men to this program has inefficiently allocated participants to a good program.

In Section 2 of the paper, we consider the problems of choosing an allocation mechanism and of evaluating profiling as a mechanism of service allocation at a conceptual level. In Sections 3 and 4 of the paper, we consider the evaluation of a particular profiling system. We use as our example the profiling of new UI claimants in the state of Kentucky to receive mandatory reemployment services based on their expected duration of UI receipt. We argue, contrary to some of the recent literature, that it is possible to predict expected durations of UI receipt. Doing so, however, requires utilizing a fairly rich set of variables measuring past earnings, past employment and local economic conditions.

In Section 4, we evaluate UI profiling in Kentucky under the assumption that its goal is efficiency. Our evidence suggests that the current profiling system is not a particularly effective means for achieving this goal. Section 5 summarizes our main points.

2. Theoretical Issues in the Design of Profiling Systems

This section first considers the general problem of choosing among alternative methods for allocating government programs. Within the context of this important general problem, it then examines key issues in the design of statistical profiling systems at both a general level and in the specific context of UI profiling in the United States.

2.1 Choosing the Optimal Assignment Mechanism

We begin by defining some notation. Let $j \in \{1,...,J\}$ index the individuals in the population of interest. Let X_j denote a vector of observable characteristics of person j and let $T_j \in \{1,...,L\}$ denote the treatment assignment for person j. This notation allows for the general case of programs that provide multiple options. For example, in the context of a job training program, we might have $T=1$ for non-participation, $T=2$ for job search assistance, $T=3$ for subsidized on-the-job training and $T=4$ for classroom training. In the context of our UI profiling example, there are two options: regular UI and UI plus a requirement to participate in reemployment services in order to continue receiving benefits.

Now consider the three basic mechanisms for allocating government programs in terms of our notation. A deterministic system allocates programs based on the observed characteristics of individuals, such as their earnings, age, geographic location or race. Such a system constitutes a mapping ϕ from X to T, or

$$\phi(X): X \rightarrow T.$$

Let $\{T_{jDET}\}$ denote the resulting allocation.

In contrast, a profiling system assigns persons based on their predicted values of the profiling variable. For example, in UI profiling it is based on the estimated probability of benefit exhaustion in some states and on the predicted duration of UI benefit receipt in others. Call the profiling variable Q and its predicted value $\hat{Q}(X)$, where X is again some (probably different) set of observable variables related to each individual. In such systems, the profiling assignment rule is given by

$$\tau(\hat{Q}(X)) : \hat{Q}(X) \rightarrow T.$$

Let $\{T_{jPROF}\}$ denote the resulting allocation.

Once the predictive model has been estimated profiling also represents an allocation scheme that is a deterministic function of X. The distinction between the two systems centers on the role of the intermediate profiling variable Q and the fact that the role of the X's in the allocation process depends on their relationship to Q rather than being determined directly through a political decision process.

The final allocation mechanism is caseworker discretion. In our notation, we can describe this allocation as $\{T_{jCW}\}$. We assume that the caseworker allocation does not depend deterministically on X, though of course caseworkers are likely to take into account at least some X's in making their allocation decisions. Case workers may also take into account other variables W representing characteristics of the individuals being assigned or of the local economic and social environment that they observe but that are not available for use in implementing either deterministic treatment rules or profiling. Examples of W variables would include the motivation shown by the individual in meetings with the caseworker or detailed knowledge of local service providers. It is, of course, the potential importance of these W variables to the allocation process that makes caseworker discretion a reasonable alternative for some programs.

Following Manski (1999), the choice among alternative treatment mechanisms then collapses to a comparison of the sets of outcomes implied by them. In conceptual terms, the outcomes are evaluated based on some measure that corresponds to the goal of the allocation process. Let S be this measure, which could be, among other things, GDP, an equity measure such as a Gini coefficient or some index of group equality, or some combination of these. The socially optimal allocation mechanism produces the maximum value of S among the available alternatives.

To make this all concrete, consider the following example of a program in which there are two treatments, participation and non-participation, corresponding to $T = 1$ and $T = 0$, respectively. If the goal of the program allocation process is maximization of GNP net of program costs, then the preferred allocation mechanism is the one that assigns the program to those who benefit the most from it, net of program costs. Let Y_1 denote earnings net of program costs conditional on participation and Y_0 denote earnings conditional on non-participation. Assume that the program affects earnings and nothing else and that the impacts do not depend on the scale of the program. In this case, if the goal of the allocation mechanism is efficiency, we can let S equal the sum of realized earnings. Then

$$S_{DET} = \sum_j \left[Y_1 T_{jDET} + Y_0 (1 - T_{jDET}) \right]$$

is the realized value of S for deterministic allocation. The realized values of S for profiling and for caseworker discretion are defined analogously. The optimal allocation mechanism provides the maximum value of S.[5]

[5] In this case, the optimal allocation assigns persons with $Y_1 > Y_0$ to $T = 1$ and everyone else to $T = 0$.

2.2 Two Key Issues That Affect the Effectiveness of Profiling

This sub-section considers the two key issues in designing an effective profiling system: what variable Q to use as the profiling variable and what variables X to use in predicting Q. Addressing these questions can be considered the initial stages in a multi-stage optimization process that leads to the choice of the optimal allocation mechanism for a given program.

First consider the question of what variable to profile on. Suppose that there is a set of candidate profiling variables, $Q_k, k = 1,..., K$, and that for each Q_k the problem of which X to use as predictors (to which we will turn shortly) has already been solved. Thus, for each Q_k we have available predicted values $\hat{Q}(X)$ to use in profiling, where the X may differ across profiling variables. Let the program being allocated be $T = 1$ while the baseline, no-program state is denoted $T = 0$, and let $\{T_{j,k}\}$ be the allocation associated with profiling variable k.

Suppose that the goal of the profiling exercise is efficiency; that is, the aim is to assign the treatment so as to maximize the sum of the outcomes, just as in the example in the preceding subsection. Assume too that program costs are equal across persons and that the program only affects earnings. Then the choice problem is quite similar to that just considered, only the choice is now over profiling variables rather than over allocation mechanisms. In terms of our notation, the optimal choice of profiling variable k becomes

$$k_{OPT} = \underset{k}{\operatorname{argmax}} \left\{ \sum_j Y_j \right\} \text{ where } Y_j = T_{j,k} Y_1 + (1 - T_{j,k}) Y_0 ,$$

where Y_1 again denotes earnings with treatment net of program costs and Y_0 again denotes earnings without treatment. In words, the optimal profiling variable is the one that maximizes the attainment of the goals of the allocation mechanism. For example, if the goal is efficiency then the optimal profiling variable is the one whose allocation of the program maximizes the total net impact of the program. If the goal is equity as defined by a Gini coefficient, then the optimal choice of profiling variables is the one that maximizes the Gini coefficient.

The optimal choice among the candidate profiling variables may depend on the set of available predictor variables X. Changing the set of available X could change the optimal choice from among the Q_k by altering their relative abilities to sort among persons based on their contribution to the goals of the allocation mechanism.

Now consider the second key issue: what X to use to predict each Q_k.[6] In practice, X is often given, and the question becomes the extent to which the available X do a good job of predicting each candidate Q_k. If the available X do not predict a particular potential profiling variable Q_k very well, then Q_k is unlikely to be selected as the optimal Q in the optimization problem just described. Put somewhat differently, profiling on Q_k when the covariance between Q_k and $\hat{Q}_k(X)$ lies close to zero will do little to advance the goals of the program allocation mechanism.[7]

[6] Selection between alternative functional forms may also be an issue. We abstract from this issue here but discuss it briefly in the context of UI profiling in Section 3.
[7] Manski (1999) investigates the conditions under which it may be better to assign treatment without the use of any profiling variable.

The remainder of this section considers issues related to the interaction between the goals of the treatment allocation and the choice of the profiling variable. We revisit the issues considered in this sub-section in the context of UI profiling in Sections 3 and 4 below.

2.3 The Fable of the Benevolent Bureaucrat

Consider a government bureaucrat who wishes to do good. In our example, the bureaucrat has the job of deciding how to allocate reemployment services to unemployment insurance claimants, but the issues she faces are completely general. In allocating government services, goals such as equity in service allocation or service to particular subgroups such as youth, minorities or transfer program participants will often conflict with other goals related to efficiency or to the size of the program budget.

In our example, the bureaucrat wishes to both help the unemployed and reduce expenditures for the Unemployment Insurance (UI) system. Our bureaucrat has already learned one important lesson: she has clearly defined the goals that she wishes to advance though the allocation of services. Clarifying the goals of the allocation mechanism allows its design to proceed with the goals in mind. Knowing what the goals are also makes it possible to evaluate the allocation mechanism relative to those goals.

Our bureaucrat hopes that the services she allocates will improve the job search skills or increase the job readiness of those who receive them, thereby speeding up their return to work and perhaps increasing their earnings when they get there. Of course, a speedier return to work also reduces the amount that the UI system spends on benefits.

Much to her dismay, however, our bureaucrat cannot offer the reemployment services to everyone: her budget is insufficient.

As in all economic problems, our bureaucrat faces a trade-off. By allocating slots to some recipients, she denies slots to other recipients. Suppose that our bureaucrat believes that the duration of a person's unemployment spell is a good indicator of their welfare: the longer the spell of unemployment the worse off is the recipient. Helping long-term recipients, which is desirable on equity grounds, may conflict with the objective of limiting the expenditures of the UI system, which may be desirable for political or efficiency reasons. In terms of our notation, the bureaucrat faces a choice between two potential Q_k's, one representing the duration of unemployment and the other representing budgetary savings.

To see the trade-off, suppose that there are two types of recipients: those who take 15 weeks to find a new job and their worse-off counterparts who take 30 weeks. Our bureaucrat may wish to target the reemployment services on the long-term unemployed for equity reasons. Because unemployment benefits in the United States are limited to 26 weeks, however, by doing so she may be limiting the impact of the services on UI expenditures. If there is a common effect across all recipients of, say, seven weeks, then targeting the long-term unemployed will cost the system money. Each long-term UI recipient who receives services saves the system only three weeks of benefit payments, compared to a full seven weeks saved for each short-term recipient served. Whether or not our bureaucrat decides to target the long-term or the short-term unemployed will depend on how she trades off between helping the worst off and saving the UI system money.

Keeping with our example, however, we can further complicate our bureaucrat's life. Suppose that the impact of treatment depends, in some nontrivial way, on the characteristics of the unemployed. This complicates things because our bureaucrat does not necessarily know what characteristics are associated with the program's effectiveness. Even if she does have this information, additional complications still arise.

Consider three simple but informative cases. First, suppose that those with longer spells of unemployment have larger treatment effects than those with shorter spells. In this happy situation, our bureaucrat's desire to help those with the greatest need is reinforced by the heterogeneity of the treatment effect. This does not necessarily completely eliminate the conflict between the desire to help the long-term unemployed and the desire to reduce expenditures. Suppose, for example, that reemployment services reduce the duration of unemployment by nine weeks for the long-term unemployed. In this case, each long-term recipient served still saves the system only five weeks of benefit payments relative to the full seven weeks saved for each short-term UI recipient served.

In the second case, we exacerbate the conflict by supposing that those with the shortest spells have the largest treatment effects. Now our bureaucrat's desire to help those most in need strongly conflicts with her desire to save the government money.

These two cases, and the third case that follows, illustrate another general lesson from our fable: tradeoffs between alternative social goals in designing profiling systems are likely to be empirically important. Related to this, the fable illustrates that the form and extent of these tradeoffs may depend on empirical relationships between the impacts of the program being allocated and the equity-related characteristics of potential participants.

Learning something about these relationships, such as the one between the impact of reemployment services and the duration of unemployment in our fable, then becomes a necessary condition for setting up an effective profiling system (or for incremental reforms of existing systems).

In the third case, suppose that the effect of the program varies by some other characteristic, Z. To keep things concrete, suppose that recipients with large values of Z have larger reductions in the duration of their unemployment spells than recipients with small values of Z. For long-term unemployed workers with large values of Z, there is no conflict between our bureaucrat's two objectives. For smaller values of Z, however, the conflict arises once again. Should our bureaucrat provide services to long-term unemployed claimants with low values of Z or short-term unemployed claimants with high values of Z?

To illustrate the trade-off, suppose that our bureaucrat chooses to minimize the expenditures of the UI system. In this case, she simply orders the recipients by their Z values and then picks the cutoff point, say Z^m, so that she just exhausts her budget by serving all recipients with $Z > Z^m$. Put differently, she implements a deterministic allocation mechanism based on Z. Of course, because our bureaucrat prefers to serve the long-term unemployed, she is willing to give up some of those cost savings to indulge her taste for helping the more needy. Toward that end, define the values Z^S and Z^L for short-term and long-term recipients, respectively. Choose these values so that they correspond to the two recipients, one long-term and one short-term, whom she is indifferent between serving, given that she serves the persons with the highest Z values first within each subgroup. It is easy to show that $Z^S > Z^m > Z^L$, or that she discriminates toward treating the long-term recipients. Similarly, it is easy to show that the number of short-term recipients in the interval $[Z^m, Z^L]$ must equal the number of long-term recipients in $[Z^S, Z^m]$. The difference in expected UI benefit payments to these two groups reflects the costs of letting our bureaucrat indulge her taste for helping the long-term unemployed.

In terms of our notation, in the third case our bureaucrat defines a variable R equal to a weighted average of each claimant's expected duration of unemployment and Z and then adopts a deterministic rule based on R. That is, if we let U equal the duration of unemployment, our bureaucrat allocates services deterministically based on $R = R(U, Z)$.

In practice the duration of unemployment is unknown in advance, and Z might represent the predicted impact of services. In that case, the allocation mechanism consists of profiling based on a weighted average of the predicted values of two profiling variables: unemployment duration and the impact of services. This mechanism illustrates the third general lesson from our fable: in cases in which multiple goals guide the allocation of services, profiling based on a weighted average of variables related to the various goals may represent an effective compromise. Such a compromise has the virtue of making explicit the relative weights assigned to the various goals in a way that informing caseworkers of the various goals and then allowing them to use their discretion in trying to meet them does not.

2.4 Lessons from the Fable for the Current UI Profiling System

Our fable provides some stark commentary about the design of the current UI profiling system. The current program offers substantial leeway to the states in terms of the predictive variables (the $X's$) to include in the profiling model. At the same time, it requires states to use benefit exhaustion or expected benefit duration as the dependent variable – so that states must profile based on this outcome. There is no allowance for profiling based on the expected impact of the reemployment services, or on some combination of expected spell duration and expected impacts.

Such a program may be justified in a couple of different ways. First, the program administrator may be concerned only with helping the long-term recipients and not place any value on reducing UI expenditures. In such an environment, we may ascertain the opportunity cost of helping the long-term unemployed by comparing the savings resulting from the current profiling system to the savings that would result from a profiling system that allocated services based on expected benefit reductions.

From our conversations with people involved with the program, however, we get the impression that they do not perceive a conflict between helping those most in need and reducing UI expenditures. To avoid such a conflict, however, two conditions must hold. First, it must be the case that we can use the length of unemployment spells as a welfare indicator for recipients. If short-term unemployed persons are worse off than long-term unemployed persons, then concern for helping those most in need would require directing services toward short-term recipients. Second, it must be the case that the effect of reemployment services for the long-term unemployed exceeds that for the short-term unemployed. Because many of the weeks that the long-term unemployed remain unemployed occur after they have exhausted their 26 weeks of benefits, to generate at least comparable cost savings the long-term unemployed must have a larger effect of services.[8]

Are these conditions true? Obviously, the second condition is one that we may verify or refute with data. Section 4 presents some initial evidence on this question. In contrast, assessing the validity of the expected length of the unemployment spell as a welfare indicator is more problematic. Because we do not observe economic welfare, we must make inferences based on theory. What does the economic theory of job search tell us about the relationship between the unemployment spell duration and economic welfare? We address that question in the next subsection.

2.5 When Does Profiling on Expected Duration Improve Equity?

Search theory tells us that the relationship between unemployment spell duration and welfare depends on the source of the variation. To see why, consider the following simple search model. Let wages be given by $w = \mu + k\varepsilon$, where μ is the mean, ε is a mean zero error term with density function $f(\varepsilon)$ and distribution function $F(\varepsilon)$, and k is a constant initially equal to one. Each firm that an unemployed person visits during his or her search costs c. Letting V_i denote the value function for searching the ith firm, we have:

$$V_i = \int_{(w_r - \mu)/k}^{\infty} w f(\varepsilon)\, d\varepsilon + \delta F\left(\frac{w_r - \mu}{k}\right) V_{i+1} - c, \tag{4}$$

[8] Machin and Manning (1999) discuss the trade-off between focusing policy on the short-term unemployed and the long-term unemployed in a European context.

where δ is the individual's discount factor and w_r is their reservation wage. If we assume that the search process is stationary in the sense that $V_i = V_{i+1} = V$ for all *i*, then we may rewrite equation (4) as

$$V = \frac{\int_{(w_r-\mu)/k}^{\infty} w\, f(\varepsilon)\, d\varepsilon - c}{1 - \delta F\left(\dfrac{w_r - \mu}{k}\right)} \tag{5}$$

Each unemployed person will select a reservation wage that maximizes his or her value function. The necessary condition reduces to

$$V\,\delta - w_r = 0. \tag{6}$$

Using equation (6), we may consider the impact of an increase in the mean of the wage distribution, μ, which yields

$$0 < \frac{d\,w_r}{d\mu} < 1. \tag{7}$$

Define $\varepsilon_r = (w_r - \mu)/k$ and notice that equation (7) implies that $d\,\varepsilon_r / d\mu < 0$. The expected number of searches to find an acceptable offer is simply $[1 - F(\varepsilon_r)]^{-1}$. A reduction in ε_r implies, therefore, a reduction in the expected number of searches and the length of time unemployed. This leaves the unemployed person better off and shortens the expected duration of the unemployed spell. This occurs because the increase in the mean of the wage offer distribution increases the opportunity cost of waiting for a relatively better wage. Thus, in a world in which persons differed only by the mean of their wage offer distributions, we could legitimately use the expected duration of the unemployment spell as a welfare measure.

Now suppose that we increase k. This increases the dispersion in wage offers without increasing the mean. Again, using equation (6), we have

$$\frac{dw_r}{dk} > 0. \tag{8}$$

Thus, unemployed persons respond by increasing their reservation wages, which implies that ε_r has also increased and therefore $[1 - F(\varepsilon_r)]^{-1}$ has increased as well. The increase in k increases their welfare, but also increases their expected duration of unemployment. In a world in which persons differ only in the dispersion of their wage offers, therefore, unemployment duration perfectly correlates with welfare, but those with shorter spells are worse off.

We do not need to rely on differences in the distribution of wage offers to generate differences in welfare. Suppose that unemployed persons face the same distribution of wages, but differ in their search costs, perhaps because some of them find the leisure associated with unemployment more appealing than others do. Again, using equation (6) we have

$$\frac{dw_r}{dc} < 0. \tag{9}$$

In response to an increase in c, the unemployed decrease their reservation wages, which means that ε_r and $[1-F(\varepsilon_r)]^{-1}$ also decrease. Once again, the expected duration of unemployment is positively associated with welfare.

In the complexity of actual markets, of course, unemployed people undoubtedly differ in their search costs, the means of their wage offer distributions, and the dispersion of their wage offer distributions as well. Inferring welfare from the expected duration of unemployment spells in such a world is extremely difficult. Indeed, the relationship need not even be monotonic. Thus, while targeting the long-term unemployed for assistance may constitute a sensible objective if equity is the primary goal of the UI profiling, it remains to be shown empirically whether targeting the long-term unemployed improves the equity of the UI system.

3. Profiling in Practice: Can Long-Term Unemployment be Predicted?

In this section9 we return to the second of the two key issues that underlie the effectiveness of profiling systems: what is the best way to use the available X to predict the profiling variable. We consider this issue in the context of the development of UI profiling models in different states in the United States, paying particular attention to the model developed in the state of Kentucky. This issue is important because a profiling system will not advance the goals of the allocation mechanism if the available X do not sort people as a function of the profiling variable.

---

9 This section draws on Berger, Black, Chandra and Allen (1997).

In June 1994, the Commonwealth of Kentucky, along with Delaware, Florida, New Jersey, and Oregon, was selected as a prototype state for the implementation of a system of profiling unemployment insurance (UI) claimants and providing them with reemployment services. In most cases, Kentucky provides only low intensity reemployment services. Almost all claimants profiled into treatment receive an assessment of their skills and interests. Claimants assessed to be "job ready" receive limited services such as referrals to employers for interviews and job search workshops. Claimants not found "job ready" may receive these services as well as referrals to education and training opportunities, such as short occupational training courses at community colleges. Noel (1998) documents that about 10 percent of Kentucky claimants profiled into services were referred to such opportunities.

The Center for Business and Economic Research (CBER) at the University of Kentucky was given the responsibility for developing and estimating the Kentucky Profiling Model (KPM). The model predicts the fraction of their 26 weeks of UI benefits that claimants will use up. CBER used five years of claimant data supplemented with data from other administrative data sources to estimate the model. The claimant data from the UI system include information on past UI benefit receipt as well as on past earnings by calendar quarter, because past earnings determine both eligibility for benefits and benefit levels. The other aggregate data from external sources used in the model include county unemployment rates and employment changes in the county of residence and nearby "commuting" counties broken down by one digit industry.

Estimation of models to predict the length of spells of UI benefit receipt is a difficult task. In Figure 1, we graph the fraction of potential benefits received for UI claimants starting spells in Kentucky in 1994. The data have a large mass of observations (almost 40 percent) at one – corresponding to the exhaustion of all potential benefits and therefore a benefit receipt spell of 26 weeks. More than 60 percent of the claimants, however, do not exhaust their benefits, and there is considerable variation among this group in the fraction that they receive. To see this variation more clearly, Figure 2 depicts the fraction of potential benefits received for those recipients who did not exhaust their benefits.[10]

Worden (1993) provides the baseline model used in profiling UI claimants. She used a logit model with UI benefit exhaustion as the dependent variable. Many other researchers (e.g., Eberts and O'Leary, 1996) have followed her lead and used binary choice models in developing profiling systems. A dichotomous model, however, treats each recipient who does not exhaust as identical, thereby ignoring much useful information in the data. For example, Berger, et al. (1997) compare high school dropout, work experience and earnings in the past year among persons who exhaust different fractions of their UI benefits. They find that claimants who use up 75 to 99 percent of their UI benefits look much more like claimants who exhaust all of their benefits than they do like claimants who exhaust less than 25 percent of their benefits.

[10] The reason for the saw-tooth pattern in Figure 2 is that UI claimants in Kentucky can file for benefits two weeks at a time. Most claimants receive the benefits for weeks 1 and 2 in their first payment, but those claimants who are slow to apply for benefits receive weeks 1 to 3 in their first payment and from then on are one week off from the other claimants.

In addition to using a binary variable for UI benefit exhaustion as its dependent variable, Worden's (1993) model also included relatively few covariates. Most other states followed Worden's lead in using relatively few variables in their models, in part due to concerns about complexity and in part to reduce the costs of assembling the data.[11] O'Leary, Decker, and Wandner (1998) note in their literature review that the model for Washington state, which is one of the larger state models, includes 36 covariates. In contrast, the model for the state of Pennsylvania uses only eight covariates. Given the use of a dichotomous dependent variable, the limited number of covariates, and the inherent difficulties in predicting the duration of unemployment, one might suspect that these models have relatively little explanatory power. This is indeed the case.

O'Leary, Decker, and Wandner (1998) estimated the Pennsylvania and Washington models, and then used their estimates to examine how well the two models distinguish long-term from short-term recipients. Table 1 reproduces their findings. For Pennsylvania and Washington, it compares the exhaustion rate among the 25 percent of claimants with the highest predicted exhaustion probabilities to that among the remaining 75 percent. Those in the top quartile are 11 to 13 percent more likely to exhaust their benefits. Interestingly, a comparison of the top and bottom 50 percent yields a smaller gap, only 8.1 percent for Pennsylvania and 7.9 percent for Washington state.

[11] Dickinson, Decker and Kreutzer (1999) summarize an evaluation of the WPRS in 12 states. That evaluation finds that two of the 12 had unrecognized implementation problems with their profiling procedures. While it is not clear that these problems were related to model complexity, the general issue of whether state information technology staff (as opposed to professors of economics, who implemented the system in Kentucky) can effectively implement a complicated profiling model is a serious one.

In contrast, the KPM appears to be much more precise in the identification of those likely to exhaust their UI benefits. Table 1 compares the exhaustion rate for those with predicted benefit receipt durations above the 60th percentile to those with predicted benefit durations below this level. The difference, 24.8 percent, is much greater than the maximum differences reported by O'Leary, Decker, and Wandner (1998) for the Pennsylvania and Washington models.[12]

Why are the differences in predictive power so large between the KPM and the prototypical profiling models from the other states? Two basic reasons suggest themselves. First, the KPM did not rely on a dichotomous variable of whether the claimant exhausted UI benefits, but rather used the fraction of benefits received as a continuous variable. It is possible that the additional information contained in the fraction of benefits received variable dramatically improves the predictive power of the estimates. Second, the KPM contains over 140 different covariates, including variables representing past earnings, industry, experience, tenure, and unemployment insurance participation.

In developing the KPM, Berger, et al. (1997) experimented with a variety of estimation procedures that allowed them to rank recipients by the likelihood that they would exhaust their UI benefits or by their expected duration of UI benefit receipt. They estimated each model on a 90 percent sample of UI recipients and used the remaining 10 percent to calculate the out-of-sample predictive power. In particular, they used the

[12] Some evidence from outside the U.S. provides mixed support for these pessimistic findings. Payne, Casey, Payne, and Connolly (1996), using data from the U.K., have limited success in predicting long-term unemployment. In contrast, Wong, Henson and Roy (1999) report moderate success in predicting unemployment spells longer than one year using Canadian data.

estimated coefficients from each model to predict the proportion of benefits used up by each recipient in the 10 percent validation sample. The 10 percent sample was then divided into subgroups predicted to consume more or less than 60 percent of their potential benefits. This cutoff was chosen because it was thought that predicted use of at least 60 percent of potential benefits would approximate the cutoff for treatment under the profiling program.[13]

Table 2 reports Berger et al.'s (1997) findings for a sample of UI recipients from the western part of Kentucky. They report the mean fraction of benefits exhausted for each of six models: a linear regression model using the fraction of benefits exhausted as the dependent variable, logit and probit models using a dichotomous dependent variable indicating that the worker exhausted benefits, a double-limit tobit using the fraction of benefits exhausted as the dependent variable, a Cox proportional hazard model using the fraction of benefits exhausted as the dependent variable, and random assignment.

The fourth column of Table 2 ranks the models in terms of the fraction of potential benefits received by the sixty percent with the highest probabilities of benefit exhaustion or longest expected durations. As can be seen, the tobit model selects the group that exhausts the most of their benefits, but its advantage over the other models is modest. For instance, using the logit model to select the group to receive reemployment services increases the fraction of benefits exhausted by 11.26 percent over random assignment. Using the tobit model rather than the logit model, however, increases the fraction of benefits exhausted by only 0.44 percent, or about a 3.9 percent improvement. Thus, the superior predictive

[13] The actual treatment rate was much higher because the booming Kentucky economy

performance of the Kentucky model results from the rich set of covariates rather than the form of the dependent variable or the estimation procedure used.

This discussion of attempts to predict UI benefit exhaustion and/or the duration of UI benefit receipt in the context of profiling UI recipients provides an important general lesson: the *X*'s matter. The results presented in this section show that the predictive power for the same or roughly the same profiling variable can vary substantially depending on the predictive variables employed. As already noted, profiling will not accomplish its goals if the variables used to predict the profiling variable do so only very poorly. Thus, the analysis in this section has important implications for the design of profiling systems in general and for the choice of the profiling variable in contexts where the available *X* are limited in particular. Moreover, our analysis suggests that while profiling on predicted UI benefit receipt durations may be a good idea in Kentucky, profiling on the probability of benefit exhaustion in states using models that only weakly predict exhaustion almost certainly is not.

4. Does Profiling Maximize Program Impacts?

We now examine the first key issue in designing a profiling system – whether or not the profiling variable is related to the goals of the allocation mechanism – in the context of UI profiling in Kentucky.

---

reduced the number of people filing UI claims.

4.1 Do claimants with longer expected durations have larger treatment effects?

We showed in the previous section that the KPM does a relatively good job of sorting claimants based on their expected duration of UI receipt. We now examine whether the earnings impact of the treatment is larger for those claimants with higher predicted durations of UI benefit receipt.

A few additional details about Kentucky's profiling system begin our discussion. As described in the preceding section, the KPM produces an estimate of the fraction of their potential UI benefits that each claimant will collect. In operating the system, these continuous estimates are collapsed into a discrete profiling score from 1 to 20. A score of 20 means that the model predicts the claimant to collect from 95 to 100 percent of their potential UI benefits, a 19 means that it predicts receipt of 90 to 95 percent and so on. In each local UI office in Kentucky in each week, claimants starting new spells are ranked by their profiling score. The office then provides the treatment to claimants starting with the highest score and continuing until it has used up its budget for the week. When there are more claimants than can be served at the marginal profiling score in a given office in a given week, the treatment is randomly assigned within the marginal group.

Overall, because of the strong Kentucky economy at this time, only 16.92 percent of the 57,779 claimants during the period covered by our data (October 1994 to June 1996) did not receive the treatment. Figure 3 depicts the empirical distribution of the profiling scores for all of the claimants in our data. The modal profiling score was 18, with 16.4 percent of claimants receiving this score. Only 9.82 percent received a score of 10 or below. The scores 13, 16, and 18 roughly divide the sample into quartiles.

We now present estimates of the impact of the profiling treatment on claimants' earnings in the six quarters after the start of their UI spell as a function of their profiling score. Importantly, the treatment consists of the requirement to receive reemployment services, not actual receipt of these services. Because many treated claimants leave UI before receiving services but after being required to receive services, the two possible treatment definitions differ from one another both conceptually and empirically.

We use the data from those persons in the marginal profiling score cells where random assignment took place in constructing these estimates. This has the advantage of providing estimates free from selection bias but the disadvantage that the estimates do not apply to the full sample without additional assumptions.[14] It also means that we can only provide impact estimates for profiling scores from 6 to 19. Everyone with a score of 20 received treatment, and there were no marginal cells with scores of 1 to 5 requiring random assignment during the period of our data.

The experimental sample includes 1,236 treated persons and 745 controls drawn form 286 marginal office-week-profiling score cells. Their average age is around 37 and their average years of schooling are just over 12, which represents completion of high school in the U.S. About 60 percent are male and over 90 percent are white. Average earnings in the year before the UI claim are just under $20,000.[15]

We obtain our experimental estimates from the a regression of the form:

[14] Black, et al. (2000) discuss the construction and interpretation of the experimental impact estimates in more detail. Non-experimental estimates of the impact of treatment as a function of the profiling score for the scores from 1 to 19 also do not reveal any systematic relationship. These estimates are available from the authors upon request.

$$Y_i = \alpha_0 + \alpha_1 X_i + \beta_6 P_{6,i} T_i + \ldots + \beta_{19} P_{i,19} T_i + \mu_j + u_i$$

where Y_i denotes the earnings of claimant i in the six quarters after the start of the UI claim, X_i is a vector of pre-random assignment characteristics, $P_{i,r}$ is an indicator variable for claimant i having profiling score r, T_i is a treatment indicator for claimant i, μ_j are fixed effects for each marginal cell, respectively, u_i is the mean zero error term, and $\alpha_0, \alpha_1, \beta_6, \ldots \beta_{19}$ are parameters to be estimated along with the fixed effects.[16] The cell fixed effects take account of the fact that the random assignment ratio differed across cells.[17]

Figure 4 graphs the impact of treatment by profiling score. There is little evidence that the magnitude of the treatment effect increases with the profiling score. From this we can conclude either that efficiency in allocation is not the goal of UI profiling in Kentucky, or that, if efficiency is the goal, the profiling system is doing little to advance it.[18]

---

[15] The characteristics of the full treated sample are surprisingly similar for these variables. See Table 1 of Black, et al. (2000).

[16] Conditioning on pre-random assignment characteristics in the regression increases the precision of the impact estimates without disrupting the random assignment, as random assignment makes these characteristics independent of treatment.

[17] In a common effect world, this regression specification is efficient. In a world where the effect of treatment differs across cells, other weighting schemes may be preferred for certain policy questions. See Black, et al. (2000) for more details.

[18] We calculated the impact estimates as a function of the predicted probability of exhausting UI benefits using the estimated impacts of the U.S. UI bonus experiments that O'Leary, Decker and Wandner (1998) present in their Table 6. In the UI bonus program, UI claimants received a bonus of US$500 if they found a job within the first part of their UI spell and kept it for a defined period (a few months). We find that the impacts of the bonus program do not vary systematically with the probability of benefit exhaustion in either of the states – Pennsylvania and Washington – that they consider. Our calculations based on their estimates are available from the authors upon request.

4.2 Can we identify those with larger treatment effects?

Allocating treatment effectively based on expected impacts requires knowing which subgroups have the largest impacts.[19] Even in cases where efficiency is not the sole goal of an allocation mechanism, this information allows the estimation of the opportunity costs, in efficiency terms, of pursuing other goals.

The empirical analysis in Section 5.1, however, suggests that identifying subgroups with large impacts may prove difficult in practice. Every 95 percent confidence interval in Figure 4 contains zero, even though the experimental sample includes almost 2000 observations.[20] Clearly, obtaining precise estimates of subgroup impacts can require very large samples, particularly when the dependent variable of interest, earnings in our case, has a high variance within the relevant population.

Black, et al. (2000) document another reason why locating those individuals with the largest impact may prove difficult in this context. They use the experimental subsample to document a sharp increase in exits from the UI system in the second week of benefit receipt among those required to receive reemployment services in order to continue

[19] We skirt the important issue of exactly which impacts ought to constitute the profiling variable in a system that attempts to profile on expected impacts. This issue in turn has several dimensions. Should only impacts on the government's budget be considered? Or should impacts on individual earnings or employment be considered separately from reductions in UI payments (or other social program expenditures)? Should short-term impacts be used or should long-term impacts (appropriately discounted) be used? The latter provide more information, but require the profiling system to be based on older data.

[20] Black, et al. (2000) present similar estimates that combine the profiling scores within the experimental sample into four groups of roughly equal size. In this case, some individual estimates do attain statistical significance. The basic conclusion, hoeweer, of no systematic relationship between the profiling score and the impact estimates remains in place.

receiving benefits. In that week, 5.4 percent of the control group exits UI, compared to 12.7 percent of the treatment group. This early exit appears to result from notification of the requirement to receive services, rather than from the services themselves, which do not start until later in the spell. This evidence suggests that a relatively small number of people exiting the program very early generate much of the overall impact of the treatment. An alternative profiling system that attempted to profile based on expected impacts would face the difficult task of trying to identify this relatively small sub-population of claimants.

5. Summary and Conclusions

Although profiling has achieved remarkable popularity in North America as a mechanism for allocating government programs, we argue that too little attention has been paid to the design and evaluation of profiling as an allocation mechanism, with the result that existing profiling systems may not be particularly effective. In this paper, we develop a conceptual framework for the choice among alternative mechanisms for allocating government benefits and obligations. In regard to profiling, we have argued that being explicit about the goal of the allocation mechanism makes possible evaluation of the allocation mechanism in light of its goals, and also aids in the selection of an appropriate profiling variable. We have highlighted possible trade-offs between equity and efficiency in the choice of allocation mechanisms based on profiling. And, finally, we have shown the critical role that the selection of the predictor variables plays in designing an effective profiling system.

Applying our views about the evaluation of profiling as an allocation mechanism to UI profiling in the U.S., we reach the following conclusions. First, many existing state UI profiling systems do a poor job of predicting the profiling variable. As our theoretical discussion shows, this failure makes it virtually impossible to advance the goals of the profiling system. Second, we show that this failure results largely from a lack of covariates in the profiling model. Predicting UI benefit exhaustion or the duration of UI benefit receipt is not impossible, as some have argued (see, e.g., OECD, 1998), but it takes more in the way of X's than most existing profiling models include. Finally, we show that even in Kentucky, where the profiling model does do a good job of predicting the profiling variable, the profiling variable is not systematically related to the impact of the treatment being allocated.

REFERENCES

Berger, Mark, Dan Black, Amitabh Chandra and Steve Allen. 1997. "Kentucky's Statistical Model of Worker Profiling for Unemployment Insurance." *Kentucky Journal of Economics and Business.* 16: 1-18.

Black, Dan, Jeffrey Smith, Mark Berger and Brett Noel. 2000. "Is the Threat of Reemployment Servies More Effective than the Services Themselves? Experimental Evidence from the UI System." Unpublished manuscript. London, Canada: University of Western Ontario.

Bloom, Howard, Larry Orr, George Cave, Stephen Bell and Fred Doolittle. 1993. *The National JTPA Study: Title II-A Impacts on Earnings and Employment at 18 Months.* Bethesda, MD: Abt Associates.

Dehejia, Rajeev. 1999. "Program Evaluation as a Decision Problem." NBER Working Paper #6954.

Dickinson, Katharine, Paul Decker and Suzanne Kreutzer. 1999. "Evaluation of Worker Profiling and Reemployment Services Systems." Unpublished manuscript.

Eberts, Randall and Chrstopher O'Leary. 1996. "Design of the Worker Profiling and Reemployment Services System and Evaluation in Michigan." W.E. Upjohn Institute for Employment Research Working Paper #96-41. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, April.

Heckman, James, Carolyn Heinrich and Jeffrey Smith. 1997. "Assessing the Performance of Performance Standards in Public Bureaucracies." *American Economic Review* 87(2): 389-395.

Heckman, James, Jeffrey Smith and Christopher Taber. 1996. "What Do Bureaucrats Do? The Effects of Performance Standards and Bureaucratic Preferences on Acceptance into the JTPA Program" in Gary Libecap, ed., *Advances in the Study of Entrepreneurship, Innovation and Economic Growth: Reinventing Government and the Problem of Bureaucracy, Volume 7.* Greenwich, CT: JAI Press. 191-217.

Machin, Stephen and Alan Manning. 1999. "The Causes and Consequences of Longterm Unemployment in Europe." In Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Volume 3C. Amsterdam: North Holland. 3085-3139.

Manski, Charles. 1999. "Statistical Treatment Rules for Heterogeneous Populations With Application to Randomized Experiments." Unpublished manuscript. Chicago, IL: Northwestern University.

Manski, Charles. 2000. "Identification Problems and Decisions Under Ambiguity: Empirical Analysis of Treatment Response and Normative Analysis of Treatment Choice." *Journal of Econometrics*. 95(2): 415-442.

Noel, Brett. 1998. *Two Essays on Unemployment Insurance*. Unpublished dissertation, University of Kentucky.

O'Leary, Christopher, Paul Decker, and Stephen Wandner. 1998. "Reemployment Bonuses and Profiling," W.E. Upjohn Working Paper #98-51. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Organisation for Economic Co-operation and Development. 1998. *Early Identification of Job Seekers at Risk of Long-Term Unemployment: The Role of Profiling*. Paris: OECD

Payne, Joan, Bernard Casey, Clive Payne, and Sara Connolly. 1996. *Long-term Unemployment: Individual Risk Factors and Outcomes.* London, UK: Policy Studies Institute.

Storey, James and Jennifer Neisner. 1997. "Unemployment Compensation in the Group of Seven Nations," in Christopher O'Leary and Stephen Wandner, eds., *Unemployment Insurance in the United States*. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research. 599-668.

Wong, Ging, Harold Henson, and Arun Roy. 1999. "Long-Term Unemployment in Canada: Outlook and Policy Options." Ottawa, Canada: Human Resources Development Canada, Ottawa, Canada.

Worden, Kellen. 1993. "Profiling Dislocated Workers for Early Referral to Reemployment Services." Unpublished manuscript. Washington, DC: U.S. Department of Labor.

TABLE 1
THE ABILITY OF PROFILING MODELS TO PREDICT BENEFIT EXHAUSTION

	Top of Distribution	Bottom of Distribution	Difference
Pennsylvania model	38.2	25.7	12.5
Washington model	35.3	24.6	10.7
Kentucky model	78.3	53.5	24.8

Notes: The division for the Pennsylvania and Washington models is the top 25 percent predicted benefit exhaustion probabilities versus the bottom 75 percent. These are the maximum differences reported by O'Leary, Decker, and Wandner (1998). For Kentucky, the division is the top 60 percent of predicted benefit receipt durations and the remaining 40 percent. Estimation and prediction for the Pennsylvania and Washington model are done using the same samples; the model for Kentucky is estimated on a 90 percent sub-sample of claimants while the prediction is performed using the remaining 10 percent.

TABLE 2
VALIDATION STUDY OF ALTERNATIVE PROFILING MODELS IN KENTUCKY

Model	Dependent Variable	Fraction of Benefits Exhausted	Rank
Tobit	Fraction benefits exhausted	78.26%	1
Ordinary Least Squares (OLS)	Fraction benefits exhausted	77.99	2
Probit	Exhaustion (binary)	77.83	3
Logit	Exhaustion (binary)	77.82	4
Cox	Fraction benefits exhausted	77.44	5
Random assignment	---	66.56	6

Notes: Authors' calculations, Kentucky UI Claims data. Estimation for Kentucky is estimated on a 90 percent sub-sample of claimants while the prediction is performed using the remaining 10 percent.

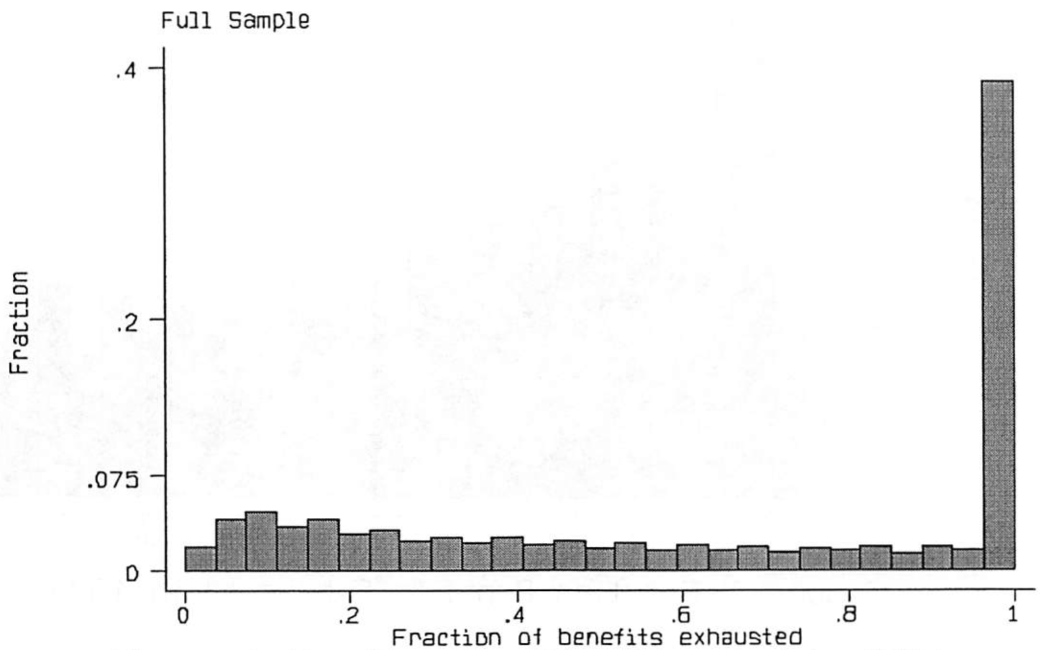

Figure 1: Fraction Benefits Exhausted in 1994

Notes: Authors' calculations, Kentucky UI Claims data.

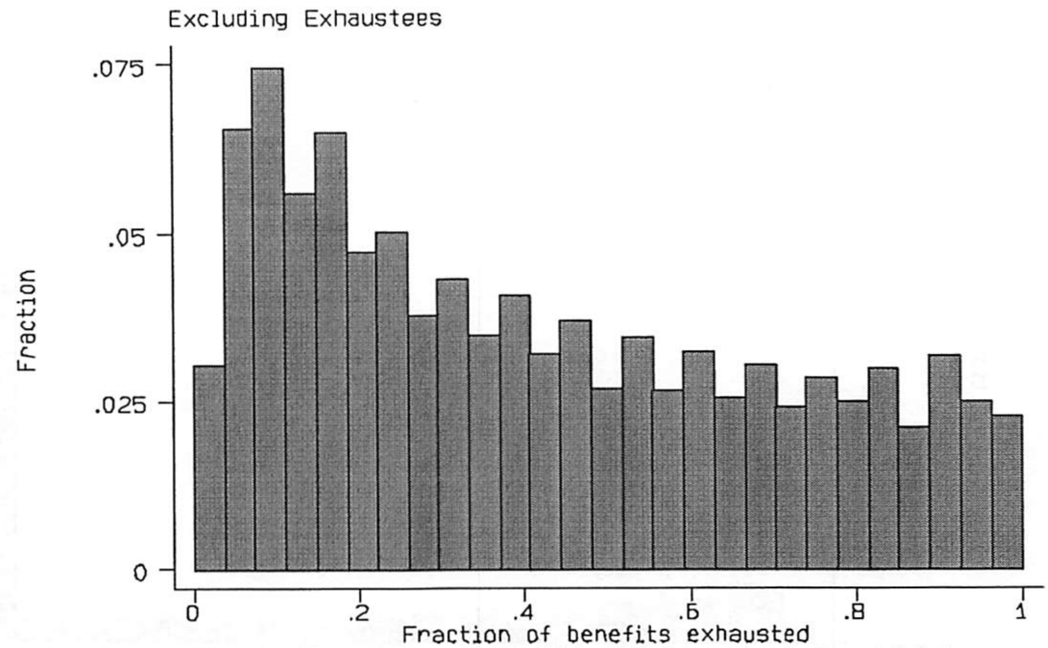

Figure 2: Fraction Benefits Exhausted in 1994

Notes: Authors' calculations, Kentucky UI Claims data.

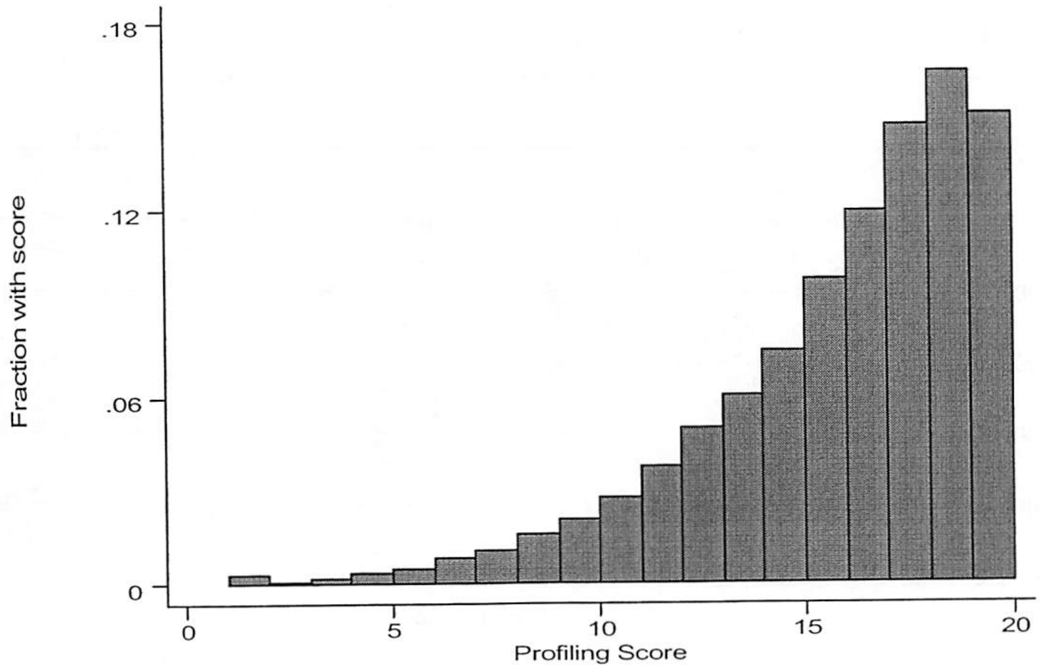

Figure 3: Empirical Distribution of Profiling Scores

Notes: Authors' calculations, Kentucky UI Claims data

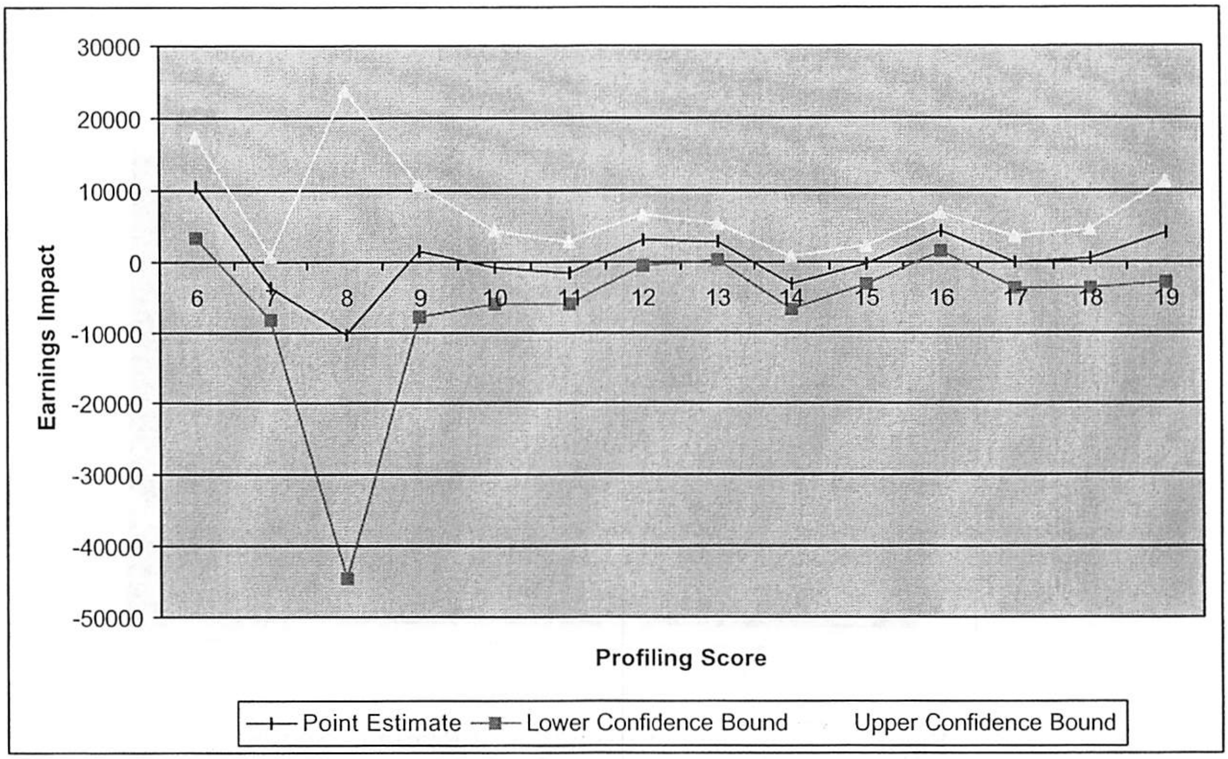

Figure 4: Estimates of Earnings Impacts by Profiling Score

Notes: Authors' calculations, Kentucky UI Claims data