To Train or Not To Train:
Optimal Treatment Assignment Rules Using Welfare-to-Work Experiments

John V. Pepper
Department of Economics
University of Virginia
P.O. Box 400182
Charlottesville, VA 22904-4182
jvp3m@virginia.edu

February 12, 2002

Abstract
To Train or Not To Train:
Optimal Treatment Assignment Rules Using Welfare-to-Work Experiments

Abstract:
Planners often face the especially difficult and important task of assigning programs or treatments to optimize outcomes. Using the recent welfare-to-work reforms as an illustration, this paper considers the normative problem of how administrators might use data from randomized experiments to assign treatments. Under the new welfare system, state governments must design welfare programs to optimize employment. With experimental results in-hand, planners observe the average effect of training on employment but may not observe the individual effect of training. If the effect of a treatment varies across individuals, the planner faces a decision problem under ambiguity (Manski, 1998). In this setting, I find a straightforward proposition formalizes conditions under which a planner should reject particular decision rules as being inferior. An optimal decision rule, however, may not be revealed. In the absence of strong assumptions about the degree of heterogeneity in the population or the information known by the planner, the data are inconclusive about the efficacy of most assignment rules.

1.   Introduction

      Decision-makers often face the especially difficult and important task of assigning programs to heterogeneous populations. In the United States welfare system, for example, state planners must design and implement programs that both provide assistance and, at the same time, meet various employment standards. In this setting, a single program is unlikely to improve the employment prospects for everyone in a caseload which includes individuals with a broad range of skills, backgrounds and challenges (Eberts, 1997; Pavetti, L. et al., 1999; Pepper, 2001).   Instead, planners must use the available information to optimally assign programs to individual recipients.[1]

      In this paper I evaluate the caseworkers' decision problem in different informational settings. Using the basic framework found in Manski (1998, 2000) and Manski, Newman and Pepper (forthcoming), I assume the planner wants to choose a treatment assignment rule to maximize an expected outcome, say the population employment probability. That is, the planner wants to maximize a utilitarian social welfare function.[2]   The planner is assumed to observe summary results from an idealized social experiment and may also know certain characteristics of the recipient. Four well-known experiments conducted by the Manpower Demonstration Research Corporation (MDRC) in the 1980s are used to illustrate the methods.

      This general informational setting in which the planner combines experiments with covariate information to make optimal treatment choices is commonplace under the new welfare regime.

---

[1] Given the recent reforms, the problems faced by welfare caseworkers are particularly germane. However, these decision problems are not unique. For example, police and judges decide how to treat (potential) offenders, educators assign students to classes and institutions, doctors and counselors prescribe treatments, editors accept or reject manuscripts, and so on. In each of these cases and many more, the planner may combine empirical evidence on the effectiveness of different treatment regimes with additional information about the individual recipient.

[2] Of course, one might consider evaluating other features of the distribution of outcomes. See, for example, Heckman, Smith and Clements (1997).

Caseworkers often observe detailed background characteristics of the recipients and experimental

analyses are thought to provide relevant data. In fact, employment and job-readiness programs like those

evaluated by the MDRC were used, in part, to motivate the federal reform, have been used to inform

planners in the new regime, and are a key component of every state program.[3]

In Section 2, I describe the MDRC demonstrations and characterize data from ideal randomized

experiments that abstract from concerns over the validity of the demonstration. Section 3 then outlines the

basic methodology that considers the normative problem of optimal treatment choice given heterogeneous

populations. Caseworkers are responsible for matching individual recipients with the "appropriate" welfare

program. The feasible assignment rules and the planner's ability to evaluate them depend on the available

information about treatments and outcomes. If the caseworker has prior information regarding how

different treatments affect each recipient, she can maximize each person's outcome. In particular, she

would assign training to everyone who "benefits" from training (that is, persons who would work if given

training but not otherwise), and cash assistance to everyone else. In general, however, planners are not

likely to have this level of detailed information.

A more pragmatic setting arises under the assumption that the planner has less extensive

information on the effectiveness of different treatments. Rather than knowing the individual response

function, the planner may only observe the employment probabilities revealed by the experiment. In this

---

[3] Michalopoulos and Schwartz (2000), evaluate 20 work-first demonstration programs, including the MDRC's SWIM
program, in a report prepared for the United States Department of Health and Human Services on the potential
effectiveness of different welfare to work programs in the new regime. The Riverside program in California has been
used to motivate work first reform at the state and national levels (Hotz, Imbens, Klerman, 2000). While most
demonstration programs, including the MDRC programs evaluated in this paper, show modest employment effects,
the Riverside program increased the employment rate from 35.3% for the control group to 49% of the treatment group.

partial information case, the best solution is to choose a single treatment that maximizes the employment probability for each observed subgroup. If the employment probability under training exceeds the employment probability under cash assistance, she would assign training.

Section 4 examines the asymmetric informational setting where a planner observes covariates that are not revealed in the experiment. In practice, caseworkers observe detailed information on recipients including demographics, work and welfare histories, schooling, neighborhood characteristics, family background, and other socio-economic indicators. The experimental outcomes for these observed sub-populations, however, might be unknown for several reasons. First, welfare-to-work experiments often do not record detailed background information on the recipients. Second, the published results from the experiments from which caseworkers are likely to base their inferences often do not report detailed covariate information. Gueron and Pauly's (1991) important evaluation of the MDRC experiments, for example, reports employment probabilities for the treatment and control groups, but does not present results for subgroups. Finally, caseworkers may be prohibited from explicitly using certain information including the age, race and gender of the recipient when assigning treatments.

In this informational setting, the experiment does not reveal the mean outcome for the subgroups identified by the planner. Thus, the caseworker may not be able to maximize the employment probability given the observed information. Rather, the planner must make a decision given ambiguous information about the optimal assignment rule.

The conclusions that can be drawn depend critically on the available data and prior information the planner can bring to bear. If data on the outcome of interest are combined with sufficiently strong assumptions, the outcome probability under different assignment rules may be identified, implying a well-defined assignment process. In practice, the most common assumption is that all persons benefit from

training (i.e., homogenous effects), in which case all persons should receive a single treatment. Parametric latent variable models describing how treatments are selected and outcomes determined may also identify the outcome probability under alternative treatment rules (see, for example, Dehejia, forthcoming).

A social planner, concerned about the credibility of her findings to policymakers and the public, might be inclined to impose more conservative assumptions. Indeed, as in Pepper (2001), I evaluate what can be learned about various assignment rules given weak assumptions on the process determining outcomes and prior information. In this conservative setting, I find a straightforward proposition formalizes conditions under which a planner should reject particular decision rules as being inferior. For instance, assigning everyone to receive cash-assistance might be strictly dominated by assigning everyone to receive training. An optimal decision rule, however, cannot be identified. In the absence of strong assumptions about the degree of heterogeneity in the population or the information known by the planner, the data are inconclusive about the efficacy of most assignment rules.

Section 5 concludes by considering practical solutions to the decision making process under ambiguity. What might the planner do when the experiment does not reveal the optimal decision rule?

2.   What Welfare-to-Work Experiments Reveal to the Caseworker

This section describes the welfare-to-work experiments. I consider experiments that evaluate two alternative treatments: the standard benefit program and assignment to a welfare-to-work training program. I assume an ideal experimental design where the subjects are randomly selected and assigned to one of the two mutually exclusive treatments, the subjects do not interact with each other, and the program administrators are not influenced by the experiment. Subjects may or may not comply with their assigned treatment so that the experiment identifies the effect of the intention-to-treat. Finally, I also assume that

the samples are large enough that the planner may abstract from sampling variability when interpreting the empirical evidence.

Section 2.1 considers what information these experiments reveal about employment outcomes under mandatory training. Section 2.2 describes the four experiments conducted by the Manpower Demonstration Research Corporation (MDRC) during the mid-1980s. As suggested above, it seems likely that program administrators and caseworkers may use experiments conducted over the past 30 years to evaluate and consider optimal assignment rules. After all, these experiments provide the only source of information for many of the innovative programs state and local governments might consider adopting.

2.1   What Do Welfare-to-Work Experiments Reveal?

What do the data reveal? To evaluate this basic questions, it is useful to distinguish between the outcome that would occur were a welfare recipient to have been assigned to training $y(1)$, and the outcome that would occur were she to have received the standard benefits, $y(0)$. In particular, let $y(\cdot)$ equal one if the individual would have participated in the labor force after the treatment period and zero otherwise. Let z denote the actual treatment received, where $z = 1$ if assigned to training and 0 otherwise.

For those who were assigned to training ($z = 1$) the employment indicator $y(1)$ is observed but $y(0)$ is latent, while for those who received the standard benefits ($z = 0$) the outcome $y(0)$ is observed but $y(1)$ is latent. Thus, the data reveal the employment probability for those who were assigned to training, $P[y(1) = 1 \mid z = 1]$, and for those who received standard benefits, $P[y(0) = 1 \mid z = 0]$.

In social experiments, the actual treatment received is randomly assigned so that the treatment, z, is statistically independent of the labor force participation indicators, $y(1)$ and $y(0)$. That is, the labor force participation probability of those who were assigned to training, $P[y(1) = 1 \mid z = 1]$, reveals the

outcome that would occur if everyone where to receive training, P[y(1) = 1]. Likewise, the employment probability of those assigned to the control group, P[y(0) = 1 | z = 0], reveals the outcome that would occur if the entire caseload received the standard benefits package, P[y(0) = 1].

In practice, whether particular experiments actually reveal the distribution of outcomes under mandatory treatment policies may be of considerable disagreement. There are many well-known and important critiques.[4] A program may not be properly implemented, so that the outcomes, y(1) and y(0), depend upon the realized treatment z. Even if properly run, a more fundamental question is whether the demonstration program operates in the same fashion as it would if it were actually implemented. Certainly, concerns about external validity are germane (Campbell and Stanley, 1966; Hotz, Imbens, and Mortimer, 1999): Macro-feedback effects (Garfinkel, Manski and Michalopolous, 1992), Hawthorne Effects, and entry effects (Heckman, 1992; Moffitt, 1992) all suggest that small scale demonstration projects may not reveal the outcomes that would occur if the program were instituted on a larger scale. Furthermore, the economy and the welfare system have undergone major changes over the last decade that may not be reflected in many of the relevant experiments. [5]

In this paper, I abstract from these concerns by assuming that the demonstrations observed by the planners identify the effects of being assigned to a job-training program. That is, the data are assumed to reveal the labor force participation probability if all recipients are assigned to training, P[y(1) = 1], or if all

---

[4] See Campbell, D. and J. Stanley. (1966), Hausman and Wise (1985) and Manski and Garfinkel (1992) for general critiques of the experimental methodology. Wiseman (1991) and Greenberg and Wiseman (1992) critically examine the MDRC demonstrations.

[5] Hotz et al. (1999) consider how variability in the population and in the programs components may compromise the external validity of the MDRC experiments. In this analysis I implicitly assume the programs components and the population dis tribution do not vary under the new regime.

recipients are given the standard benefits, P[y(0) = 1]. Maintaining this best-case assumption, I focus on the resulting decision problems that have received almost no attention in the literature.

2.2   MDRC Experiments

During the mid-1980s, the Manpower Demonstration Research Corporation (MDRC) evaluated four welfare-to-work training programs: the Arkansas WORK Program, the Baltimore Options Program, the San Diego Saturation Work Initiative Model (SWIM), and the Virginia Employment Services Program (ESP).   The MDRC randomly selected samples of size 1127, 2757, 3211, and 3150, in Arkansas, Baltimore, San Diego and Virginia, respectively.   For each program, welfare recipients were randomly assigned to either participate in a basic work or training activity, or to receive the standard benefits package.   For each subject, the data reveal the treatment received – training or the standard welfare benefits - and numerous labor market and welfare participation outcome measures. In this paper, the outcome variable of interest is whether or not the respondent participated in the labor force two years after treatment.

These MDRC experiments appear particularly relevant for evaluating the types of welfare and training programs which might be adopted under new regime. These evaluations are generally regarded as well designed and implemented social experiments (see, for example, Greenberg and Wiseman (1992), and Wiseman (1991)).   Furthermore, each evaluation was broad based and mandatory. All single parent families receiving AFDC and whose children were at least 6 years of age were required to participate.[6]

---

[6] Noncompliance with the assigned treatment led to sanctions or possible expulsion from the program. Over 50 percent of the caseload complied with the assigned training. Those that did not comply either exited the welfare system, were sanctioned, or were excused for reasonable cause.

Finally, the programs generally stressed labor force participation rather than human capital development. The modest training programs focused on supervised job search and unpaid work assignments. Educational activities were only offered in limited cases. For additional details on these experiments see Gueron and Pauly (1991) and Friedlander and Burtless (1995).

Table 1 displays the estimated employment probability for the treatment and control groups. In each case, the estimates confirm the well-established results that job-training programs slightly increase the probability of employment. Under the Virginia Employment Service Program (ESP), for instance, the labor force participation probability if all welfare recipients receive training is estimated to be 39.0 percent. If instead, all welfare recipients receive the standard benefits, these data suggest that 33.9 percent would be working after two years. Thus, the ESP increases the probability of employment by 0.051.

3.   The Caseworker's Problem

The caseworker's problem is to optimally assign treatments using the employment probabilities revealed by a welfare-to-work demonstration. To evaluate this problem it is useful to divide the population into those that are and are not affected by treatment. The treatment assignment process only affects some individuals. In particular, a fraction $P[\, y(1) = 1 \mid y(0) = 0 \,]$ of the caseload "benefits" from job training, while a fraction $P[\, y(1) = 0 \mid y(0) = 1 \,]$ "benefits" from the standard program. The remainder are unaffected by the treatment policy, with some fraction $P[\, y(1) = 1 \mid y(0) = 1 \,]$ participating in the labor force regardless of the treatment and some fraction $P[\, y(1) = 0 \mid y(0) = 0 \,]$ unemployed regardless.

Suppose that the planner wants to choose a treatment rule, z_m, to maximize the employment probability, $P[\, y\,(z_m) = 1 \,]$. Ideally, the planner would assign training to those that benefit from training, and cash assistance to everyone else. A fundamental identification problem occurs in that data cannot reveal the counterfactual outcomes of interest (see Manski, 1997; Pepper, 2001). One cannot, for

example, observe the labor market outcomes of an individual assigned to receive training if she were to have instead received the standard benefits. Likewise, we cannot observe how a recipient who was given cash-assistance would have behaved had she instead received job training. Thus, the outcome will depend critically on the planner's knowledge of treatment response.

In this section, I consider two extreme information settings. Section 3.1 considers the optimal assignment rule in the case where the planner fully observes the response functions, $\{y(1), y(0)\}$, for each individual. Section 3.2 considers the optimal assignment rule when the planner knows nothing more than is revealed by the experiments, namely the employment probability under mandatory training, $P[\ y(1) = 1]$, and mandatory standard benefits, $P[\ y(0) = 1]$.

3.1   Treatment Assignment with Full Information

Consider the best case scenario where caseworkers know the employment indicators$\{\ y_i(1), y_i(0)\ \}$ for each individual and select the treatment, job training or standard benefits, to maximize the labor force participation probability. That is, for each individual i, the caseworker assigns training so that

(1.) $y_i(z_m) = \max\ [y_i(1),\ y_i(0)].$

In this full information scenario, the treatment selection policy will maximize the labor force participation probability.

Since some fraction of the caseload will remain unemployed regardless of the assignment process, the realized employment probability is

(2.) $P[y(z_m) = 1] = \ 1 - P[\ y(1) = 0 \cap y(0) = 0\ ].$

To implement this optimal assignment rule, the planner uses prior information regarding the response functions of each recipient. The planner does not use the experiment to inform her decisions.

Still, the experiments might reveal ex-ante information about the optimal employment rate revealed in (2). Although the experiment cannot reveal what fraction of the caseload remains unemployed regardless, information on the employment probabilities under uniform treatments can be use to bound this joint probability (Manski, 1997, Pepper, 2001). Intuitively, under this assignment rule administrators can do no worse in terms of maximizing the employment probability of welfare recipients than what would have occurred if all recipients were assigned to job training, and no better than the sum of the two employment probabilities (see Frechet, 1951 and Manski, 1997). That is,

(3.) $\text{Max}\{P[\ y(1) = 1\ ],\ P[\ y(0) = 1\ ]\ \} \ \#\ P[\ y(z_m) = 1\ ]\ \#\ \text{Min}\{1,\ P[\ y(1) = 1\ ] + P[\ y(0) = 1\ ]\ \}$

(Manski, 1997, Proposition 6).

Even in this best-case model, where planners with rational expectations maximize the expected outcome, there remains much uncertainty. Consider, for instance, Virginia's ESP program. If planners combine this program with an outcome optimization assignment rule, the estimated bounds imply that at least 39.0 percent and at most 72.9 percent of welfare recipients will be employed after two years.

3.2   Treatment Assignment with Partial Information

In practice, caseworkers are unlikely to be able to distinguish among the outcomes $y_i(1)$ and $y_i(0)$ for each individual, i. A planner who does not have advanced knowledge of the outcomes of the response

functions for particular individuals may not implement the optimal assignment rule in Equation (1). Instead, treatment decisions must be made with some degree of uncertainty.

Assume the caseworker observes covariates X for each member of the population and knows the employment probability for each sub-population, $P[\ y(")=1\mid x\ ]$. The planner cannot, however, distinguish among persons with the same covariates and so cannot implement treatment rules that effectively differentiate among these persons. In this second best world, the planner is only concerned with the distribution of outcomes across the observed sub-populations, not with the experiences of particular persons. Treatment choice must be based on the employment probability, $P[\ y(")=1\mid x\ ]$, alone. Formally, a caseworker that observes certain covariates X assigns training so that

(4.) $P[\ y(z_m)=1\mid x\ ]=\max\{P[y(1)=1\mid x],\ P[\ y(0)=1\mid x\ ]\ \}.$[7]

Each member of a sub-population defined by X receives the same treatment.

If, for example, the caseworker does not observe any covariates, the optimal assignment rule either gives all persons training or all standard benefits: $P[\ y(z_m)=1\ ]=\max\{P[y(1)=1\ ],\ P[\ y(0)=1]\}$. In this setting, the results in Table 1 suggest that all recipients should be assigned to training, in which case about one-third of the caseload will work under the Baltimore, San Diego and Virginia programs.

Although the optimal assignment rule in the partial information setting maximizes the employment probability conditional on the observed covariates, the assignment rule in Equation (4) is weakly dominated by the full information assignment rule described in Equation (1). To see this, note that the employment

---

[7]This decision rule is implicit in statistical profiling models used to target services. See, for example, Berger, Black, and Smith (2000), and Eberts (1997).

probability under this partial information setting equals the lower bound (Equation 3) under full information. The two assignment rules are equal if being assigned to a particular treatment, say training, never reduces the likelihood of participating in the labor force. That is, $P[\ y(1) = 0 \cap y(0) = 1 \mid x\ ] = 0$. After all, the planner can do no better in terms of maximizing the employment probability than assigning everyone to training and no worse than assigning everyone standard benefits. If instead, the effects of training are heterogeneous, with some fraction unaffected, some fraction employed only if assigned to training, and some fraction employed only if given standard benefits, the second best allocation rule is not optimal. Without full information, some persons will benefit from the assignment rule in Equation (4) and others will be hurt.

4.   Treatment Choice with Asymmetric Information

Thus far, I have considered two extreme informational settings. Either the planner knows the effect of treatment for each recipient, in which case she can optimize outcomes for each individual, or the planner has no prior information about treatment response, in which case she must base treatment choice on the available empirical evidence alone. In this section, I explore the implications of a middle ground situation which is likely to apply in practice. In particular, the planner knows the effect of training on the employment probability from a welfare-to-work experiment. She also observe characteristics, X, of each member of the population and knows the population distribution, P(x), of these observed covariates. These characteristics, however, are not revealed for the experimental subjects or at least by the published summary results.

Observing covariates enables the planner to choose a treatment rule that differentiates among members of the population. However, without prior knowledge of the employment probabilities conditional

on these covariates, the planner cannot systematically use this information to choose a better treatment rule and may unintentionally choose a worse treatment rule.[8]

What treatment rule should the planner choose to optimize employment when the response functions are not observed? Section 4.1 evaluates this question in the general setting where the planner can uniquely identify each individual and thus can consider any heterogeneous assignment rule. Section 4.2, as in Manski (1998), illustrates the general method by considering the special case where the planner only observes a binary covariate, say whether or not the recipient has relevant labor market experience. In both settings, the planner faces a decision problem under ambiguity (see Manksi, 1998, 2000). In ambiguous settings, there is generally no optimal way to decide among the alternatives. Rather, the best the planner can do is rule out certain options.

4.1   Treatment Assignment Rules With Complete Covariate Information

Suppose the caseworker uniquely observes each member, i, of the caseload but does not know the individual response function, $y_i(\cdot)$. Rather, from the experiment, the planner knows $P[y(1)]$ and $P[y(0)]$. The planner's objective is to maximize employment among the caseload given the experiment and information identifying each recipient. In this setting, the caseworker must decide among any permutation

---

[8] There are many situations where the response probability, $P[y(")]$ may not be observed. At the most basic level, we draw inferences on the employment probability using a sample of observations, not the population. That is, rather than observing the true response function we only observe consistent estimate of the employment probabilities (see Manski, 2000). Even if we abstract from concerns of statistical variability, a large literature on identification problems suggest that both observational and experimental data alone may not fully reveal the response functions of interest. In observational studies, one must address the selection problem that the decision to participate in a training program is not random and thus the observed data alone cannot reveal $P[y(") = 1]$. Experiments in which the selection process is designed to be random may still not reveal the response function if there remain concerns about internal and external validity. In this paper, I abstract from concerns about statistical variability as well as concerns over the validity of the experiments (see Section 2) and focus instead on the ambiguity created by asymmetric information between the planner and the experiment.

of assignment rules from the extreme mandatory training or mandatory cash assistance rules evaluated in the second best informational setting above, to heterogeneous assignment rules like those adopted in the full information setting.

To illustrate how the identification problem faced by the planner complicates the assignment process, suppose a caseworker is faced with assigning the ESP program evaluated in Virginia. The estimates displayed in Table 1 imply that the labor force participation probability would be 39.0 percent under mandatory training and 33.9 percent under standard benefits. Many joint distributions are consistent with these marginal probabilities. Consider the two extreme cases. Table 2A displays the joint distribution with the strongest positive correlation between the two outcomes so that the treatment has the smallest possible influence on employment. Only 5.1 percent of respondents benefit from training, and no one benefits from the standard program. Given this distribution, at most 39.0 percent of the caseload will participate in the labor force while at least 33.9 percent will participate. Table 2B displays the other extreme where the outcomes exhibit the strongest negative correlation so that the treatment has the largest possible effect on outcomes. Regardless of the selection method, at least 27.1 percent of the caseload will be unemployed. The labor force participation outcomes of the remaining 72.9 percent of the caseload depend upon the assignment process.

Using these extreme distributions, one can evaluate what is known about different assignment policies. Suppose, for example, that the planner considers four different possible rules: training everyone, training half the caseload, training one-quarter of the caseload, and not training anyone. Under the two uniform assignment rules the experiment reveals the employment probability. If all persons were trained, 39.0 percent of the caseload would work. If all persons were given standard benefits, 33.9 percent would work.

If assignment is heterogeneous, the results are less certain. Suppose the planner assigns half the caseload to be trained. In the best case, all of the recipients assigned to training would benefit from training so that the labor force participation probability would be 72.9 percent (P[y(1) = 1] + P[y(0) = 1]), or the upper bound in the full information setting. In the worst case, the planner assigns training to the 33.9 percent who benefit from cash assistance, and the entire caseload might be unemployed. Thus, under this assignment rule, between [0, 72.9] percent of the caseload will be employed. Likewise, under an assignment rule where 25 percent are assigned to training, the employment probability will lie between [8.9, 58.9].

Clearly, mandatory training is preferred to mandatory standard benefits. For that matter, rules that train less 5.1 percent of the caseload are dominated by a mandatory training rule. Other heterogeneous rules are undominated. The planner cannot know whether the optimal policy is to train everyone, to train half the caseload, or to train substantially fewer. The answer to this question depends on two unknown factors: the fraction of the caseload affected by treatment and efficacy of the assignment policy. If there is substantial heterogeneity in treatment response (Table 2b), the planner can do much better or much worse than a mandatory training policy.

To formalize this planner's problem, it is useful to split the decision into two parts: first determine the optimal fraction to train and then decide which respondents to train. Since the planner has no prior information that effectively leads her to pick those who are trained and those who are given standard benefits, the first stage decision problem completely summarizes the choice problem.

Assume the planner decides on the fraction of welfare recipients who receive job training in order to maximize employment. Let $p = P[z_m = 1]$ be the fraction of recipients who will receive training, and 1-

$p = P[z_m = 0]$ be the fraction who will receive standard benefits. Then, using the law of total probability, we know that

(5.) $\qquad P[\ y(z_m) = 1\ ] = P[\ y(1) = 1\ |\ z_m = 1]\cdot p\ +\ P[\ y(0) = 1\ |\ z_m = 0]\cdot(1\text{-}p).$

The planner does not know the employment probabilities for those who will be trained, $P[\ y(1) = 1\ |\ z_m = 1]$, nor for those who will receive standard benefits, $P[\ y(0) = 1\ |\ z_m = 0]$. After all, the planner does not know the response functions for each individual, i.

Information on the observed outcomes under uniform treatment policies, however, can be used to bound the labor force participation probability. The experiment reveals the probability a recipient works if everyone is assigned to training $P[y(1) = 1\ ]$ while our interest is in learning the labor force participation probability for those who will be assigned to training. The relationship between these two probabilities is highlighted using the law of total probability to write

(6.) $\qquad P[y(1) = 1\ ]\ =\ P[y(1) = 1^{\star}\ z_m = 1]\cdot p + P[y(1) = 1^{\star}\ z_m = 0]\cdot(1\text{-}p).$

Since the unknown probability $P[y(1) = 1^{\star}\ z_m = 0]$ lies in the interval $[0, 1]$ we can bound the labor force participation probability for recipients who will be assigned to training. In particular,

(7.) $\qquad \text{Max}[\ 0, (P[y(1) = 1\ |\ x]\text{-}1 + p)/p\ ]\ \#\ P[y(1) = 1^{\star}x, z_m = 1]\ \#\ \text{Min}[\ 1, P[y(1) = 1\ |\ x]\ /\ p\ ].$

Analogous bounds can be derived for the labor force participation probability for those who will receive

standard benefits, $P[y(0) = 1^\star z_m = 0]$.

From Equations (5) and (7) it follows that

$$\max\{\ 0, P[\ y(1) = 1\ ] - (1-p)\ \} + \max\{\ 0, P[\ y(0) = 1\ ] - p\}$$

(8.) $$\#\ \ P[\ y_m = 1\ ]\ \ \#$$

$$\min\{\ p, P[\ y(1) = 1\ ]\ \}\ \} + \min\{\ 1- p, P[\ y(0) = 1\ ]\ \}$$

Manski (1997, Proposition 7). Notice that as the fraction trained approaches one, the bounds center

around the outcome that would be observed if all recipients are assigned to training, $P[\ y(1) = 1\ ]$, while as

the fraction approaches zero the bounds center around the outcome that would be observed if all recipients

receive standard benefits, $P[\ y(0) = 1\ ]$.

The planner's problem is to determine the fraction, p, which optimizes the employment probability

under an arbitrary assignment rule m. Assume, without loss of generality, the employment rate under

mandatory training exceeds the employment rate under standard benefits. In this case, the lower bound in

(8) is maximized when all persons receive training, $p = 1$.

By maximizing the lower bound, this maximin rule serves as a benchmark for all alternative

assignment rules, $p < 1$. In this informational setting, no other assignment rule can strictly dominate

mandatory training. However, rules that lead to bounds that include the maximin outcome are

undominated; the planner cannot determine whether the outcome under mandatory training will be better

or worse than the outcome under the alternative heterogeneous assignment rule. Rules where the upper

bound in (8) is less than the mandatory training outcome are dominated. That is, an assignment rule is

dominated by mandatory training iff $p + P[\ y(0) = 1] < P[\ y(1) = 1]$. Thus, we have

Proposition 1: For Bernoulli random variables $y(1)$ and $y(0)$, let $P[\ y(1) = 1]$ and $P[\ y(0) = 1\ ]$ be known. Assume, without loss of generality, that $P[\ y(1) = 1\ ]\ \$P[\ y(0) = 1\ ]$. Let $P[\ z_m = 1] = p$. Then rules where $p < P[\ y(1) = 1] - P[\ y(0) = 1]$ are dominated. All other rules are undominated.

Proposition 1 reveals that in this asymmetrical informational setting the planner cannot determine an optimal treatment policy. Rather, the best she can do is rule out policies that train less than the revealed effect of training, that is, $P[\ y(1) = 1] - P[\ y(0) = 1]$. So, for example, Table 1 reveals that a planner using the Baltimore program should at least train 1.1 percent of the caseload, whereas a planner using the San Diego program should at least train 6.5 percent of the caseload. Any assignment rule that fails to satisfy these conditions cannot be optimal. Any assignment rule satisfying these conditions may or may not be optimal.

In this informational setting, the caseworker has the power to assign individualistic treatments but does not have the information to optimally implement a heterogeneous treatment policy. Remarkably, in contrast to the partial information setting examined above, additional covariates in asymmetric settings may degrade the quality of decision-making. While the planner can choose a treatment rule that differentiates among the members of the population, the absence of information on the outcomes may lead to sub-optimal decisions rules (Manski, 1998). By chance, the planner might replicate the optimal assignment policy in full information, namely assigning treatment such that the only persons not working are those that will be unemployed regardless. In this best case scenario, the employment probability would equal $1 - P[\ y(1) = 1 \cap y(0) = 0]$. By chance, the planner might instead assign all those who benefit from training to receive cash assistance so that the only persons working are those that would be employed regardless. In

this worst case scenario, the employment probability would equal $P[\ y(1) = 1 \cap y(0) = 1]$. Formally, the

bounds in Equation 8 imply that the employment probability may lie between

(9.) $\max\{0, P[\ y(1) = 1\ ] + P[\ y(0) = 1] - 1\ \}\# \ P[\ y_m = 1 \mid x]\ \#\ \min\{\ P[\ y(1) = 1\ ]+ P[\ y(0) = 1\ ],\ 1\ \}.$

Table 3 presents these bounds for each of the four MDRC programs. Clearly, the class of

undominated assignment policies may have an extremely wide range of consequences for the employment

probability. The upper bound occurs if the caseworker happens to implement the optimal assignment policy

under full information and the joint outcome distribution has the strongest negative correlation (see Table

2B). The full unemployment case results if the planner inadvertently makes poor treatment choice

decisions and the joint outcome distribution has the strongest negative correlation (see Table 2B). Thus,

the realized employment probability depends on two unknown factors: the heterogeneity in treatment

response and the efficacy of the assignment rule.

4.2   Illustration: Treatment Assignment with Partial Covariate Information

Rather than uniquely observing each individual and having complete freedom to assign treatments,

a planner might only observe and/or focus on a subset of covariates. Assume that the planner observes a

binary covariate X, taking the values $x = 0$ and $x = 1$, and knows the distribution of the covariate, $P[\ X =$

1]. Say, for example, that $X = 1$ for respondents with recent labor market experience and $X = 0$

otherwise. The caseworker also observes the results from a welfare-to-work experiment that reveals the

outcome under the mandated training regime, $P[y(1) = 1]$, and the mandated cash assistance regime, $P[$

$y(0) = 1]$.

As in the partial information setting, the planner cannot distinguish among persons with the same observed covariates and does not implement treatment rules that systematically differentiate among observationally equivalent persons. Thus, in this setting, there are only four different treatment rules to consider: A. everyone receives training; B. no one receives training; C. respondents with experience receive training; D. respondents without experience receive training.[9]

The bounds in Equation 8 determine which treatment rules are dominated. In particular, Proposition 1 reveals that a rule is inferior if the fraction trained is less than the treatment effect, $p < P[y(1) = 1] - P[y(0) = 1]$.

Table 4A displays the estimated employment probability under the four assignment rules. The first column displays the fraction who work in the fourth-quarter prior to assignment, $P(X=1)$, while the remaining columns display the possible outcomes under the four assignment rules.[10] There is much uncertainty about the optimal treatment assignment policy. In all cases, the fraction with and without experience exceeds the effect of training so that the only dominated rule is a policy of mandatory standard benefits. The other three rules may or may not be optimal. Depending on both the heterogeneity of response and efficacy of the assignment process, either of the two heterogeneous assignment policies might do substantially better than a mandatory training regime and either of them might do substantially worse. Still, the uncertainty is less than in the prior setting where the planner observed all covariates (see Table 3). Restricting the choice set serves to not only lower the best possible outcome but also increase the worst possible outcome.

---

[9] This is the same basic decision problem considered by Manski (1998).

[10] Although not typically reported (see Gueron and Pauly, 1991), the MDRC data actually include this employment indicator.

Since the detailed MDRC data files actually include covariates on work experience one-year prior to assignment, I can use these data to determine the optimal treatment assignment policy under partial information (Section 3.2). Table 4B displays the actual employment rate for each of the four assignments rules, and highlights the optimal rule. For the Arkansas and Baltimore programs, training has a slight negative effect on the employment probability of persons with experience. Thus, a heterogeneous assignment rule training those without experience (D) leads to employment rates that slightly exceed those under a mandatory training regime. In contrast, mandatory training rules are optimal for the San Diego and Baltimore programs.

For purposes of illustration, Table 5 displays the estimated bounds under the assumption that 5 percent of the caseload has an observed characteristic; $P(X=1) = 0.05$. As revealed in Equation 8, the bounds center around $P[\ y(0) = 1]$ as the fraction receiving standard benefits approaches one. Thus, in this setting, there is far less uncertainty about the optimal rules and outcomes. For the San Diego and Virginia programs, where the effect of training exceeds 0.05, a planner can rule out only training to those with experience (rule C). Thus, the 95 percent of recipients without experience should all be trained. For the Arkansas and Baltimore programs, mandatory cash assistance remains the only dominated rule. Still, with the employment probability under either heterogeneous treatment policy lying within a 10 point range, there is much less uncertainty about the outcomes.

5. Conclusions: What is a Caseworker to Do?

In this paper, I consider the normative question of how planners should assign treatments to optimize expected outcomes in different informational settings. With full information on the effect of treatment for each individual, the planner can maximize the employment probability. With partial information, the planner would assign uniform treatments for each observed sub-group. Finally, in the likely

case where empirical evidence on the effectiveness of different treatments is combined with additional information about the individual recipient, the planner faces a decision problem under ambiguity. In this setting, the data do not reveal the optimal decision rule. Depending on the unknown fraction of the caseload affected by treatment and the efficacy of the treatment assignment rule, the additional information observed by the planner can improve or degrade the decision making process.

What then is a decision maker to do? She might impose additional assumptions that are strong enough to reveal the optimal rule. For instance, an assumption that training never reduces than chances of being employed implies that mandatory training is an optimal decision rule. Likewise, one might assume the planners always have rational expectations: planners learn the response functions given the observed information and can implement the optimal decision rule (under partial information). The problem, of course, is that imposing assumptions that are not credible does not eliminate the ambiguity in the evaluation problem (Manski, 2000; Manski, Newman and Pepper, forthcoming).

If stronger assumptions are not imposed, the only way to resolve an indeterminate finding is to collect richer data. In principal, for example, the problem could be resolved if the randomized experiments included and/or the summary reports displayed extensive covariate data for the subjects.

In the absence of stronger assumptions or richer data, the planner is confronted with making decisions in ambiguous settings. An indeterminate finding, however, does not imply that the planner should be unwilling or unable to make decisions. A planner, for instance, might formally appeal to alternative decision criteria (Manski, 2000). One might adopt a Bayesian approach by placing a subjective distribution on the different possible outcome distributions and maximize expected welfare with respect to this distribution (see, for example, Dehejia, forthcoming). A maximin rule that selects the treatment rule with the largest lower bound employment probability would lead to a mandatory training rule (Wald, 1950). Concerns over equity may lead other rules (Berger et al., 2000). A planner using a Bayesian, maximin or

some other criteria, will ultimately decide among the various undominated decision rules. The planner

simply cannot assert that the chosen rule optimizes the employment probability.

References:

Berger, M, D. Black, and J. Smith. (2000). "Evaluating Profiling as a Means of Allocating Government Services," *University of Western Ontario*, manuscript.

Campbell, D. and J. Stanley. (1966). Experimental and Quasi-Experimental Designs for Research. Boston: Houghton Mifflin.

Dehejia, R. H. (forthcoming). "Program Evaluation as a Decision Problem," *Journal of Econometrics.*

Eberts, R.W. (1997). "The Use of Profiling to Target Services in State Welfare-to-Work Programs: An Example of Process and Implementation," *W.E. Upjohn Institute for Employment Research Working Paper.*

Frechet, M. (1951). "Sur Les Tableaux de Correlation Donte les Marges Sont Donnèes." *Annals de Universitè de Lyon A*, ser. 3, 14: 53-77.

Friedlander, D. and G. Burtless. (1995). Five Years After: The Long Term Effects of Welfare-to-Work Programs. New York: Russell Sage Foundation.

Garfinkel, I., C. Manski, and C. Michalopolous. (1992). "Micro-Experiments and Macro Effects." In C. Manski and I Garfinkel, eds., Evaluating Welfare and Training Programs. Cambridge, MA: Harvard University Press.

Greenberg, D. and M. Wiseman. (1992). "What did the OBRA Demonstrations Do?" In C. Manski and I Garfinkel, eds., Evaluating Welfare and Training Programs. Cambridge, MA: Harvard University Press.

Gueron, J.M. and E. Pauly. (1991). From Welfare to Work. New York: Russell Sage Foundation.

Hausman, J. and D. Wise, eds. (1985). Social Experimentation. University of Chicago Press.

Heckman, J. (1992). "Randomization and Social Policy Evaluation." In C. Manski and I Garfinkel, eds., Evaluating Welfare and Training Programs. Cambridge, MA: Harvard University Press.

Heckman, J., J. Smith and N. Clements. (1997). "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Proramme Impacts," *The Review of Economic Studies*, 64, 487-536..

Hotz, V.J., G.W. Imbens, and J.A. Klerman. (2000). "The Long-Term Gains from GAIN: A Re-Analysis of the Impacts of the California GAIN Program," NBER Working Paper, 8007.

Hotz, V.J., G.W. Imbens, and J.H. Mortimer. (1999). "Predicting the Efficacy of Future Training Programs Using Past Experiances," NBER Working Paper, T0238.

Manski, C. (1997). "The Mixing Problem in Program Evaluations." *The Review of Economic Studies*, 64, 537-553.

Manski, C. (1998). Treatment Choice in Heterogeneous Populations Using Experiments Without Covariate Data. in G. Cooper and S. Moral, Eds., Uncertainty in Artificial Intelligence, Proceedings of the Fourteenth Conference, Morgan Kaufmann, San Francisco, 379-385.

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

Manski, C. and I. Garfinkel, eds. (1992). Evaluating Welfare and Training Programs. Cambridge, MA: Harvard University Press.

Manski, C., J. Newman, J. Pepper (1998). "Using Performance Standards to Evaluate Social Programs with Incomplete Outcome Data," Thomas Jefferson Center Discussion Paper 312.

Michalopoulos, C. and C. Schwartz (2000). "What Works Best for Whom: Impacts of 20 Welfare-to-Work Programs by Subgroup," in The National Evaluation of Welfare_to_Work Strategies, Manpower Demonstration Research Corporation.

Moffitt, R. (1992). "Evaluation Methods for Program Entry Effects." In C. Manski and I Garfinkel, eds., Evaluating Welfare and Training Programs. Cambridge, MA: Harvard University Press.

Pavetti, LaDonna, Krista Olson, Demetra Nightingale, Amy_Ellen Duke, and Julie Isaacs. (1997). *Welfare_to_Work Options for Families Facing Personal and Family Challenges: Rationale and Program Strategies,* Urban Institute Press, Washington, D.C.

Pepper, J.V. (2001) "Using Experiments to Evaluate Performance Standards: What Do Welfare-to-Work Demonstrations Reveal to Welfare Reformers?" Thomas Jefferson Center Discussion Paper, University of Virginia.

Wald, A. (1950), Statistical Decision Functions, New York: Wiley.

Wiseman, M. (1991). Ed. "Research and Policy: A Symposium on the Family Support Act of 1988," *The Journal of Policy Analysis and Management, 10(4), pp.588-666.*

Table 1: Estimated Probability of Employment Eight Quarters After Treatment

	Control Group: P[y(0) = 1]	Treatment Group: P[y(1) = 1]
Arkansas	20.5% (0.175, 0.230)	23.8% (0.213, 0.265)
Baltimore	37.7% (0.255, 0.392)	38.8% (0.373, 0.412)
San Diego	28.5% (0.265, 0.302)	35.0% (0.331, 0.370)
Virginia	33.9% (0.315, 0.362)	39.0% (0.373, 0.408)

Note: Bootstrapped 90 percent confidence intervals are in parentheses below the estimates.

Table 2A: A Joint Distribution Consistent with the Labor Force Participation Probabilities for the Virginia Employment Services Program: Strongest Positive Association Between the Outcomes

Treatment

	Training	
Outcomes	y(1) = 0	y(1) = 1
Standard Benefit y(0) = 0	61.0%	5.1%
y(0) = 1	0.0%	33.9%

Table 2B: A Joint Distribution Consistent with the Labor Force Participation Probabilities for the Virginia Employment Services Program: Strongest Negative Association Between the Outcomes

Treatment

	Training	
Outcomes	y(1) = 0	y(1) = 1
Standard Benefit y(0) = 0	27.1%	39.0%
y(0) = 1	33.9%	0.0%

Table 3: Estimated No Assumption Bounds on the Probability of Employment Eight Quarters After Treatment Under Assignment Policy m, P[$y_m = 1$]

	Lower Bound	Upper Bound
Arkansas	0.0%	44.2% (0.474)
Baltimore	0.0%	76.5% (0.796)
San Diego	0.0%	63.5% (0.661)
Virginia	0.0%	72.9% (0. 759)

Note: The 0.95 quantiles of bootstrapped upper bound are in parentheses below the estimates.

Table 4A: No Assumption Bounds on the Probability of Employment Eight
              Quarters After Treatment Under Assignment Policies A-D.

[X = 1 if respondent worked in 4th quarter prior to assignment; 0 otherwise]

State [P(X=1)]	A. Z = 0 for all X	B. Z = 1 for all X	C. Z = 1 for X = 1, Z =0 for X = 0	D. Z = 1 for X = 0, Z =0 for X = 1
Arkansas [15.0%]	20.5%	23.8%	[5.5, 35.5]	[8.8, 38.8]
Baltimore [29.7%]	37.7%	38.8%	[8, 67.4]	[9.1, 68.5]
San Diego [24.7%]	28.5%	35.0%	[3.8, 53.2]	[10.3, 59.7]
Virginia [20.0%]	33.9%	39.0%	[13.9, 53.9]	[19.0, 59.0]

Table 4B: Probability of Employment Eight Quarters After Treatment Under
              Assignment Policies A-D

[X = 1 if respondent worked in 4th quarter prior to assignment; 0 otherwise]

State [P(X = 1)]	A. Z = 0 for all X	B. Z = 1 for all X	C. Z = 1 for X = 1, Z =0 for X = 0	D. Z = 1 for X = 0, Z =0 for X = 1
Arkansas [15.0%]	20.5%	23.8%	19.5%	24.5%
Baltimore [29.7%]	37.7%	38.8%	37.5%	39.0%
San Diego [24.7%]	28.5%	35.0%	29.8%	33.7%
Virginia [20.0%]	33.9%	39.0%	34.9%	37.8%

Note: Highlighted cells indicate optimal employment probability under partial information.

Table 5:   Hypothetical No Assumption Bounds on the Probability of Employment Eight
                Quarters After Treatment Assuming 5% of the Caseload has $X = 1$.

State	A. $Z = 0$ for all X	B. $Z = 1$ for all X	C. $Z = 1$ for $X = 1$, $Z = 0$ for $X = 0$	D. $Z = 1$ for $X = 0$, $Z = 0$ for $X = 1$
Arkansas	20.5%	23.8%	[15.5, 25.5]	[18.8, 28.8]
Baltimore	37.7%	38.8%	[32,7, 42.7]	[33.8, 43.8]
San Diego	28.5%	35.0%	[23.5, 33.5]	[30.0, 40.0]
Virginia	33.9%	39.0%	[28.9, 38.9]	[34.0, 44.0]