1

Utilitarian Treatment of Heterogeneous Populations

1.1 Studying Treatment Response to Inform Treatment Choice

An important practical objective of empirical studies of treatment response is to provide decision makers with information useful in choosing treatments. Often the decision maker is a planner who must choose treatments for a heterogeneous population. In the utilitarian tradition of welfare economics, the planner may want to choose treatments whose outcomes maximize the welfare of this population.

Consider, for example, a physician choosing medical treatments for a population of patients. The physician may observe each patient's demographic attributes, medical history, and the results of diagnostic tests. He may then choose a treatment rule that makes treatment a function of these covariates. If the physician acts on behalf of his patients, the outcome of interest may measure his patients' health status, and welfare may measure health status minus the cost of treatment, in comparable units.

Or consider a judge choosing sentences for a population of convicted offenders. The judge may observe each

offender's past criminal record, demeanor in court, and other attributes. Subject to legislated sentencing guidelines, she may consider these covariates when choosing sentences. If the judge acts on behalf of society, the outcome of interest may measure recidivism, and social welfare may decrease with recidivism and the cost of carrying out a sentence.

Empirical studies of treatment response are useful to physicians, judges, and other planners, to the extent that they reveal how outcomes vary with treatments and observable covariates. There are fundamental, practical, and volitional reasons why studies of treatment response do not provide all the information that planners would like to have. An obvious but fundamental reason is that outcomes can be observed only for treatments that have already been received. Hence, a planner cannot know prospectively how persons will respond to alternative treatments. Moreover, observation of treatment response in a study population that has previously been treated can at most reveal the outcomes that these persons experienced under the treatments that they actually received. The counterfactual outcomes that members of the study population would have experienced under other treatments are logically unobservable.

Practical problems of data collection enlarge the gap between the information that planners would like to have and the evidence that empirical studies of treatment response provide. The mundane fact that data collection is costly may constrain researchers to study small samples of survey respondents or experimental subjects. Planners may want to learn long-term outcomes of treatments, whereas studies of treatment response may only measure short-term outcomes. Survey respondents may refuse to answer or may respond inaccurately to questions about the treatments that they have received and

the outcomes that they have experienced. Experimental subjects may not comply with assigned treatments or may drop out of trials before their outcomes are measured.

The volitional reasons are the researchers' choices that limit the usefulness of their work to planners. Much empirical research on treatment response tests hypotheses that bear only a remote relationship to the treatment choice problems that planners face. Researchers frequently study populations that differ substantially from those that planners treat. Researchers rarely report how treatment response varies with the covariates that planners observe. Research findings often rest on untenable assumptions. See Section 1.3 for further discussion.

To learn how studies of treatment response can be most helpful to planners, I find it productive to eliminate the distinction between researcher and decision maker. That is, I maintain the perspective of a planner who can perform his own research in the service of treatment choice. The planner observes a study population, combines this empirical evidence on treatment response with assumptions that he deems credible, and then chooses treatments for the population of interest. This monograph examines how a planner may reasonably go about this task.

1.2 The Planning Problem

To move beyond generalities, it is necessary to pose the treatment choice problem that I shall presume a planner faces. As in Manski (2000, 2001, 2002, 2003, Chapter 7, 2004, 2005), I assume that a planner must choose treatments for the members of a heterogeneous population. Each member of the population has a response function that maps treatments into an outcome of interest.

The planner may observe some covariates that differentiate members of the population. The observed covariates determine the set of treatment rules that are feasible to implement. These are functions that map the observed covariates into a treatment allocation.

I assume that the planner wants to choose a treatment rule that maximizes population mean welfare; that is, he wants to maximize a utilitarian social welfare function. This problem has a simple solution—the optimal treatment rule assigns to each member of the population a treatment that maximizes mean welfare conditional on the person's observed covariates. However, the planner does not have all the knowledge of treatment response needed to implement the optimal rule. What the planner does have is the ability to observe a study population in which treatments have already been selected and outcomes have been realized. The planner's problem is to use the available empirical evidence and credible assumptions to make treatment choices.

1.2.1 The Choice Set

To formalize the planning problem, suppose that there is a finite set T of mutually exclusive and exhaustive treatments. Each member j of the treatment population, denoted J^*, has a response function $y_j(\cdot) : T \to Y$ mapping treatments $t \in T$ into outcomes $y_j(t) \in Y$. The planner is concerned with the distribution of outcomes across the population, not with the outcomes of particular persons. Hence, it is convenient to make the population a probability space (J^*, Ω, P). Then the probability distribution $P[y(\cdot)]$ of the random function $y(\cdot) : T \to Y$ describes treatment response across the population.

A planner must choose a treatment rule assigning a treatment to each member of J^*. A fully specified treatment rule is a function $\tau(\cdot) : J \to T$ that assigns a

treatment to each person. Person j's outcome under rule $\tau(\cdot)$ is $y_j[\tau(j)]$. I assume that treatment is individualistic; that is, a person's outcome may depend on the treatment he is assigned, but not on the treatments assigned to others.

The planner observes certain covariates $x_j \in X$ for each member of the population; thus, $x : J \to X$ is the random variable mapping persons into their observable covariates. To simplify analysis, I suppose that the covariate space X is finite and that $P(x = \xi) > 0$, $\forall \xi \in X$. The planner can differentiate persons with different observed covariates, but cannot distinguish among persons with the same observed covariates. Hence, a feasible treatment rule is a function that assigns all persons with the same observed covariates to one treatment or, more generally, a function that randomly allocates such persons across the different treatments.

Formally, let Z denote the space of functions that map $T \times X$ into the unit interval and that satisfy the adding-up conditions: $z(\cdot\,,\cdot) \in Z \;\Rightarrow\; \sum_{t \in T} z(t, \xi) = 1$, $\forall \xi \in X$. Then the feasible treatment rules are the elements of Z. An important subclass of Z are the *singleton* rules that assign all persons with the same observed covariates to one treatment; that is, $z(\cdot\,,\cdot)$ is a singleton rule if, for each $\xi \in X$, $z(t, \xi) = 1$ for some $t \in T$ and $z(s, \xi) = 0$ for all $s \neq t$. Nonsingleton rules randomly allocate persons with covariates ξ across multiple treatments, with assignment shares $[z(t, \xi),\, t \in T]$. This definition of nonsingleton rules does not specify which persons with covariates x receive each treatment, only the assignment shares. Designation of the particular persons receiving each treatment is immaterial because assignment is random and the planner's objective is to maximize population mean welfare.

In some settings, a planner may not be permitted to use certain covariates (say race or gender) to assign treatments. If so, the present description of the choice set remains accurate if x is defined to be the covariates that the planner is permitted to use, rather than the full vector of covariates that the planner observes.

1.2.2 The Objective Function and the Optimal Treatment Rule

The planner wants to choose a feasible treatment rule that maximizes population mean welfare. The welfare from assigning treatment t to person j is

$$u_j(t) \equiv u[y_j(t), t, x_j],$$

where $u(\cdot\,,\cdot\,,\cdot) : Y \times T \times X \to R$ is the welfare function. The planner knows the form of $u(\cdot\,,\cdot\,,\cdot)$ and observes x_j. However, he does not observe the potential treatment outcomes $[y_j(t),\ t \in T]$.

Welfare may, for example, have the additive "benefit–cost" form

$$u[y(t), t, x] = y(t) + c(t, x),$$

where $c(t, x)$ is the real-valued cost of assigning treatment t to a person with covariates x, and $y(t)$ is the real-valued benefit of this treatment. In the case of a physician, $y_j(t)$ may measure the health status of patient j following receipt of treatment t, and $c(t, x_j)$ may be the (negative-valued) cost of treatment. At the time of treatment choice, the physician may know the costs of alternative medical treatments but not their health outcomes. Similarly, in the case of a judge, $y_j(t)$ may measure the criminal behavior of offender j following receipt of sentence t, and $c(t, x_j)$ may be the cost of carrying out the sentence. Again, the judge may know the costs of alternative sentences but not their criminality outcomes.

For each feasible treatment rule z, the population mean welfare that would be realized if the planner were to choose rule z is

$$U(z,P) \equiv \sum_{\xi \in X} P(x=\xi) \sum_{t \in T} z(t,\xi) E[u(t) \mid x=\xi]. \quad (1.1)$$

The planner wants to solve the problem

$$\max_{z \in Z} U(z,P). \quad (1.2)$$

Let S denote the unit simplex in $R^{|T|}$. The maximum in (1.2) is achieved if, for each $\xi \in X$, the planner chooses the treatment allocation $z(\cdot,\xi)$ to solve the problem

$$\max_{z(\cdot,\xi) \in S} \sum_{t \in T} z(t,\xi) E[u(t) \mid x=\xi]. \quad (1.3)$$

The maximum in (1.3) is achieved by a singleton rule that allocates all persons with covariates ξ to a treatment that solves the problem

$$\max_{t \in T} E[u(t) \mid x=\xi]. \quad (1.4)$$

There is a unique optimal rule if problem (1.4) has a unique solution for every $\xi \in X$. There are multiple optimal rules if (1.4) has multiple solutions for some $\xi \in X$. In the latter case, all rules that randomly allocate persons with the same covariates among their optimal treatments are optimal. In any event, the population mean welfare achieved by an optimal rule is

$$U^*(P) \equiv \sum_{\xi \in X} P(x=\xi) \Big\{ \max_{t \in T} E[u(t) \mid x=\xi] \Big\}. \quad (1.5)$$

1.2.3 The Value of Covariate Information

The population welfare achievable by an optimal treatment rule depends on the observed covariates. The set

of feasible treatment rules grows as more covariates that differentiate members of the population are observed. Hence, the optimal welfare cannot fall, and may rise, as more covariates are observed.

In particular, compare $U^*(P)$ with the welfare achievable when no covariates are observed. In that case, the optimal feasible treatment rule yields welfare $U^0(P) \equiv \max_{t \in T} E[u(t)]$. Hence, the value of observing covariates x is the nonnegative quantity $U^*(P) - U^0(P)$. If observation of x is costly and welfare is measured in commensurate units, $U^*(P) - U^0(P)$ is the amount that the planner should be willing to pay to observe x.

The value of observing x is positive whenever optimal treatments vary with x. It is zero if there exists a common optimal treatment, that is, a t^* that solves (1.4) for all values of x. Thus, observable heterogeneity in treatment response is relevant to treatment choice if and only if optimal treatments vary with the observed covariates.

1.2.4 Partial Knowledge of Treatment Response

A planner who knows the treatment-response distributions $P[y(t) \mid x]$, $t \in T$, can choose an optimal treatment rule. Economic theorists studying social choice have long assumed that planners know the response distributions and have sought to characterize the resulting optimal treatment rules. See, for example, Mirrlees (1971) on optimal income taxation, Polinsky and Shavell (1979) on optimal fines, and Shavell and Weiss (1979) on optimal unemployment benefits.

My concern is a planner who does not know the response distributions but who can observe a study population in which treatments have been selected and outcomes realized. I consider how such a planner may use the available empirical evidence and credible assumptions to choose treatments reasonably.

If observation of a study population is to yield information useful in treatment choice, a planner must be able to extrapolate from the study population to the treatment population. With this in mind, I assume that the study population, denoted J, is identical in distribution to the treatment population J^*. Thus, J is a probability space whose probability measure P is the same as that of J^*. The only difference between J and J^* is that some *status quo treatment rule* has already been applied and outcomes experienced in the former population, whereas a treatment rule is yet to be chosen in the latter.

It is often optimistic to suppose that a planner can observe a study population that is distributionally identical to the treatment population. Nevertheless, treatment choice is a formidable task even in this benign setting. I focus on two ubiquitous problems. First, outcomes are observable only for the treatments that members of the study population received under the status quo treatment rule; the outcomes of counterfactual treatments are necessarily unobservable (Chapter 2). Second, outcomes may be observed only for a random sample drawn from the study population, in which case the planner must perform statistical inference from this sample to the population (Chapters 3 and 4).

While the chapters ahead differ in their specification of the available empirical evidence, they all ask how a planner with partial knowledge of treatment response may reasonably make treatment choices. Decision theorists have studied various criteria for decision-making with partial information, but no consensus prescription has emerged and it may be that none will ever emerge. I apply the Wald (1950) development of statistical decision theory and, within Wald's framework, I focus most attention on the minimax-regret criterion proposed by Savage (1951). The minimax-regret criterion is appealing in

principle and yields treatment rules that seem quite reasonable in the applications that I have studied. Nevertheless, I do not assert that a planner with partial knowledge of treatment response should necessarily make treatment choices in this way. I discuss other criteria as well.

1.2.5 Nonseparable Planning Problems

An important property of the optimal treatment rule (1.4) is that it is separable across covariate values. That is, the optimal rule for persons with covariates ξ is invariant with respect to the situations of persons with other values of x. This separability, which greatly simplifies analysis of treatment choice, rests on three assumptions: the planner is utilitarian, the set of feasible treatments is rectangular, and treatment is individualistic. To close this section, I call attention to important planning problems that do not satisfy separability and that, consequently, are not studied here.

Nonutilitarian Objective Functions

The idea of a planner with a utilitarian social welfare function carries forward a long tradition in public economics. Nevertheless, I would not assert that the utilitarian perspective is necessarily realistic in all settings. Analysis of treatment choice from nonutilitarian perspectives would be welcome, but is beyond the scope of this monograph. Nonutilitarian objective functions generically make social welfare depend on the relative positions of different members of the population. A consequence is that optimal nonutilitarian treatment rules are not generally separable across covariate values.

Nonrectangular Sets of Feasible Treatment Rules

I assume that the set of feasible treatment rules is rectangular; that is, the set of treatments that are jointly

feasible to assign across the population are the Cartesian products of the treatments that are feasible to assign to each member of the population. Budgetary or technological constraints may render a treatment set nonrectangular, in which case optimal treatment rules generally are not separable.

Suppose, for example, that there is a budgetary upper bound on the total cost of treating the population. Then the feasible treatment rules satisfy the inequality

$$\sum_{\xi \in X} P(x = \xi) \sum_{t \in T} z(t, \xi) c(t, \xi) \leqslant K,$$

where K is the budget and $c(t, \xi)$ is the cost of assigning treatment t to a person with covariates ξ. The budget constraint binds if the total cost of the optimal rules determined in (1.4) exceeds K. If so, application of (1.4) is infeasible.

Social Interactions

Individualistic treatment means that each person's outcome depends only on the treatment that he is assigned, not on the treatments assigned to others. Social interactions occur when personal outcomes do depend on the treatments assigned to others. The present analysis does not cover problems of treatment choice with social interactions.

1.3 Practices that Limit the Usefulness of Research on Treatment Response

In Section 1.1, I stated four ways in which choices made by researchers studying treatment response often limit the usefulness of their work to planners. Having posed the planning problem, I can now elaborate.

1.3.1 Hypothesis Testing

Empirical research on treatment response has been strongly influenced by the classical theory of hypothesis testing, especially by the idea of testing the null hypothesis of zero average treatment effect; that is, equality of $E[y(t)]$ and $E[y(t')]$ for specified treatments t and t'. This null hypothesis is prominent in experimental design, where researchers use norms for statistical power to choose sample sizes. Research findings may go unreported or may be deemed to be "insignificant" if they do not meet test-based criteria for statistical precision.

Hypothesis testing has been particularly influential in medical research using randomized clinical trials. A standard reference on the design and analysis of clinical trials gives this prescription for choice of sample size in a trial (Meinert 1986, p. 74): "With a sample size calculation, the investigator sets out to determine the number of patients required to detect a designated treatment difference with specified levels of type I and type II error protection." Many medical researchers (e.g. Halpern, Karlawish, and Berlin 2002) consider it unethical to conduct "underpowered" trials, in which the sample size does not make the probability of a type II error sufficiently small, given a specified value for the probability of a type I error. Testing the hypothesis of zero average treatment effect is institutionalized in the US Food and Drug Administration (FDA) drug approval process, which calls for comparison of a new treatment under study ($t = b$) with a placebo or an approved treatment ($t = a$). FDA approval of the new treatment normally requires rejection of the null hypothesis of zero average treatment effect $\{H_0 : E[y(b)] = E[y(a)]\}$ in two independent clinical trials (Fisher and Moyé 1999).

Hypothesis testing is remote from treatment choice. The classical practice of handling the null and alternative

hypotheses asymmetrically, fixing the probability of a type I error, and seeking to minimize the probability of a type II error, makes no sense from the perspective of treatment choice. Moreover, error probabilities at most measure the chance of choosing a suboptimal rule; they do not measure the damage resulting from a suboptimal choice. For these and other reasons, research reporting the results of hypothesis tests yields little information of use to a planner.

1.3.2 The Study Population and the Treatment Population

Much research on treatment response downplays the importance of correspondence between the study population and the population to be treated. Donald Campbell argued that studies of treatment effects should be judged primarily by their *internal validity* and only secondarily by their *external validity* (e.g. Campbell and Stanley 1963; Campbell 1984). By internal validity, Campbell meant the credibility of findings within the study population, whatever it may be. By external validity, he meant the credibility of extrapolating findings from the study population to another population of interest.

Rosenbaum (1999, p. 263) recommends that observational studies of human subjects aim to approximate the conditions of laboratory experiments:

> In a well-conducted laboratory experiment one of the rarest of things happens: The effects caused by treatments are seen with clarity. Observational studies of the effects of treatments on human populations lack this level of control but the goal is the same. Broad theories are examined in narrow, focused, controlled circumstances.

Rosenbaum, like Campbell, downplays the importance of having the study population be similar to the population of interest, writing (Rosenbaum 1999, p. 259): "Studies of samples that are representative of populations may be quite useful in describing those populations, but may be ill-suited to inferences about treatment effects."

In accord with Campbell and Rosenbaum, many researchers concerned with the evaluation of social programs analyze treatment response in easy-to-study populations that differ fundamentally from the populations that planners must treat. A common practice has been to report the "effect of treatment on the treated," where "the treated" are the members of a study population who actually received a specified treatment (see, for example, Bloom 1984; Angrist 1990; Gueron and Pauly 1991; Dubin and Rivers 1993). Attempting to cope with the problem of noncompliance in randomized experiments, Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996) recommend that treatment effects be reported for the subpopulation of "compliers," these being persons who would comply with their designated experimental treatments whatever they might be.

From the perspective of treatment choice, analysis of treatment response in an easy-to-study population is sensible if treatment response is homogeneous. Then planners can be confident that research findings can be extrapolated to the populations they must treat. In human populations, however, homogeneity of treatment response may be the exception rather than the rule. Whether the context be medical, educational or social, it is reasonable to think that persons vary in their response to treatment. To the degree that treatment response is heterogeneous, a planner must take care when extrapolating research findings from a study population to a

treatment population, as optimal treatments in the two may differ. Hence, correspondence between the study population and the treatment population assumes considerable importance.

1.3.3 Reporting Observable Variation in Treatment Response

To inform treatment choice, research on treatment response should aim to learn how treatment response varies with covariates that planners can observe. If all persons respond to treatment in the same manner, then it is best to treat all persons uniformly. However, if treatment response varies with observable covariates, then planners can do better by implementing treatment rules in which treatment varies appropriately with these covariates. For example, judges may be able to lower recidivism among criminal offenders by sentencing some offenders to prison and others to probation. Social workers may be able to increase the life-cycle earnings of welfare recipients by placing some in job training and others in basic skills classes. In these and many other cases, the key to success is determining which persons should receive which treatments.

Nevertheless, the prevalent research practice has been to report treatment response in the population as a whole or within broad subpopulations, rather than conditional on the covariates that planners may observe. An article reviewing evaluations of training programs for the economically disadvantaged exemplifies the problem (Friedlander, Greenberg, and Robins 1997). Throughout their extended discussion of inferential problems that arise in evaluating training programs, the authors assume that all potential trainees respond uniformly to treatment. Their subsequent discussion of empirical findings presents separate estimates of treatment effects only for

the very broad demographic groups of adult men, adult women, and youth. The authors do not, even in their concluding "Agenda for Future Evaluations," ask how response to training may vary with schooling, work experience, or other covariates that the administrators of training programs may observe.

The Friedlander et al. article faithfully portrays the literature that it reviews, so I do not intend to single it out for criticism. Similar inattention to observable heterogeneity in treatment response is prevalent in other major literatures. Consider the vast body of medical research through clinical trials. Physicians commonly have much information—medical histories, diagnostic test findings, and demographic attributes—about the patients they treat. Yet the medical journal articles that report on clinical trials typically present estimates of treatment effects aggregated to broad demographic groups.

An article on a clinical trial comparing alternative psychosocial treatments for cocaine dependence provides an apt illustration. Crits-Christoph et al. (1999) report on a National Institute on Drug Abuse study randomly placing 487 cocaine-dependent patients in one of four treatment groups, each designated treatment combining group drug counseling (GDC) with another form of therapy. In some respects, the article is attentive to the possibility of heterogeneity in treatment response. The authors call attention to the fact that previous findings on the relative merits of psychotherapy and drug counseling for treatment of opiate-dependent patients do not hold up in the context of cocaine dependence. They provide much descriptive information on the characteristics of the subjects, including measures of race, sex, age, education, employment status, type and severity of drug use, psychiatric state, and personality. They test hypotheses that treatment effects do not vary with patient

psychiatric state or personality. However, the article does not report outcomes conditional on any of the patient covariates observed by the researchers. Indeed, its formal "Conclusion" section makes no reference to the possibility that treatment response might vary with observable covariates, stating simply (Crits-Christoph et al. 1999, p. 493): "Compared with professional psychotherapy, a manual-guided combination of intensive individual drug counseling and GDC has promise for the treatment of cocaine dependence."

Why have researchers done so little to analyze observable heterogeneity in treatment response? Some researchers may firmly believe that, in their study settings, treatment response is homogeneous across the population. If so, then covariate information has no value. However, it is difficult to imagine many cases in which it is credible to assume homogeneous treatment response, without empirical investigation.

I conjecture that the primary reason why researchers neglect to analyze observable heterogeneity in treatment response is concern for the statistical precision of their estimates of treatment effects. (I can only conjecture this because researchers rarely state explicit reasons for what they do not do.) As discussed above, conventional ideas about what constitutes adequate statistical precision for an empirical finding to be of interest have been strongly influenced by the theory of hypothesis testing. Conditioning on covariates generally reduces the statistical precision of estimates of treatment effects, often to the point where findings are "statistically insignificant" by conventional criteria. Hence, researchers often restrict their attention to estimation of population-wide average treatment effects, or effects within major subpopulations such as adult women or youth.

Chapter 3 will show that, if researchers wish to inform treatment choice, they should not view statistical insignificance as a reason to refrain from studying observable heterogeneity in treatment response. A planner must be concerned with the quantitative variation of outcomes with treatments and covariates. Hypothesis tests simply do not address this question.

1.3.4 Untenable Assumptions

Powerful incentives often influence researchers studying treatment response to maintain assumptions far stronger than they can persuasively defend, in order to draw strong conclusions. The scientific community tends to reward those who produce unambiguous findings. The public, impatient for solutions to its pressing concerns, tends to reward those who offer simple analyses leading to unequivocal policy recommendations.

Especially commonplace has been research using untenable assumptions to infer the outcomes that members of study populations would have experienced under counterfactual treatments. Heckman and Robb (1985) provide a compendium of assumptions that logically suffice to infer counterfactual outcomes from data on observed outcomes. However, researchers applying these assumptions are rarely able to provide much substantive justification for them.

Research findings based on untenable assumptions are not much use to a planner facing the treatment choice problem of Section 1.2. The objective of such a planner is to maximize *actual* social welfare, not the social welfare that would prevail if untenable assumptions were to hold.