Econometrics for Credible Policy Choice under Uncertainty

Charles F. Manski

Northwestern University

Final Version: September 8, 2025

Forthcoming in *Journal of Political Economy Micro*

Abstract

The importance of public policy creates a concern that policy analysis be credible and relevant. There has been concern with the credibility of analysis using econometric models that are point-identified under strong assumptions. A response of microeconomists has been to use quasi-experimental data to estimate treatment effects in treatment-determined study populations. Such research lacks policy relevance. I have studied what research can credibly learn about policy impacts in populations of welfare-economic interest. The findings are partial identification of treatment effects. I use decision theory to address policy choice under uncertainty, focusing on the minimax-regret criterion. This paper explains.

Acknowledgements: Prepared for the JPE Micro conference at the Department of Economics, University of Chicago, April 18-19, 2025. I am grateful for the helpful comments of Stéphane Bonhomme. "Edited by Alex Torgovitsky.

1. Credibility and Policy Relevance in Econometrics

From its beginnings, econometrics has sought to develop methodology for empirical research that aims to inform choice of public policy. In the early 20th century, econometricians developed deterministic structural models of economies and attempted to explain deviations between model predictions and observed data solely as consequences of measurement error. In a fundamental conceptual breakthrough, Haavelmo (1944) developed models in which outcomes are determined jointly by observed and unobserved factors. He proposed a probabilistic structure that characterizes uncertainty by placing restrictions on the probability distribution of unobserved quantities conditional on specified observed quantities. This probabilistic structure has subsequently been used throughout econometric research.

In the mid-20th century, econometricians mainly studied the use of observational data to estimate linear structural models in which mean-independence assumptions are the primary probabilistic restriction. Concerned with the restrictiveness of these models, econometricians from the 1970s onward have studied a wide spectrum of parametric, semiparametric and nonparametric structural models. This research has weakened the assumptions of earlier structural models, but it has mainly maintained assumptions strong enough to point-identify the models under study.

From its beginnings through today, structural econometric research has admirably sought to be policy relevant. However, the use of untenably strong assumptions to estimate point-identified models has sacrificed credibility in order to draw strong conclusions. Structural researchers use point estimates of policy impacts to determine ostensibly optimal policies. However, realistic optimization is elusive if estimates of policy impacts are not credible. Hence, policy analysis using structural models has been criticized.

Many econometricians have written on these matters. I have addressed them in Manski (1995, 2003, 2007, 2021) and elsewhere. I have emphasized the tension between the strength of assumptions and the credibility of inference, described in Manski (2003, p. 1) as: "*The Law of Decreasing Credibility*: The credibility of inference decreases with the strength of the assumptions maintained." This 'Law' implies that

analysts face a dilemma as they decide what assumptions to maintain. Stronger assumptions yield conclusions that are more powerful but less credible.

The Law of Decreasing Credibility expresses a dilemma in research using given data. The problem may be mitigated, albeit not entirely eliminated, if it is possible to improve credibility by enriching the available data and knowledge of the process generating it. This broad idea has stimulated considerable interest in the conduct and analysis of experiments with random assignment of treatments.

The classical argument for randomized experimentation, often attributed to Fisher (1935), can be phrased as follows:

> Let random samples of persons be drawn and formed into treatment groups. Let all members of a treatment group be assigned the same treatment and assume that treatment response is individualistic. Suppose that each subject complies with the assigned treatment. Then the distribution of outcomes experienced by the members of a treatment group should be the same, up to random sampling error, as would be observed under a program in which the treatment in question is received by all members of the population.

Design-Based Inference

Based on this argument, some researchers have argued that so-called *design-based inference* of data from randomized experiments or quasi-experiments eliminates any need for structural econometric research and maximizes the credibility of study of treatment response; see, for example, Bassi and Ashenfelter (1986) and LaLonde (1986). Angrist and Pischke (2010) used the term *credibility revolution* to advocate for design-based analysis of experimental data and against structural econometrics.

An early example of design-based inference is Fisher's permutation test, which is based on randomization of treatment assignment in a given sample of subjects, without reference to the population from which subjects were drawn. Modern advocacy of design-based inference has roots in the work of Donald Campbell and collaborators; e.g., Campbell and Stanley (1963). Campbell distinguished between the internal and external validity of studies of treatment response. A study is said to have *internal validity* if it has credible findings for the study population, whatever it may be. It has *external validity* if an invariance assumption permits

credible extrapolation to a population of substantive interest.

Campbell argued that studies should be judged primarily by their internal validity and secondarily by their external validity. This perspective has been used to argue for the primacy of experimental research over observational studies, whatever the study population may be. The appeal of an ideal randomized experiment is its internal validity.[1] Such experiments have no inherent advantage in terms of external validity.

Researchers studying experimental data commonly fail to seriously consider external validity. Analyses focus on the outcomes measured with the treatments assigned in the study population. Researchers may offer verbal conjectures on external validity in the discussion sections of their papers, but they do not assess external validity quantitatively. Thus, policy relevance is sacrificed.[2]

---

[1] In practice, randomized experiments may deviate from Fisher's ideal in ways that compromise the internal validity valued by Campbell. Subjects may selectively not comply with assigned treatments, or they may leave the experiment before their outcomes are observed. Fisher's argument for randomized treatment assignment also breaks down if treatment response is not individualistic. Social interactions may make a full-scale program yield different outcomes than a small-scale experiment.

Deviation of actual experiments from the ideal was a prominent concern of labor economists analyzing data from the income-maintenance experiments of the 1970s, who sought to use the experimental data to estimate structural econometric models of labor supply (e.g., Cain, 1986; Hausman and Wise, 1979). Many of the contributors to the volumes edited by Hausman and Wise (1985) and Manski and Garfinkel (1992) found that Fisher's argument for random assignment was not credible when applied to experimental evaluations of income maintenance, welfare, training, and other social programs performed in the 1970s and 1980s.

[2] This critique does not imply that an experiment has no policy relevance if it is performed on a study population that differs from the population of policy concern. It may have partial relevance if the former population has an informative known relationship to the latter. Manski (1996, Section IV) studied the simple

Campbell's doctrine of the primacy of internal validity has been extended from randomized trials to observational studies. When considering the design and analysis of observational studies of treatment response, Campbell and his collaborators recommended that researchers aim to emulate as closely as possible the conditions of a randomized experiment, even if this requires focus on a study population that differs materially from the population of interest. This has led to development of various methodologies for research on so-called *quasi-experiments*.

Thistlethwaite and Campbell (1960) introduced *regression-discontinuity* analysis of policies that use observable institutional rules to determine treatment assignment. At most, such analysis credibly point-identifies treatment response in the sub-population of persons who are close to the threshold determining treatment assignment. Economists have performed *difference-in-difference* analyses of temporal changes in policy (e.g., Card and Krueger, 1994). Such analysis maintains a questionable *parallel trends* assumption and, even when this assumption is credible, only identifies average treatment effects in treatment units where treatment assignment changed over time. As with analysis of randomized experiments, these approaches to study of observational data have been promoted for their internal validity, leaving external validity as a secondary concern. Again, policy relevance has been sacrificed.

Since the mid-1990s, the Campbell perspective has been championed by applied microeconomists who advocate study of a *local average treatment effect* (LATE). This is defined as the average treatment effect within the sub-population of so-called *compliers*, these being persons whose received treatments would be modified by hypothetically altering the value of an instrumental variable; see Imbens and Angrist (1994) and Angrist, Imbens, and Rubin (1996). Local average treatment effects are not policy-relevant estimands;

---

case of experimentation on a sub-population of the population of policy concern, the sub-population comprising a known fraction of the full population. Then an ideal experiment point-identifies the distribution of treatment response in the sub-population. Even in the absence of any external validity, the experiment partially identifies the distribution on the full population because the full population is a probability mixture of the study population and its complement.

see Manski (1996, 2007), Deaton (2009), and Heckman and Urzua (2009). Their study has been motivated by the fact that they are point-identified given certain assumptions that are sometimes thought credible.

Considering the practice of empirical microeconomic policy analysis in the mid-1990s, I argued in Manski (1996) that serious credibility problems afflict point-identified structural econometric research. I expressed strong concern that the LATE concept lacks policy relevance. I regarded LATE to be only an algebraic curiosity, a modest extension of the Bloom (1984) analysis of *the effect of treatment on the treated.* Writing in 1996, I did not foresee that applied microeconomists would embrace LATE, making it central to a large body of research.

As I see the matter now, research on LATE has become a prime example of the common reluctance of researchers to face up to uncertainty in policy analysis (Manski, 2011, 2020, 2024). Researchers often are aware that they cannot form a credible point estimate of a policy-relevant estimand. They could face up to uncertainty and determine what they can credibly infer about this quantity, perhaps obtaining an informative bound. Instead, they change the objective and focus on another estimand that is not of substantive interest but that can be point-estimated credibly. Thus, they sacrifice relevance for certitude.

Notable scientists have critiqued this practice, but it persists. John Tukey wrote (Tukey, 1962, p. 13-14): "Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise." Many cite versions of the joke about the drunk and the lamppost. Noam Chomsky has been quoted as putting it this way (Barsky, 1998, p. 95): "Science is a bit like the joke about the drunk who is looking under a lamppost for a key that he has lost on the other side of the street, because that's where the light is."

Why are so many researchers content to look under a lamppost? It is rare to find an explicit rationale. One study that tried to motivate the practice was the Angrist (1990) study of effect of treatment on the treated. Here, treatment b denoted being drafted into American military service during the period of the Vietnam War draft lottery, and treatment a denoted not being drafted. Taking the outcome to be a person's lifetime earnings, Angrist estimated $E[y(b)|z = b] - E[y(a)|z = b]$, the average effect of the draft on the lifetime earnings of those who were drafted. He motivated interest in this quantity by writing (p. 313): "A

central question in the debate over military manpower policy is whether veterans are adequately compensated for their service."

Both within and outside of economics, scientists sometimes motivate research as an effort to improve our "understanding" of a subject, and argue that this is a worthwhile objective even if there are no further implications. A notable example occurs in a text on statistical methods in epidemiology. Fleiss (1981, p. 92) states that the retrospective studies of disease that are a staple of medical research do not yield policy-relevant predictions and so are "necessarily useless from the point of view of public health." Nevertheless, Fleiss goes on to say that "retrospective studies are eminently valid from the more general point of view of the advancement of knowledge."

Credible Policy Analysis

Justifications of the Angrist or Fleiss sort do not appear in my research. I have sought to develop methodology enabling empirical research to be both credible and policy relevant. In the 1990s I studied partial identification of treatment response using observational or experimental data; see Manski (1990, 1995, 1997a, 1997b), Horowitz and Manski (2000), and Manski and Pepper (2000). I formally connected partial identification analysis with policy choice in Manski (2000), writing (p. 416):

> "This paper connects decisions under ambiguity with identification problems in econometrics. Considered abstractly, it is natural to make this connection. Ambiguity occurs when lack of knowledge of an objective probability distribution prevents a decision maker from solving an optimization problem. Empirical research seeks to draw conclusions about objective probability distributions by combining assumptions with observations. An identification problem occurs when a specified set of assumptions combined with unlimited observations drawn by a specified sampling process does not reveal a distribution of interest. Thus, identification problems generate ambiguity in decision making."

I borrowed the term *ambiguity* from Ellsberg (1961), who was concerned with uncertainty that may not be expressible by a probability distribution.

The above quoted paragraph sketches the basic problem: combining available data with credible assumptions may not reveal whether one policy outperforms another in terms of the social welfare it yields. When this occurs, a social planner faces a problem of policy choice under uncertainty. The statistical imprecision of inference using finite samples adds further uncertainty. To cope with the latter problem, I have brought to bear statistical decision theory, beginning in Manski (2004). From these beginnings, I have developed a program of research on credible social planning under uncertainty, with journal articles developing specific ideas and the program of work exposited comprehensively in the book Manski (2024).

My research on planning under uncertainty differs profoundly in philosophy and practice from design-based study of treatment effects. Research of the latter type prizes point identification of average treatment effects in some treatment-determined sub-population of a study population.[3] It does not seek to predict how prospective treatment choices would affect a population of substantive concern.

In contrast, the objective of my research is to evaluate policy prospectively. I specify a social welfare function for a population of interest and study how combination of available data with credible assumptions yields conclusions about the welfare yielded by alternative policy choices. Adhering to the prevalent practice in public economics and benefit-cost analysis, I have mainly studied utilitarian welfare functions. However, my broad concern with planning under uncertainty is not limited to utilitarianism. It applies just as much if the social welfare function expresses non-utilitarian notions of paternalism, justice, or fairness.

My work has the same objective as econometric research using structural models to predict policy impacts, but it differs in practice. Structural econometrics has used strong assumptions to point-identify

[3] In regression-discontinuity analysis, the sub-population is the observable group comprising persons (or other treatment units) who were close to the threshold determining treatment assignment. In difference-in-differences analysis, it is the observable group comprising those whose treatment changed over time. In LATE analysis, it is the unobservable group comprising those whose treatment would change if the value of an instrumental variable were to hypothetically change.

policy impacts and has studied optimal policy under those assumptions. The weaker assumptions maintained in my research partially identify policy impacts. Hence, a planner must choose policy under uncertainty.

The most familiar economic prescription for decision making under uncertainty has been maximization of subjective expected utility. In applications to policy choice, a planner would place a probability distribution on a specified space of feasible states of nature and choose an action that maximizes the expected value of welfare with respect to this distribution. Considering settings where it is not credible to express uncertainty through a subjective probability distribution, decision theorists have studied criteria that, in some sense, works uniformly well over all of the state space. Two prominent interpretations of this broad idea are the maximin and minimax-regret (MMR) criteria. The maximin criterion maximizes the minimum welfare attainable across the state space. In the MMR approach, one evaluates a policy by its maximum regret over all states and selects a policy that minimizes maximum regret. The maximum regret of a policy measures its maximum distance from optimality across states.

Being concerned with planning in the absence of a subjective distribution on the state space, I have studied planning using the minimax-regret criterion to cope with the ambiguity generated by partial identification of treatment effects. The maximin and MMR criteria both provide ex ante evaluations of the worst result that a decision maker may experience ex post, but they do so in different ways. Whereas maximin considers the worst absolute outcome that an action may yield across states, MMR considers the worst outcome relative to the best achievable. The criteria are equivalent only in special cases, particularly when optimal welfare is invariant across states.

I have argued in Manski (2021, 2024) and elsewhere for the conceptual appeal of the MMR criterion. The term “maximum regret” is shorthand for the maximum sub-optimality of a policy choice across the

feasible states of nature. A decision with small maximum regret is uniformly near optimal across all states. I think this a desirable property.[4]

The remainder of this paper demonstrates how my research differs from design-based analysis of treatment effects and structural econometric study of optimal planning. I think that an apt demonstration, which should be of broad interest to applied microeconomists, is to focus on a simple finding reported in Manski (2007, Chapter 11) and Manski (2009). The finding is that a planner who uses the MMR criterion to choose a treatment for each member of an observational identical population will diversify treatment, assigning a positive fraction of the population to each treatment. I have subsequently applied this finding to address policy choice in Manski (2015), Cassidy and Manski (2019), and Manski (2025).

Section 2 exposits the general analysis. Section 3 discusses the application to choice of treatment dosage in Manski (2025), which uses a credible structural assumption to partially identify a social welfare function. Section 4 concludes.

2. Diversified Treatment under Ambiguity

When it is feasible to treat persons differentially, a planner may make a *singleton* allocation, assigning all observationally identical persons to the same treatment. Or the planner may choose a *fractional* allocation, randomly assigning positive fractions of these persons to different treatments. Fractional allocations cope with ambiguity through diversification.

Suppose that there are two feasible treatments, A and B. Diversification enables a decision maker to balance two types of potential error. A Type A error occurs when treatment A is chosen but is inferior to B, and a Type B error occurs when B is chosen but is inferior to A. The singleton allocation assigning everyone

[4] This argument for the appeal of the MMR criterion is prescriptive. A separate empirical question, warranting investigation, is how actual planners and other decision makers cope with uncertainty in practice.

to treatment A avoids type B errors but may yield Type A errors, and vice versa for singleton assignment to treatment B. Fractional allocations make both types of errors but reduce their potential frequencies. Thus, treatment diversification enables a planner to avoid gross errors that would occur if all persons were given the inferior treatment.

Treatment diversification is analogous to financial diversification. Diversification enables an investor facing uncertain asset returns to limit the potential negative consequences of placing "all eggs in one basket." It is well known that an investor seeking to maximize subjective expected utility chooses to diversify if the probability distribution of returns has sufficient spread and if the investor is sufficiently risk averse, utility being a sufficiently concave function of the return to the investment. Treatment allocation by a planner can be studied in the same manner. It can also be studied using the maximin or minimax-regret criterion, as shown below.

2.1. Utilitarian Treatment Choice

Manski (2007, 2009) considered a planner who chooses a treatment for each member j of a population J of observationally identical persons. Persons who are observationally identical need not respond uniformly to treatment. Being observationally identical means only that all persons share the same observed covariates. The objective is to maximize utilitarian social welfare, adding personal welfare outcomes across the population. Equivalently, the planner wants to maximize mean welfare in the population.

Person j has a response function $y_j(\cdot)$ mapping treatments t into outcomes $y_j(t)$. $P[y(\cdot)]$ denotes the population distribution of treatment response. Let the population be a continuum, with $P(j) = 0$ for all $j \in J$. This technical regularity assumption implies that, when treatment assignment is fractional, the realized fractions of persons who randomly receive different treatments have the same distribution of treatment response. The task is to allocate the population to treatments A and B. An allocation $\delta \in [0, 1]$ randomly assigns a fraction δ of the population to treatment B and the remaining $1 - \delta$ to A.

Let $u_j(t) \equiv u_j[y(t), t]$ be the cardinal welfare of person j when this person receives treatment t and realizes outcome $y_j(t)$. Let $\alpha \equiv E[u(A)]$ and $\beta \equiv E[u(B)]$ be mean personal welfare if all members of the population were to receive treatment A or B. Mean welfare with allocation δ is

$$(5.1) \qquad w(\delta) = \alpha(1 - \delta) + \beta\delta = \alpha + (\beta - \alpha)\delta.$$

$\delta = 1$ is optimal if $\beta \geq \alpha$ and $\delta = 0$ if $\beta \leq \alpha$. Thus, optimal treatment choice is determined by the sign of the population-wide average treatment effect $\beta - \alpha$.

The problem of treatment choice under uncertainty arises when the planner has partial knowledge of (α, β). The planner may have partial knowledge due to the many identification problems faced in empirical analysis of treatment response; Manski (2007) provides a textbook exposition. Structural econometric modeling deals with identification problems by making assumptions strong enough to point-identify (α, β), sacrificing credibility. Design-based analysis deals with them by changing the estimand from (α, β) to one that can be credibly point-identified, sacrificing policy relevance. Both research approaches avoid recognition of uncertainty.

My study of diversified treatment choice assumed only that the planner knows (α, β) to lie in a bounded set $[(\alpha_s, \beta_s), s \in S]$. In partial identification analysis, this set is called the *identification region* or *identified set* for (α, β). In decision theory, it is called the *state space*. Let $\alpha_L \equiv \min_{s \in S} \alpha_s$, $\beta_L \equiv \min_{s \in S} \beta_s$, $\alpha_U \equiv \max_{s \in S} \alpha_s$, and $\beta_U \equiv \max_{s \in S} \beta_s$.

The planner faces ambiguity if $\alpha_s > \beta_s$ for some values of s and $\alpha_s < \beta_s$ for other values, implying that the optimal treatment is not known. Several criteria for choice under ambiguity have been prominent. These include

Maximization of Subjective Expected Welfare (aka Bayesian Planning)

The planner places a subjective distribution π on S and solves

(1) $$\max_{\delta \in [0,1]} E_\pi(\alpha) + [E_\pi(\beta) - E_\pi(\alpha)]\delta,$$

where $E_\pi(\alpha) = \int \alpha_s d\pi$ and $E_\pi(\beta) = \int \beta_s d\pi$. The planner chooses $\delta = 0$ if $E_\pi(\beta) < E_\pi(\alpha)$ and $\delta = 1$ if $E_\pi(\beta) > E_\pi(\alpha)$. He is indifferent among all δ if $E_\pi(\beta) = E_\pi(\alpha)$.

Maximin Planning

A maximin planner solves

(2) $$\max_{\delta \in [0,1]} \min_{s \in S} \alpha_s + (\beta_s - \alpha_s)\delta.$$

In general, the policy choice depends on the structure of S. If (α_L, β_L) is feasible, the decision is $\delta = 0$ if $\beta_L < \alpha_L$, $\delta = 1$ if $\beta_L > \alpha_L$, and all δ if $\beta_L = \alpha_L$.

Minimax-Regret Planning

The regret of δ in state s is the difference between the maximum achievable welfare and the welfare achieved with allocation δ. Maximum welfare in state s is $\max(\alpha_s, \beta_s)$. The minimax-regret criterion is

(3) $$\min_{\delta \in [0,1]} \max_{s \in S} \{\max(\alpha_s, \beta_s) - [\alpha_s + (\beta_s - \alpha_s)\delta]\}.$$

Let $S(A) \equiv \{s \in S: \alpha_s > \beta_s\}$ and $S(B) \equiv \{s \in S: \beta_s > \alpha_s\}$. Let $M(A) \equiv \max_{s \in S(A)} (\alpha_s - \beta_s)$ and $M(B) \equiv \max_{s \in S(B)} (\beta_s - \alpha_s)$. Manski (2007, 2009) showed that the MMR allocation to treatment B is $\delta_{MMR} = M(B)/[M(A) + M(B)]$. This is an interior fraction when treatment is under ambiguity, as $M(A) > 0$ and $M(B) > 0$. If (α_L, β_U) and (α_U, β_L) are feasible, the MMR allocation to B simplifies to $\delta_{MMR} = (\beta_U - \alpha_L)/[(\alpha_U - \beta_L) + (\beta_U - \alpha_L)]$.

The basic idea underlying this result is that singleton and fractional treatment rules have different regret properties across states of nature. Each singleton rule is the best rule in some states of nature and the

worst in the others. Hence, singleton rules have zero regret in some states but high maximum regret. In contrast, fractional rules yield intermediate social welfare in all states of nature. They have regret in all states, but smaller maximum regret. Thus, the minimax-regret rule is fractional rather than singleton.

2.2. Welfare Increasing in Mean Personal Welfare

A social welfare function may increase with mean personal welfare rather than equal mean welfare. Thus, consider $w(\delta) = f[\alpha + (\beta - \alpha)\delta]$, where $f(\cdot)$ is strictly increasing.

It is well known that the Bayesian decision is singleton if $f(\cdot)$ is convex and may be fractional if $f(\cdot)$ has concave segments. In finance, this is the well-known finding that a risk-seeking investor, whose utility is convex in income, does not diversify. A risk-averse investor, whose utility is concave in income, may diversify.

The shape of $f(\cdot)$ does not affect the maximin decision. The maximin criterion only uses ordinal, not cardinal properties of the welfare function.

Manski (2007, 2009) showed that the MMR allocation is fractional if $f(\cdot)$ is continuous and the planner faces ambiguity. If $f(\cdot) = \log(\cdot)$ and $\{(\alpha_L, \beta_U), (\alpha_U, \beta_L)\}$ are feasible, the MMR allocation has an explicit form, this being $\delta_{MMR} = [\alpha_U(\beta_U - \alpha_L)]/[\alpha_U(\beta_U - \alpha_L) + \beta_U(\alpha_U - \beta_L)]$.

2.3. Diversification and Equal Treatment of Equals

Proposing that an investor may want to choose a diversified portfolio is uncontroversial. It is uncontroversial to suggest that a firm diversify when making production decisions. However, I have found it controversial to propose diversification of treatments to humans. The concern is that diversification violates the ethical principle calling for *equal treatment of equals*. Utilitarian planning does not address this ethical concern. Equal treatment of equals is a deontological consideration. It supposes that actions have intrinsic value, apart from their consequences.

Manski (2009) emphasized that it is important to distinguish ex ante and ex post interpretations of equal treatment of equals. Diversification is consistent with the equal-treatment principle in the *ex-ante* sense that all members of the population have the same probability of receiving a particular treatment. It violates the principle *ex post*: Different persons ultimately receive different treatments.

2.4. Adaptive Diversification

I have thus far considered a static planning environment. Now consider a planner who makes treatment decisions in a sequence of periods, facing a new population each period. Assume that each successive population has the same distribution of treatment response. The planner may observe the outcomes of early decisions and use this evidence to inform treatment later. Diversification is advantageous for learning treatment response because it generates randomized experiments. As evidence accumulates, the planner can revise the fraction of persons assigned to each treatment in accord with the available knowledge. I have called this *adaptive diversification*.

It has long been appreciated that variation in policy promotes learning. What is new in the concept of adaptive diversification is recognition that policy variation at a point in time yields the benefit of diversification in reducing gross errors, which occurs even in the absence of learning over time.

Adaptive diversification can be achieved by a Bayesian planner who updates his subjective distribution on treatment response after observing the outcomes of each population. I suggested that a planner who does not want to specify a subjective distribution on treatment response might use the *adaptive minimax-regret* (*AMR*) criterion. In each period, this criterion applies the static MMR criterion using the information available at the time. It is adaptive because successive populations may receive different allocations as knowledge of treatment response increases over time. The AMR criterion treats each population as well as possible in the MMR sense, given the available knowledge. It does not ask persons in one period to sacrifice for the benefit of future persons. Nevertheless, diversification enables learning about treatment response.

Implementation of adaptive diversification may be possible in centralized health-care systems where there exists a planning entity who chooses treatments for a broad patient population. Examples are the Military Health System in the United States and the National Health Service in the United Kingdom. Cassidy and Manski (2019) suggested adaptive diversification of testing for and treatment of tuberculosis in public-health clinics in developing countries.

3. Using Trial Evidence to Choose Treatment Dosage when Efficacy and Toxicity Increase with Dose

Medical decision making offers a fertile setting for studying social planning under uncertainty. The planner is a clinician or a public health agency, prescribing treatments to patients. Medical ethics suggests a utilitarian welfare function, the planner aiming to maximize mean health in the patient population.

Medical economists have adopted this perspective and have studied optimal patient care under the assumption that the planner knows the distribution of treatment response in the patient population. See, for example, Phelps and Mushlin (1988) and Manski, Mullahy, and Venkataramani (2023). However, studies of optimal care have limited credibility. Identification problems in analysis of treatment response imply that medical planners commonly have only partial knowledge of treatment response.

I have used the MMR criterion to study several classes of problems of medical decision making under ambiguity, ranging from personalized clinical treatment (Manski, 2013, 2018) to the public-health problem of vaccination of broad populations (Manski, 2010, 2017). I discuss here my recent study of choice of treatment dosage using limited experimental evidence (Manski, 2025). This research characterized the identification problem when a credible structural assumption is maintained and showed how to compute MMR decisions. The analysis distinguished the choice problem of a clinician, who treats one patient at a time, and that of a health agency, which can diversify treatment across a population of patients. The latter analysis applied the work described in Section 2.

3.1. Background

It has been standard in medicine to base dosage of drug treatments on evidence in randomized trials. Yet trial evidence has commonly been limited to comparison of at most a few dose levels. Evidence on dose response has been particularly limited in Phase III trials performed to obtain FDA approval to market new drugs. These trials specify some dose of a new drug and compare it with an established therapy or placebo. When evaluating cancer drugs, the specified dose is often the *maximum tolerated dose*. To determine this dose, pharma firms may conduct "dose-finding trials" in Phase II studies, but Phase II sample sizes are usually small and findings unpublished.[5]

Trials provide no direct information about patient outcomes with doses not included in the study design. I write "no direct information" because the standard practice in analysis of experimental data has been to view each treatment arm as qualitatively different. Consider treatments (A, B, C). In standard experimental analysis, trial findings with treatment A are not used to draw conclusions about B and C, and vice versa.

Standard trial analysis of treatment dosage would be appropriate if different dose levels were qualitatively different treatments. However, medical researchers generally find it credible to characterize a dose level by two attributes: efficacy in reducing the disease of primary concern and toxicity generating secondary adverse events (AEs) in patients. Medical researchers often think it credible to assume that efficacy and AEs increase with dose. These structural assumptions have identifying power.

[5] The long-run solution to the limited evidence on dose response is to perform new trials that enrich the available data. Unfortunately, there is scant reason to expect that pharma firms will voluntarily enhance Phase III trials to have multiple dose arms. Given a fixed total sample size, a multi-armed trial reduces sample size per arm, lowering statistical power and making it harder to obtain drug approval. Given that FDA drug approval is based primarily on drug efficacy rather than toxicity, pharma firms have a statistical incentive to seek approval of the designated maximum tolerated dose rather than a smaller dose.

With this background, Manski (2025) studied the use of limited trial evidence to choose treatment dosage when efficacy and AEs increase with dose.[6] Identification analysis showed that findings on outcomes in a trial comparing a limited set of doses imply informative bounds on outcomes at other dose levels for which no experimental evidence exists. Study of MMR decision making showed that, in some cases, a clinician or public health agency may reasonably choose a dose level for which no experimental evidence exists. I explain below.

3.2. The Dosage Choice Problem

Let dose be an integer $t \in (0, . ,T)$, where T is a specified maximum dose. Let the objective be to maximize mean patient welfare, which is a function of treatment efficacy, AEs, and cost. To simplify, consider illness and AEs to be binary events rather than outcomes that vary in severity. Suppose that dosage is a one-time choice rather than a dynamic decision. Suppose that trial evidence is available for $K < T + 1$ dose levels.

Let j label a patient. Dose-dependent outcomes $[d_j(t), e_j(t)]$ determine patient welfare. Given dose t, patient j may experience a disease [$d_j(t) = 1$ if yes, $d_j(t) = 0$ if no] and/or an AE [$e_j(t) = 1$ if yes, $e_j(t) = 0$ if no]. Thus, $[d_j(t), e_j(t)]$ takes one of the values [(0, 0), (1, 0), (0, 1), (1, 1)]. Patient welfare is $w_j[d_j(t), e_j(t)]$. Treatment cost, including monetary and non-monetary costs, is a function $g_j(t)$ of dose level, measured in

[6] The methodology developed in this research may be applied as well to other settings where treatments can be administered in different magnitudes, with benefits and costs that increase with the magnitude of the treatment. Consider, for example, provision of mathematics tutoring to students. Here dosage is the amount of time devoted to tutoring. A larger intensity of tutoring should be more effective in helping students learn math, but it will have AEs. Time spent tutoring crowds out other activities for the student and the tutor. See, for example, Higbee (2023).

the same units as welfare. Indexing outcomes, welfare, and cost by j permits heterogeneity across the patient population.

The objective is to maximize welfare net of cost. Interpretation of optimal dosage depends on the available knowledge. If a clinician has perfect foresight, an optimal dose for patient j solves

$$\max_{t = 0, \ldots, T} w_j[d_j(t), e_j(t)] - g_j(t).$$

Medical economists have studied optimal utilitarian care, assuming objectively correct probabilistic expectations (aka rational expectations). Assume that a clinician observes some patient covariates. Viewing j as a member of a population of patients with the same covariates, assume that the clinician knows the dose-dependent distribution of (welfare, cost). Then an optimal dose solves

$$\max_{t = 0, \ldots, T} E\{w[d(t), e(t)]\} - E[g(t)].$$

Standard data collection in trials yields evidence on efficacy and AEs, but not on welfare as a function of these outcomes. Medical economists sometimes conduct separate studies that sample patients and aim to learn their welfare functions, by questioning them about the choices they would make if they were to experience hypothetical disease and AE outcomes. The patients in these studies typically are different than the subjects in trials. Estimation of dose-dependent mean welfare may be feasible if it is credible to assume that welfare functions $w(\cdot, \cdot)$ are mean-independent of disease and AE outcomes. Then mean welfare with dose t is

(4) $E\{w[d(t), e(t)]\} =$

$$E[w(0, 0)] \cdot p[d(t) = 0, e(t) = 0] + E[w(1, 0)] \cdot p[d(t) = 1, e(t) = 0]$$
$$+ E[w(0, 1)] \cdot p[d(t) = 0, e(t) = 1] + E[w(1, 1)] \cdot p[d(t) = 1, e(t) = 1] - E[g(t)].$$

Assume that $E[w(\cdot, \cdot)]$ and $E[g(\cdot)]$ are known. The problem is that the available trial data do not fully reveal the probabilities p of different (disease, AE) outcomes. A K-armed trial, assigning subjects to dosages $(t_k, k = 1, . , K)$, enables one to estimate $p[d(t_k), e(t_k)]$ for $k = 1, . , K$, but not $p[d(t), e(t)]$ for other dose levels. Suppose that the trial sample size is large enough that statistical imprecision is negligible. Then one

can optimize dosage over (t_k, k = 1, . , K), but not over t = 0, . , T.

3.3. Identification Assuming Monotone Dose Response

Now introduce the structural assumption that efficacy and AEs weakly increase with dose. Formally:

Monotone Dose Response: Consider doses (s, t) with t > s. For each patient j, assume

Monotone Efficacy (ME): $d_j(s) = 0 \Rightarrow d_j(t) = 0$.

Monotone AE (MT): $e_j(s) = 1 \Rightarrow e_j(t) = 1$.

Assumptions ME and MT place substantive restrictions on dose response. They are often credible, but they may not hold universally. Either may fail if there exist biological interactions between disease and AEs.[7]

It is analytically convenient to express monotone dose response using the concept of latent patient-specific threshold dose levels (t_{dj}, t_{ej}):

Monotone Efficacy (ME): Either $d_j(t) = 1$, all $t \in \{0, . , T\}$, or there exists a threshold dose $t_{dj} \in \{0, . ,T\}$ such that $d_j(t) = 1$ for $t < t_{dj}$ and $d_j(t) = 0$ for $t \geq t_{dj}$. When $d_j(t) = 1$, all $t \in \{0, . , T\}$, it suffices to define an infeasible threshold dose $t_{dj} = T + 1$.

[7] Monotonicity assumptions have been used in structural econometrics, but their use here differs from elsewhere. Assumptions ME and MT are reminiscent of the *monotone treatment response* assumption studied in Manski (1997). There I analyzed partial identification using observational data when a univariate outcome varies monotonically with treatment intensity. The present setting differs because the (efficacy, AE) outcome is bivariate, with increasing dosage increasing both efficacy and AEs.

Monotone AE (MT): Either $e_j(t) = 0$, all $t \in \{0, . , T\}$ or there exists a threshold dose $t_{ej} \in \{0, . , T\}$ such that $e_j(t) = 0$ for $t < t_{ej}$ and $e_j(t) = 1$ for $t \geq t_{ej}$. When $e_j(t) = 0$, all $t \in \{0, . ,T\}$, it suffices to define an infeasible threshold dose $e_{dj} = T + 1$.

Latent threshold dose levels are reminiscent of reservation wages in the labor economics theory of job search. The two versions of assumptions (ME)−(MT) are equivalent, as $d_j(t) = 1[t < t_{dj}]$ and $e_j(t) = 1[t \geq t_{ej}]$.

Let $q(t_d, t_e)$ denote the population distribution of threshold dose levels. Proposition 1 of Manski (2025) showed that $q(t_d, t_e)$ solves $3K + 1$ non-redundant linear equations in $(T + 2)^2$ unknowns, plus linear inequalities requiring probabilities to be non-negative. Thus, the feasible $q(t_d, t_e)$ form a $(T + 2)^2 – (3K + 1)$ dimensional convex polygon. (disease, AE) probabilities are linear in $q(t_d, t_e)$, so the feasible ones form a convex polygon. Mean welfare is linear in (disease, AE) probabilities, so the feasible values form an interval.

This Proposition illustrates a recurrent theme in partial identification analysis, wherein the identification region for an estimand of interest comprises the values of the estimand that satisfy a set of linear equalities and inequalities. The proof is instructive, so I include it in this paper as an Appendix.

3.4. Minimax-Regret Dosage Choice

Consider a clinician treating one patient and a health planner treating a population. The clinician must choose a single treatment. The planner can choose a fractional treatment allocation $\delta(t)$, $t \in \{0, . ,T\}$. Thus, the planner has a richer set of options than the clinician.

The Bayesian prescription is to assert a subjective probability distribution on $q(t_d, t_e)$ and maximize subjective expected welfare. The absence of a credible subjective distribution motivates study of decision making under ambiguity.

I evaluated options by their maximum regret. To explain, I first pose the MMR criterion abstractly and then apply it to dosage choice. Consider a planner who faces choice set C and believes that the true state of

nature lies in specified state space S. An objective function $f(\cdot,\cdot)$: $C \times S \rightarrow R^1$ maps actions and states into welfare. The planner wants to maximize true welfare but does not know the true state.

In settings without statistical imprecision, the MMR criterion solves

$$(5) \qquad \min_{c \in C} \max_{s \in S} [\max_{b \in C} f(b, s) - f(c, s)].$$

Here $\max_{b \in C} w(b, s) - w(c, s)$ is the regret of action c in state s; that is, the degree of suboptimality.

In the present setting, the state space indexes the feasible threshold distributions $q(t_d, t_e)$ derived in Proposition 1; thus, S = Q. In the clinical case, the choice set comprises all feasible doses; thus, C = {0, . . , T}. In the planning case, the choice set comprises all fractional allocations; thus, C = Δ.

In the clinical case, the objective function in state s is $f[t, q_s(t_d, t_e)] = E_s\{w[d(t), e(t)]\}$, where $q_s(t_d, t_e)$ is the distribution indexed by state s and E_s denotes expectation with respect to this distribution. In the planning case, the objective function is the expectation of the mean welfare function with respect to the treatment allocation; thus, $f[\delta(\cdot), q_s(t_d, t_e)] = \sum_{t = 0. . . , T} \delta(t) \cdot E_s\{w[d(t), e(t)]\}$.

3.5. Computation

Exact computation of the MMR decision is tractable in the clinical case when T is not extremely large. An alternative expression of (5) reverses the two max operations to obtain

$$(5') \qquad \min_{c \in C} \max_{b \in C} \max_{s \in S} [f(b,s) - f(c,s)].$$

Holding c and b fixed, consider the inner maximization over S. When c = b, the maximum is 0. When $c \neq b$, $f(b, s) - f(c, s)$ is linear in $q_s(\cdot, \cdot)$ in the setting of this analysis. $\max_{s \in S} [f(b, s) - f(c, s)]$ is a linear programming problem, solvable with standard algorithms. Considering all $c \neq b$, determination of the MMR dosage requires solution of $T \cdot (T + 1)$ linear programming problems and comparison of the solutions.

Performing computations in the T = 2 and T = 3 settings, I found that a clinician using the MMR criterion sometimes chooses a dose level that was not studied in the trial.

In the planning case, $\max_{s \in S} [f(b, s) - f(c, s)]$ remains a linear programming problem, but $f(b, s) - f(c, s)$ is a different linear function of $q_s(\cdot, \cdot)$. The set of feasible dosage allocations is the entire simplex Δ on R^{T+1} rather than only the vertices, which place probability one on single treatments. When T > 2, this makes exact computation of the MMR dosage allocation infeasible. However, it is feasible to use a finite grid to approximate Δ and solve the associated finite number of linear programming problems.

3.6. Population-Health Dosage Allocation when T = 2

Computation of the public health planner's MMR allocation is simple when T = 2, K = 2, $t_1 = 0$, and $t_2 = 2$. Thus, there is no trial evidence for t = 1. This case is important in dosage choice for cancer drugs. As discussed earlier, Phase III trials commonly compare zero dose and a maximum tolerable dose, without evaluating smaller positive doses.

Mean welfare net of treatment cost is point-identified for t = 0 and t = 2. Let them be denoted $\omega_0 \equiv E\{w[d(0), e(0)]\} - E[g(0)]$ and $\omega_2 \equiv E\{w[d(2), e(2)]\} - E[g(2)]$. Mean welfare is partially identified for t = 1, being known to lie in a computable interval, say $[\omega_{1L}, \omega_{1U}]$, determined by the identification analysis. Dosage choice is a problem of decision making under ambiguity when $\omega_{1L} < \max(\omega_0, \omega_2) < \omega_{1U}$.

The analysis of MMR allocation of a population between two treatments, discussed in Section 2, applies when there is ambiguity. Dosage t = 1 and either t = 0 or t = 2 are undominated. The MMR allocation assigns positive fractions of patients to t = 1 and to the other dose with the higher value of ω. If $\omega_2 > \omega_0$, t = 0 is dominated and receives zero allocation. The fraction of the population assigned to t = 1 is $(\omega_{1U} - \omega_2)/(\omega_{1U} - \omega_{1L})$ and the fraction assigned to t = 2 is $(\omega_2 - \omega_{1L})/(\omega_{1U} - \omega_{1L})$. Symmetrically, if $\omega_2 < \omega_0$, t = 2 is dominated and receives zero allocation. The fraction assigned to t = 1 is $(\omega_{1U} - \omega_0)/(\omega_{1U} - \omega_{1L})$ and that assigned to t = 0 is $(\omega_0 - \omega_{1L})/(\omega_{1U} - \omega_{1L})$.

4. Conclusion

The importance of public policy creates a concern that the findings of policy analysis be widely credible. The weaker are the assumptions imposed, the more credible are the findings. There has been considerable concern with the credibility of policy analysis with structural econometric models that are point-identified under untenably strong assumptions. A response of applied microeconomists has been to perform design-based analysis, which estimates average treatment effects in treatment-determined sub-populations of a study population. There has been a proliferation of research studying regression discontinuities, difference-in-differences, and local average treatment effects. The estimands may be credibly point-identified, but they lack policy relevance.

My research has aimed to specify problems of policy choice for populations of welfare-economic interest and determine what research can credibly learn about policy impacts in these populations. The findings have been partial identification of treatment effects. I have used basic theory on decision making under ambiguity to address reasonable policy choice in partially identified settings, with emphasis on the MMR criterion. I have used statistical decision theory to recognize the statistical imprecision that is inevitable with use of finite-sample data. This paper has used my study of diversified treatment under ambiguity and treatment dosage to illuminate some central themes.

There are computational and conceptual challenges to the study of partial identification of treatment response and MMR policy choice. Consider computation. In settings such as the clinical one discussed in Section 3, the identification region for mean treatment response is an easily computed convex set determined by a vector of linear equalities and inequalities, However, computation is complex in other settings. An example is the setting of the health planner in Section 3 when $T > 2$. Another complex setting was studied in Li, Litvin, and Manski (2023), where mean treatment response satisfies a vector of bi-linear equations. Given characterization of the identification region, there remains the problem of computation of MMR policy choices. MMR problems have tractable analytical solutions in some cases, but computation

commonly requires numerical methods to find approximate solutions. See Manski (2021) and Dominitz and Manski (2025) for discussion.

A conceptual challenge when studying partial identification is to decide what assumptions are sufficiently credible that they should be maintained when considering policy choice. The Law of Decreasing Credibility expresses the general dilemma, but it does not prescribe how to cope with the dilemma in practice. Maintaining stronger assumptions shrinks the state space. This reduces the maximum regret of alternative actions if the stronger assumptions are correct, but the benefit is illusory if the assumptions are incorrect. Credibility can be difficult to assess. In Manski (2007) I wrote (p. 48): "Credibility is a property of an assumption and the person contemplating it. An assumption is credible to the degree that someone thinks it so."

In this paper, the discussion of diversified treatment in Section 2 did not address how a planner should specify the state space in a particular setting. I assumed only that the space is bounded. The discussion of treatment dosage in Section 3 did address specification of the state space. I maintained assumptions ME and MT because they are broadly credible with drug treatment. I imposed no other assumptions because I am not aware of any that have similar credibility.

To help policy analysts cope with the Law of Decreasing Credibility, I have recommended performance of a "layered" analysis that begins with weak and highly credible assumptions and then adds stronger, less credible ones. Illustrative applications include the Manski and Nagin (1998) study of criminal sentencing and recidivism and the Manski (2014) study of the impact of income tax policy on labor supply. In his 2015 Cowles Lecture, Elie Tamer suggested a complementary "top down" approach that begins with a point-identified parametric structural model and then relaxes assumptions deemed particularly worrisome.[8] Both approaches can be useful in practice.

[8] https://tamer.scholars.harvard.edu/sites/g/files/omnuum7756/files/tamer/files/cowles-2015.pdf, accessed April 20, 2025.

Appendix: Proof of Proposition 1in Manski (2025)

The non-negativity of probabilities and the Law of Total Probability imply that $q(t_d, t_e)$ satisfies these inequalities and equality:

(A1) $q(t_d = h, t_e = i) \geq 0, h \in \{0, . , T + 1\}, i \in \{0, . , T + 1\}.$

(A2) $1 = \sum_{h=0}^{T+1} \sum_{i=0}^{T+1} q(t_d = h, t_e = i).$

The dose-dependent outcome distributions $p[d(t), e(t)]$, $t \in \{0, . , T\}$ have $4(T + 1)$ component probabilities. Equations (A3a)−(A3d) express outcome probabilities in terms of $q(t_d, t_e)$. For $t \in \{0, . , T\}$,

(A3a) $p[d(t) = 0, e(t) = 0] = q(t_d \leq t, t_e > t) = \sum_{h=0}^{t} \sum_{i=t+1}^{T+1} q(t_d = h, t_e = i).$

(A3b) $p[d(t) = 1, e(t) = 0] = q(t_d > t, t_e > t) = \sum_{h=t+1}^{T+1} \sum_{i=t+1}^{T+1} q(t_d = h, t_e = i).$

(A3c) $p[d(t) = 0, e(t) = 1] = q(t_d \leq t, t_e \leq t) = \sum_{h=0}^{t} \sum_{i=0}^{t} q(t_d = h, t_e = i).$

(A3d) $p[d(t) = 1, e(t) = 1] = q(t_d > t, t_e \leq t) = \sum_{h=t+1}^{T+1} \sum_{i=0}^{t} q(t_d = h, t_e = i).$

These preliminaries yield the identification regions Q, P, and W. Q comprises all distributions $q(t_d, t_e)$ that satisfy (A1), (A2), and, for $k = 1, . , K$,

(A4a) $p[d(t_k) = 0, e(t_k) = 0] = q(t_d \leq t_k, t_e > t_k) = \sum_{h=0}^{t_k} \sum_{i=t_k+1}^{T+1} q(t_d = h, t_e = i),$

(A4b) $p[d(t_k) = 1, e(t_k) = 0] = q(t_d > t_k, t_e > t_k) = \sum_{h=t_k+1}^{T+1} \sum_{i=t_k+1}^{T+1} q(t_d = h, t_e = i),$

(A4c) $p[d(t_k) = 0, e(t_k) = 1] = q(t_d \leq t_k, t_e \leq t_k) = \sum_{h=0}^{t_k} \sum_{i=0}^{t_k} q(t_d = h, t_e = i),$

(A4d) $p[d(t_k) = 1, e(t_k) = 1] = q(t_d > t_k, t_e \leq t_k) = \sum_{h=t_k+1}^{T+1} \sum_{i=0}^{t_k} q(t_d = h, t_e = i).$

P comprises all distributions $p[d(t), e(t)]$, $t \in \{0, . , T\}$ that satisfy (A3a)−(A3d) for some element of Q. W comprises all means $E\{w[d(t), e(t)]\}$, $t \in \{0, . , T\}$ that satisfy (1) for some element of P.

Q. E. D.

References

Angrist, J. (1990), "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review*, 80, 313-336.

Angrist, J. and J. Pischke (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24, 3-30**.**

Barsky, R. (1998), *Noam Chomsky: A Life of Dissent,* Cambridge: MIT Press.

Bassi, L., and O. Ashenfelter (1986), "The Effect of Direct Job Creation and Training Programs on Low-Skilled Workers," In S. Danziger and D. Weinberg, eds. *Fighting Poverty*. Cambridge, Mass.: Harvard University Press.

Bloom, H. (1984), "Accounting for No-Shows in Experimental Evaluation Designs," *Evaluation Review*, 8, 225-246.

Cain, G. (1986), "The Issues of Marital Stability and Family Composition and the Income Maintenance Experiments," in A. Munnell. ed., *Lessons from the Income Maintenance Experiments*, Boston: Federal Reserve Bank of Boston.

Campbell, D. and J. Stanley (1963), *Experimental and Quasi-Experimental Designs for Research*, Boston: Houghton Mifflin.

Card, D. and A. Krueger (1994), "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review,* 84, 772–793.

Cassidy, R. and C. Manski (2019), "Tuberculosis Diagnosis and Treatment under Uncertainty," *Proceedings of the National Academy of Sciences*, 116, 22990-22997.

Deaton, A. (2009), " Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development," National Bureau of Economic Research Working Paper 14690.

DerSimonian, R., and N. Laird (1986), "Meta-Analysis in Clinical Trials." *Controlled Clinical Trials*, 7, 177-188.

Dominitz, J. and C. Manski (2025), "Comprehensive OOS Evaluation of Predictive Algorithms with Statistical Decision Theory," *Quantitative Economics*, forthcoming.

Ellsberg, D. (1961), "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, 75, 643-69.

Fleiss, J. (1981), *Statistical Methods for Rates and Proportions,* New York: Wiley.

Haavelmo, T. (1944), "The Probability Approach in Econometrics," *Econometrica,* 12, Supplement, iii-vi and 1-115.

Hausman, J. and D. Wise (1979), "Attrition Bias in Experimental and Panel Data: The Gary Income Maintenance Experiment," *Econometrica*, 47, 455-473.

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

Heckman, J. and S. Urzua (2009), "Comparing IV with Structural Models: What Simple IV Can and Cannot Identify," National Bureau of Economic Research Working Paper 14706.

Higbee, S. (2023), "Policy Learning with New Treatments," https://samuelhigbee.github.io/assets/papers/policy_learning_new.pdf

Horowitz, J., and C. Manski (2000), "Nonparametric Analysis of Randomized Experiments with Missing Attribute and Outcome Data," *Journal of the American Statistical Association*, 95, 77-84.

Imbens, G. and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-476.

LaLonde, R. (1986), "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 76, 604-620.

Li, S., V. Litvin, and C. Manski (2023), "Partial Identification of Personalized Treatment Response with Trial-reported Analyses of Binary Subgroups," *Epidemiology*, 34, 319-324.

Manski, C. (1990), "Nonparametric Bounds on Treatment Effects," *American Economic Review Papers and Proceedings*, 80, 319-323.

Manski, C. (1995), *Identification Problems in the Social Sciences*, Cambridge, MA: Harvard University Press.

Manski, C. (1996), "Learning about Treatment Effects from Experiments with Random Assignment of Treatments," *Journal of Human Resources,* 31, 707-733.

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

Manski, C. (1997b), "Monotone Treatment Response," *Econometrica*, 65, 1311-1334.

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 (2003), *Partial Identification of Probability Distributions,* Springer-Verlag: New York.

Manski, C. (2004), "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 72, 221-246.

Manski, C. (2007), *Identification for Prediction and Decision*, Cambridge, MA: Harvard University Press.

Manski, C. (2010), "Vaccination with Partial Knowledge of External Effectiveness," *Proceedings of the National Academy of Sciences*, 107, 3953-3960.

Manski, C. (2011), "Policy Analysis with Incredible Certitude," *The Economic Journal*, 121, F261-F289.

Manski, C. (2013), "Diagnostic Testing and Treatment under Ambiguity: Using Decision Analysis to Inform Clinical Practice," *Proceedings of the National Academy of Sciences*, 110, 2064-2069.

Manski, C. (2014), "Identification of Income-Leisure Preferences and Evaluation of Income Tax Policy," *Quantitative Economics*, 5, 145-174.

Manski, C. (2015), “Randomizing Regulatory Approval for Diversification and Deterrence," *Journal of Legal Studies*, 44, S367-S385.

Manski, C. (2017), “Mandating Vaccination with Unknown Indirect Effects,” *Journal of Public Economics Theory*, 19, 603-619.

Manski, C. (2018), “Credible Ecological Inference for Medical Decisions with Personalized Risk Assessment,” *Quantitative Economics*, 9, 541-569.

Manski, C. (2020), “The Lure of Incredible Certitude,” *Economics and Philosophy*, 36, 216-245.

Manski, C. (2021), “Econometrics for Decision Making: Building Foundations Sketched by Haavelmo and Wald,” *Econometrica*, 89, 2827-2853.

Manski, C. (2024), *Discourse on Social Planning under Uncertainty*, Cambridge: Cambridge University Press.

Manski, C. (2025), “Using Limited Trial Evidence to Choose Treatment Dosage when Efficacy and Toxicity Weakly Increase with Dose,” *Epidemiology*, 36, 60-65.

Manski, C. and I. Garfinkel, editors (1992), *Evaluating Welfare and Training Programs*, Cambridge, Mass.: Harvard University Press.

Manski, C., J. Mullahy, and A. Venkataramani (2023), “Using Measures of Race to Make Clinical Predictions: Decision Making, Patient Health, and Fairness,” *Proceedings of the National Academy of Sciences*, 120, https://www.pnas.org/doi/10.1073/pnas.2303370120

Manski, C., and D. Nagin (1998), "Bounding Disagreements about Treatment Effects: A Case Study of Sentencing and Recidivism," *Sociological Methodology,* 28, 99-137.

Manski, C. and J. Pepper (2000), "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 997-1010.

Phelps, C. and A. Mushlin (1988), "Focusing Technology Assessment Using Medical Decision Theory," *Medical Decision Making* 8, 279-289.

Thistlethwaite, D. and D. Campbell (1960), "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment," *Journal of Educational Psychology,* 51, 309-317.

Tukey, J. (1962), "The Future of Data Analysis," *Annals of Mathematical Statistics,* 33, 1-67.