Diagnostic testing and treatment under ambiguity: Using decision analysis to inform clinical practice

Charles F. Manski[1]

Department of Economics and Institute for Policy Research, Northwestern University, Evanston, IL 60208

Partial knowledge of patient health status and treatment response is a pervasive concern in medical decision making. Clinical practice guidelines (CPGs) make recommendations intended to optimize patient care, but optimization typically is infeasible with partial knowledge. Decision analysis shows that a clinician's objective, knowledge, and decision criterion should jointly determine the care he prescribes. To demonstrate, this paper studies a common scenario regarding diagnostic testing and treatment. A patient presents to a clinician, who obtains initial evidence on health status. The clinician can prescribe a treatment immediately or he can order a test yielding further evidence that may be useful in predicting treatment response. In the latter case, he prescribes a treatment after observation of the test result. I analyze this scenario in three steps. The first poses a welfare function and characterizes optimal care. The second describes partial knowledge of response to testing and treatment that might realistically be available. The third considers decision criteria. I conclude with reconsideration of clinical practice guidelines.

Partial knowledge of patient health status and treatment response is a pervasive concern in medical decision making. To improve decision making, modern societies invest in research and in translational activities seeking to make the research useful to clinical practice. Translational activities often become embodied in clinical practice guidelines (CPGs). A recent report (1) by the Institute of Medicine (IOM) gave this definition for CPGs (ref. 1, p. 4): "Clinical practice guidelines are statements that include recommendations intended to optimize patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options."

Clinical Practice Guidelines. Medical research makes much use of biological science, technology, and quantitative inferential methods. However, translational activities such as CPGs typically are informal and qualitative. The recent IOM report is illustrative. The report repeatedly calls for the development of rigorous CPGs. However, the eight standards proposed by the IOM committee (ref. 1, pp. 6–9) are, to this reader, uncomfortably vague.

Although the IOM report aims to inform medical decision making, it brings to bear no formal decision analysis. It discusses decision analysis only briefly, stating the following (ref. 1, p. 171): "A frontier of evidence-based medicine is decision analytic modeling in health care alternatives' assessment. . . . . Although the field is currently fraught with controversy, the committee acknowledges it as exciting and potentially promising, however, decided the state of the art is not ready for direct comment."

This statement is surprising. The foundations of decision analysis were largely in place more than 50 y ago and applications have since become common. Applications within medicine have been promoted for over 30 y by the Society for Medical Decision Making. However, the IOM report refers to decision analysis as "a frontier of evidence-based medicine" in which "the state of the art is not ready for direct comment."

I consider use of decision analysis to be a prerequisite for development of the rigorous CPGs cited as the objective of the IOM report. A serious shortcoming of the report is its supposition that CPGs make recommendations intended to "optimize patient care." Optimization of care may be infeasible when one has only partial knowledge of patient health status and treatment response. Then various care options may be reasonable, each using the available knowledge in its own way. (I semantically distinguish between evidence and knowledge. Evidence is synonymous with data. Knowledge is the set of conclusions that one draws by combining evidence with assumptions about unobserved quantities.)

If the IOM report were to embrace decision analysis, it would observe that rigorous decision making regarding patient care requires one to specify explicitly (*i*) the objective one wants to achieve, (*ii*) the knowledge one has of patient health status and treatment response, and (*iii*) the decision criterion one uses when partial knowledge makes optimization infeasible. Moreover, the report would observe that medical research speaks only to the second of these factors. Research may help a clinician measure health status and predict treatment response, but it cannot tell the clinician what objective he should want to achieve and what decision criterion he should use when optimization is infeasible.

Analysis of Testing and Treatment Decisions. To demonstrate the conduct of decision analysis, this paper studies a class of medical decisions that are straightforward to describe yet subtle to resolve. I consider a common scenario regarding diagnostic testing and treatment. A patient presents to a clinician, who obtains initial evidence on health status. The clinician either prescribes a treatment immediately or orders a test that may yield further evidence on health status. In the latter case, he prescribes a treatment after observation of the test result. (My analysis applies not only to diagnostic testing of symptomatic patients but also to medical screening of unsymptomatic ones. I use the term testing rather than screening throughout.)

For example, clinicians often decide between aggressive treatment of an illness and active surveillance (a.k.a. watchful waiting). Before prescribing treatment, they may order a diagnostic test. A common practice is to choose aggressive treatment if the test result is positive and active surveillance if it is negative or if the patient is not tested. I call this practice aggressive treatment with positive testing (ATPT) and use it to illustrate general ideas.

Given a specified objective for patient care and sufficient knowledge of response to testing and treatment, one can optimize care. Medical decision analysis of this type appears to have originated with ref. 2. This study assumed that the clinician knows the objective probability distributions of test results and of patient outcomes under alternative treatments, conditional on observed covariates and test results. It also assumed that the objective is to maximize expected utility. In this context, the usefulness of testing is expressed by the expected value of information, defined succinctly in ref. 3 (p. 119) as "the change in expected utility with the collection of information."

My concern is with settings in which the clinician lacks the knowledge of objective probability distributions required to compute the expected value of information and is unable to credibly assert subjective distributions in their stead. Then the clinician faces a problem of decision making under ambiguity, where maximization of expected utility is infeasible. (Use of the term ambiguity to describe settings where a decision maker does not assert a subjective distribution on unknown decision-relevant quantities originates with ref. 4. The term uncertainty describes settings in which a decision maker places a subjective distribution on unknowns.) The plan that the clinician should adopt depends on his objective, knowledge, and decision criterion. Decision analysis does not prescribe one best plan. It rather shows how the preferred plan depends on the objective, knowledge, and decision criterion.

To flesh out these ideas, *Optimal Testing and Treatment* studies optimal testing and treatment when a clinician has knowledge enabling optimization. *Identification of Response to Testing and Treatment When ATPT Is Standard Practice* explains why this knowledge may not be available in practice. To exemplify the broad issue, I study identification of response to testing and treatment from observation of a study population where ATPT was the standard clinical practice. *Patient Care Under Ambiguity* considers how a clinician might make decisions with partial knowledge. *Developing Clinical Guidelines Under Ambiguity* returns to the development of CPGs, drawing implications from the analysis of the previous three sections.

Background for the Analysis. The joint analysis of testing and treatment in this paper builds on my previous research that studied treatment decisions without considering the possibility of ordering a test before treatment (refs. 5–8 inter alia). The decision problems that I have studied share a simple structure. A decision maker, termed a planner, must choose a treatment for each member of a population. The planner observes some covariates for each person. Treatment response may vary with these covariates and also across persons having the same covariates. In medical applications, the planner may be a clinician treating patients. The covariates may be demographic attributes, medical histories, health status observed in office examinations, and the results of diagnostic tests.

I have supposed that the objective is to maximize the mean of some outcome of interest across the treated population. The optimization problem has an elementary solution. The planner should divide the population into groups having the same covariates. Within each group, he should assign everyone to the treatment with the highest within-group mean outcome. Thus, persons with different covariates may receive different treatments, but persons in the same group should be treated uniformly. Differential treatment of persons with different observed covariates is called profiling or personalized treatment.

Ideally, the planner might learn treatment response by performing a randomized trial or an observational study where treatment selection emulates a trial. However, statistical imprecision and identification problems limit knowledge in practice. Small sample sizes limit the precision of inference. Identification problems are the inferential difficulties that persist even when sample size grows without bound.

Identification problems often are the dominant difficulty. The unobservability of counterfactual treatment outcomes creates a fundamental identification problem when attempting to draw conclusions from observational studies, where treatment selection may be related to treatment response. Identification problems also complicate inference from randomized trials, which typically do not attain the ideal that persons have in mind when they refer to them as the "gold standard" for research.

The first task of decision analysis is to characterize the knowledge possessed by a planner who observes specified evidence and combines it with specified assumptions. For this purpose, I have

brought to bear my research on partial identification of treatment response, exposited in ref. 6. Partial identification means that the available evidence and maintained assumptions imply a bound on treatment response but not a precise value. (Formally, a quantity is partially identified if the evidence and assumptions imply that the quantity lies in some informative set of values but do not enable one to determine its precise value. The set of feasible values is called the identification region.) I have found that partial identification is the norm when realistically available evidence is combined with credible assumptions.

The second task is to study treatment choice with partial knowledge. Consider a planner choosing between two treatments. Partial identification may prevent the planner from knowing which treatment is better. Specifically, he may know only a bound on the treatment effect that covers zero. There is no optimal treatment choice in this setting, but decision theory suggests various decision criteria that one might view as reasonable. I have found it illuminating to study the properties of expected utility maximization, the maximin criterion, and the minimax-regret criterion. See refs. 7 and 8 for applications to medical decision making. [As far as I am aware, research on medical decision making has otherwise not explicitly studied decision making under ambiguity. However, scattered studies discuss some related issues. Meltzer (3) discusses bounding the value of information in the absence of subjective probability distributions.]

Whereas the analysis in this paper focuses on ambiguity induced by identification problems, clinicians may also face ambiguity stemming from imprecision in drawing inferences from samples of study populations. See refs. 9–11 for studies of treatment choice under this type of ambiguity.

Optimal Testing and Treatment

To substantiate the idea of optimal patient care, we need to specify a decision maker, a set of feasible actions, and a welfare function that embodies the objective of the decision maker. This done, we can ask what action maximizes welfare and what knowledge the decision maker needs to solve the optimization problem.

Caring for a Population of Patients poses a problem of optimal testing and treatment by a clinician who cares for a population of patients. *Optimal Testing and Treatment* studies its solution. As we proceed, I point out simplifying assumptions made to illuminate central issues and provide a tractable foundation for the analysis ahead in *Identification of Response to Testing and Treatment when ATPT is Standard Practice*. However one views the realism of these assumptions, an important virtue of decision analysis is that it forces one to make explicit assumptions and derive logically valid conclusions.

Caring for a Population of Patients. Discussions of health care often suppose that a clinician should optimize care for each patient in isolation, without reference to his care of other patients. However, it is feasible to optimize care for a single patient only if the clinician knows enough about individual treatment response to be certain what treatment is best for this patient. Clinicians typically lack this knowledge, particularly when considering whether to order a diagnostic test. After all, the medical purpose of a diagnostic test is to provide evidence on health status that may be useful in choosing a treatment. If the clinician were to already know what treatment is best, there would be no medical reason to contemplate a test. (I write that a clinician has no medical reason to order a test if he already knows what treatment is best. The adjective "medical" is necessary because a clinician may have other reasons. A health insurance plan may require the test as a condition for reimbursement of the cost of treatment. A legal reason is that a clinician may want to use a test to exhibit due diligence before choosing a treatment. These considerations may play roles in clinical practice, but I abstract from them.)

Rather than consider each patient in isolation, I suppose that the objective of the clinician is to optimize care on average across the patients in his practice. Optimization in this sense does not require that the clinician be certain what treatment is best for each patient. It only requires knowledge of mean treatment response within groups of patients having the same observed covariates.

Basic notation for testing and treatment. To formalize the decision problem, I consider a clinician who cares for the population of patients who present to him. I consider this patient population to be predetermined and I assume that patients always comply with the clinician's decisions. (In actuality, the patients who present to a clinician may depend on his testing and treatment policies. If patients can choose among alternative clinicians, they may do so in part on the basis of their testing and treatment policies. I ignore this possibility, which substantially complicates analysis.)

When a patient presents, the clinician initially observes covariates that may include demographic attributes, medical history, indicators of health status, and patient statements of preferences regarding care and outcomes. The clinician can prescribe a treatment immediately or order a test that may yield further evidence. In the latter case, the clinician prescribes a treatment after observation of the test result.

The notation x denotes the initially observed covariates of a patient and t denotes a treatment. I suppose for simplicity that there are two feasible treatments, t = A and t = B. I use s to indicate whether the clinician orders the diagnostic test, with s = 1 if he orders the test and s = 0 if he does not. I use r to denote the test result and suppose that r can take one of the two values p (positive) or n (negative). Medical professionals commonly call a test result "positive" if it indicates illness and "negative" otherwise.

With this notation, the actions that the clinician may choose and the knowledge of patient covariates accompanying each action may be expressed as a decision tree. The clinician chooses s = 0 or s = 1 with knowledge of x. If he chooses s = 0, he chooses t = A or t = B with knowledge of x. If he chooses s = 1, he chooses t = A or t = B with knowledge of (x, r).

Feasible testing and treatment allocations. When the clinician makes testing decisions, patients with the same value of x are observationally identical, whereas those with distinct values are observationally distinct. Hence, the clinician can use x to profile, making systematically different testing decisions for groups of patients with different values of x. The clinician cannot profile within the group of patients having the same value of x but he can randomly differentiate within the group, ordering testing for some fraction and not testing the remainder.

To formalize these ideas, let $\delta_S(x)$ be the fraction of the patients with covariates x who are tested and $1 - \delta_S(x)$ be the fraction not tested. The clinician can choose $\delta_S(x)$ to be any fraction in the interval [0, 1]. This done, he tests a randomly drawn fraction $\delta_S(x)$ of the patient group and does not test the remainder. I suppose for simplicity that the patient group is large. Then randomization implies that the tested and untested subgroups have approximately the same distribution of response to testing and treatment.

Applying similar reasoning, the clinician can profile treatment across groups of patients with different observed covariates and randomly differentiate treatment among patients with the same observed covariates. When considering treatment, we need to distinguish three types of patients. Patients who are not tested have observed covariates x when they are treated. Patients who are tested have observed covariates (x, r) when treated, where r equals n or p. Among patients who are not tested, let $\delta_{T0}(x)$ be the fraction of the patients with covariates x who receive treatment B and $1 - \delta_{T0}(x)$ be the fraction who receive treatment A. Among those who are tested, let $\delta_{T1}(x, r)$ be the fraction of the patients with covariates (x, r) who receive B and $1 - \delta_{T1}(x, r)$ be the fraction who receive A.

Welfare function. Having specified the feasible actions, it remains to specify the welfare function. The definition of CPGs in the IOM report (1) stated that recommendations intended to optimize patient care should be informed by "an assessment of the benefits and harms of alternative care options" (ref. 1, p. 4). I assume that the clinician aggregates the benefits and harms of making a particular testing and treatment decision for a given patient into a scalar welfare measure denoted y. Thus, y(s, t) summarizes the clinician's overall assessment of the benefits and harms that would occur if he were to make testing decision s and treatment decision t. The welfare measure may take into account not only health outcomes but also patient preferences and financial costs. Patients may respond heterogeneously, so y(s, t) may vary across patients.

Mean welfare across the population of patients is determined by the fraction of those in each covariate group that the clinician assigns to each testing–treatment option. Suppose that x lies in a finite set X of possible covariate values. For each $x \in X$, let P(x) denote the fraction of patients with covariate value x. For $r \in \{p, n\}$, let f(r|x) denote the fraction of patients with covariates x who would have test result r if they were to be tested.

For each possible value of (s, t), let E[y(s, t) |x] be the mean welfare that would result if all patients with covariates x were to receive (s, t). Let E[y(s, t) |x, r] be the mean welfare that would result if all patients with covariates x and test result r were to receive (s, t). Let $\delta = [\delta_S(x), \delta_{T0}(x), \delta_{T1}(x, r), x \in X, r \in \{p, n\}]$ denote any specified testing–treatment allocation. Then the mean welfare $W(\delta)$ that would result if the clinician were to choose allocation δ is obtained by averaging the various mean welfare values E[y(s, t) |x] and E[y(s, t) |x, r] across the groups who receive them. Thus,

$$W(\delta) = \sum_{x \in X} P(x) \big[[1 - \delta_S(x)][1 - \delta_{T0}(x)]E[y(0, A)|x] \\ + [1 - \delta_S(x)]\delta_{T0}(x)E[y(0, B)|x] \\ + \sum_{r \in \{p,n\}} f(r|x)\{\delta_S(x)[1 - \delta_{T1}(x, r)]E[y(1, A)|x, r] \\ + \delta_S(x)\delta_{T1}(x, r)E[y(1, B)|x, r]\} \big].$$ **[1]**

Optimal Testing and Treatment. An optimal testing and treatment allocation is any δ that maximizes $W(\delta)$. A derivation in *SI Text, section A* shows that an optimal allocation is

$$\delta_S(x) = 1 \text{ if } \sum_{r \in \{p,n\}} f(r|x)[\max\{E[y(1, A)|x, r], E[y(1, B)|x, r]\}] \\ \geq \max\{E[y(0, A)|x], E[y(0, B)|x]\}, = 0$$ **[2A]**

otherwise.

$$\delta_{T0}(x) = 1 \text{ if } E[y(0, B)|x] \geq E[y(0, A)|x], = 0$$

otherwise. **[2B]**

$$\delta_{T1}(x, p) = 1 \text{ if } E[y(1, B)|x, p] \geq E[y(1, A)|x, p], = 0$$

otherwise. **[2C]**

$$\delta_{T1}(x, n) = 1 \text{ if } E[y(1, B)|x, n] \geq E[y(1, A)|x, n], = 0$$

otherwise. **[2D]**

Each maximum is unique when the stated inequality is strict, whereas all allocations yield the same welfare when the values are equal.

The optimal treatment allocations given in Eqs. **2B–2D** are transparent. Whatever testing decisions the clinician may make, he should choose treatments that maximize mean welfare conditional on the observed covariates and test results.

The optimal testing allocation given in Eq. **2A** is more subtle. Suppose first that testing has no direct effect on the welfare of patients with covariates x. Thus, suppose that y (1, t) = y(0, t) for both treatments and all patients. Jensen's inequality implies that

$$\sum_{r\in\{p,n\}} f(r|x)[\max\{E[y(1,A)|x,r], E[y(1,B)|x,r]\}] \geq \max\{E[y(1,A)|x], E[y(1,B)|x]\}. \qquad [3]$$

Moreover,

$$y(1,T) = y(0,T) => \max\{E[y(1,A)|x], E[y(1,B)|x]\} = \max\{E[y(0,A)|x], E[y(0,B)|x]\}. \qquad [4]$$

Hence, it is optimal to test all patients with covariates x.

Why test if testing has no effect on the welfare of any patient? The explanation is that testing yields evidence on health status, which may be useful in decision making. Jensen's inequality shows that the mean welfare achieved by conditioning treatment on (x, r) is at least as high as that achieved by conditioning only on x. [Basu and Meltzer (12) use this implication of Jensen's inequality to recommend that clinicians elicit the preferences of their patients to individualize care. In their analysis, the test is a questionnaire eliciting patient preferences. Responding to the questions does not affect welfare directly.]

This conclusion continues to hold if testing has no negative effect on welfare, that is, if y(1, t) ≥ y(0, t) for all treatments and patients. Hence, a decision not to test can be optimal only if testing sometimes negatively affects welfare. This may happen if testing is invasive, is costly, or may harm patients by delaying treatment.

The subtlety is that testing may be optimal even if it always has a negative direct welfare effect, that is, even if y(1, t) < y(0, t) for all treatments and patients. The explanation again is that testing yields evidence on health status. Eq. **2A** makes precise the circumstances in which the information value of testing outweighs any direct negative effect that testing may have on welfare.

It remains to characterize when testing has positive information value for decision making. The form of Jensen's inequality given in Eq. **3** is only a weak inequality. Testing has positive information value for decision making when the inequality is strict. This occurs if and only if (*a*) the test result is uncertain in the absence of testing and (*b*) the optimal treatment decision varies with the test result.

Illustration: aggressive treatment vs. active surveillance. To illustrate, consider a choice between aggressive treatment of a possible illness (say t = B) and active surveillance (say t = A). The ATPT practice chooses aggressive treatment if the result of a test is positive and active surveillance if the result is negative or if the patient is not tested. Under the assumptions maintained in this section, ATPT is optimal if E[y(0, A)|x] > E[y(0, B)|x], E[y(1, A)|x, n] > E[y(1, B)|x, n], and E[y(1, B)|x, p] > E[y(1, A)|x, p].

What about the decision to test? The mean-response inequalities stated above imply that optimal treatment varies with the test result. Hence, testing has positive information value, a necessary condition for testing to be optimal. However, positive information value does not suffice to conclude that testing is optimal. It is optimal if the inequality in Eq. **2A** holds. Then the information value of testing is large enough to exceed any negative effect on welfare that testing may have.

Identification of Response to Testing and Treatment When ATPT Is Standard Practice

Determination of optimal treatments for untested and tested patients requires sufficient knowledge of E[y(0, t) |x] and E[y (1, t)|x, r], respectively, for t ∈ {A, B}, r ∈ {n, p}, and x ∈ X. One need not know the precise values of these quantities, but one must know

whether the inequalities in Eqs. **2B–2D** hold. Optimal testing also requires enough knowledge of the distribution f(r|x) of test results to conclude whether the inequality in Eq. **2A** holds.

In principle, one might obtain this knowledge by performing a randomized trial or an observational study that emulates a trial. An ideal trial with four arms, one for each value of (s, t), would reveal mean response and the distribution of test results. Murphy (13) studies estimation of optimal plans, using such data. However, an ideal trial often is infeasible. Available evidence may come from observational studies and/or imperfect trials. Hence, partial knowledge is common in practice.

A large body of research has analyzed identification of treatment response without considering the possibility of ordering testing before treatment. Here I provide new analysis of identification of response to testing and treatment. To study a reasonably realistic setting, I suppose that one observes a study population where ATPT was standard clinical practice. (I do not attempt to explain how ATPT became standard practice in the study population. In particular, I do not assume that ATPT solves the optimization problem in *Optimal Testing and Treatment*.) I suppose that no other evidence is available. For example, the clinician is unable to learn over time by observing the outcomes of his own testing and treatment decisions and drawing lessons for future patients.

I analyze the identification problem that arises from the unobservability of counterfactual testing and treatment outcomes. To focus attention on this core difficulty, I abstract from others that may arise in practice. I suppose that one observes the entire study population rather than just a sample. This enables study of identification per se, without having to also cope with statistical imprecision. I suppose that the study population has the same composition as the population to be treated. Hence, retrospective findings about response to testing and treatment in the study population imply the same prospective findings in the population to be treated. I also suppose that outcomes are bounded and, to simplify notation, measure welfare on the unit interval.

As in previous research, I find it illuminating to first analyze identification with relatively weak maintained assumptions and then consider stronger assumptions that may be credible in some applications. *Basic Analysis* assumes only that the population to be treated has the same composition as the study population. Combining the evidence with this assumption partially identifies some of the quantities relevant to optimization of testing and treatment but reveals nothing about others. *Random Testing*, *Test Result as a Monotone Instrumental Variable*, and *Monotone Response to Testing* add further assumptions that sharpen inference.

Basic Analysis. To optimize care, the clinician wants to learn enough about {E[y(0, t) |x], E[y(1, t) |x, r], f(r|x)} for t ∈ {A, B}, r ∈ {n, p}, x ∈ X to evaluate the inequalities in Eq. **2**. Thus, there are eight relevant quantities for each value of x. Given the basic assumption alone, observation of the outcomes yielded by the ATPT practice reveals nothing about three quantities and partially identifies the other five.

The evidence reveals nothing about E[y(0, B) |x], E[y(1, B) |x, n], and E[y(1, A) |x, p]. It is uninformative about E[y(0, B) |x] and E[y(1, B) |x, n] because ATPT prescribes treatment A for all patients who are untested or who are tested and have negative results. The evidence is uninformative about E[y(1, A) |x, p] because ATPT prescribes B for all tested patients with positive results.

When outcomes are bounded, the basic assumption and the evidence partially identify E[y(0, A) |x], E[y(1, A) |x, n], E[y(1, B) | x, p], f(r = p|x), and f(r = n|x). To analyze each quantity, we need to distinguish the members of the study population who were and were not tested. Thus, let z = 1 if a person was tested and z = 0 if he was not. The present basic analysis assumes nothing about the association of z with potential outcomes.

The identification region for each partially identified quantity is a bound, given below in Eqs. **5–8**. The proofs are collected

in *SI Text*, section B. In what follows, $g(n, x) \equiv f(r = n|x, z = 1)P(z = 1|x) + P(z = 0|x)$ and $g(p, x) \equiv f(r = p|x, z = 1)P(z = 1|x) + P(z = 0|x)$:

$$E[y(0, A)|x, z = 0]P(z = 0|x) \leq E[y(0, A)|x]$$
$$\leq E[y(0, A)|x, z = 0]P(z = 0|x) + P(z = 1|x). \quad [5]$$

$$E[y(1, A)|x, n, z = 1]f(r = n|x, z = 1)P(z = 1|x)/g(n, x)$$

$$\leq E[y(1, A)|x, n] \leq \{E[y(1, A)|x, n, z = 1]f(r = n|x, z = 1)P(z = 1|x)$$
$$+ P(z = 0|x)\}/g(n, x). \quad [6]$$

$$E[y(1, B)|x, p, z = 1]f(r = p|x, z = 1)P(z = 1|x)/g(p, x)$$

$$\leq E[y(1, B)|x, p] \leq \{E[y(1, B)|x, p, z = 1]f(r = p|x, z = 1)P(z = 1|x)$$
$$+ P(z = 0|x)\}/g(p, x). \quad [7]$$

$$f(r = n|x, z = 1)P(z = 1|x) \leq f(r = n|x)$$
$$\leq f(r = n|x, z = 1)P(z = 1|x) + P(z = 0|x). \quad [8A]$$

$$f(r = p|x, z = 1)P(z = 1|x) \leq f(r = p|x)$$
$$\leq f(r = p|x, z = 1)P(z = 1|x) + P(z = 0|x). \quad [8B]$$

All bounds are generally informative, each having a width that is less than one except in special cases.

Random Testing. Each of the five quantities that is partially identified in the basic analysis becomes point identified if testing is random conditional on x. Random testing may occur through performance of a randomized trial of testing, which is not prohibited by the ATPT practice. Or it may occur without an explicit trial if clinicians caring for the study population make testing decisions that are statistically independent of test results and of response to testing and treatment.

Random testing implies these equalities:

$$E[y(0, A)|x] = E[y(0, A)|x, z = 0]. \quad [9A]$$

$$E[y(1, A)|x, n] = E[y(1, A)|x, n, z = 1]. \quad [9B]$$

$$E[y(1, B)|x, p] = E[y(1, B)|x, p, z = 1]. \quad [9C]$$

$$f(r = n|x) = f(r = n|x, z = 1). \quad [9D]$$

$$f(r = p|x) = f(r = p|x, z = 1). \quad [9E]$$

The evidence reveals the right-hand side of Eq. **9A** if $P(z = 0|x) > 0$ and it reveals the right-hand sides of Eqs. **9B–9E** if $P(z = 1|x) > 0$. Hence, all of the quantities that were partially identified in the basic analysis are now point identified provided only that positive fractions of the study population are untested and tested.

The assumption of random testing does not help to identify $E[y(0, B)|x]$, $E[y(1, B)|x, n]$, and $E[y(1, A)|x, p]$. However, other assumptions are informative about these quantities. I consider two such assumptions below.

Test Result as a Monotone Instrumental Variable. Patients with negative results on a diagnostic test are often thought to be healthier than ones with positive results. Hence, a clinician may find it credible to predict that patients with negative test results have better future prospects, on average, than do patients with positive results. Consider, for example, use of a PET/computer-assisted tomography (CT) scan to detect metastasis of a cancer diagnosed at a primary site. A clinician may reasonably predict better prospects, on average, for patients with negative scans than for those with positive ones.

Formally, the clinician may find it credible to assume that the inequality

$$E[y(s, t)|x, n] \geq E[y(s, t)|x, p] \quad [10]$$

holds for one or more specified values of (s, t). This inequality asserts that the test result is a monotone instrumental variable (MIV), as defined in ref. 14.

Assertion of Eq. **10** for (s, t) = (1, A) yields an upper bound on $E[y(1, A)|x, p]$. This assumption alone implies that $E[y(1, A)|x, p]$ is no larger than the upper bound on $E[y(1, A)|x, n]$ given in Eq. **6**. If one also assumes random testing, then $E[y(1, A)|x, p]$ is no larger than the known value of $E[y(1, A)|x, n]$ given in Eq. **9B**.

Analogously, assertion of Eq. **10** for (s, t) = (1, B) yields a lower bound on $E[y(1, B)|x, n]$. This assumption alone implies that $E[y(1, B)|x, n]$ is no smaller than the lower bound on $E[y(1, B)|x, p]$ given in Eq. **7**. If one also assumes random testing, then $E[y(1, B)|x, n]$ is no smaller than the known value of $E[y(1, B)|x, p]$ given in Eq. **9C**.

Monotone Response to Testing. In some settings, a clinician may believe that testing cannot directly improve welfare but may decrease it. For example, he may think that testing has no therapeutic effect but may be invasive or costly. Formally, the clinician may find it credible to assume the inequality

$$y(0, t) \geq y(1, t) \quad [11]$$

for specified values of t and for all patients. This is a monotone-response assumption as defined in ref. 15.

Assertion of Eq. **11** for t = B yields a lower bound on $E[y(0, B)|x]$, the bound depending on what other assumptions are imposed. To begin, it follows from this assumption that

$$E[y(0, B)|x] \geq E[y(1, B)|x]. \quad [12]$$

In the absence of other assumptions, we may extend *Basic Analysis* to derive an informative lower bound on $E[y(1, B)|x]$ and, hence, a lower bound on $E[y(0, B)|x]$. The derivation is somewhat complex, so I omit it.

A simple finding emerges if one combines the monotone-response assumption with Eq. **10** for (s, t) = (1, B). The latter assumption implies that $E[y(1, B)|x] \geq E[y(1, B)|x, p]$. This and Eq. **12** imply that $E[y(0, B)|x] \geq E[y(1, B)|x, p]$. Hence, $E[y(0, B)|x]$ is no smaller than the lower bound on $E[y(1, B)|x, p]$ given in Eq. **7**. If one also assumes random testing, then $E[y(0, B)|x]$ is no smaller than the known value of $E[y(1, B)|x, p]$ given in Eq. **9C**.

Patient Care Under Ambiguity

Identification of Response to Testing and Treatment When ATPT Is Standard Practice demonstrated that identification problems occur regularly when studying response to testing and treatment. Combining available evidence with credible assumptions may yield informative bounds on response but not precise conclusions. Hence, identification problems make partial knowledge common in practice.

Clinicians must somehow use the knowledge they have to make decisions about patient care. This section discusses how decision theory can contribute to clinical practice. *States of Nature and Dominance* describes basic principles and *Decision Criteria* considers specific decision criteria.

States of Nature and Dominance. Suppose that a clinician wants to choose a testing–treatment allocation that maximizes the welfare function in Eq. **1**. However, he has only partial knowledge of the quantities that determine welfare. He knows that they jointly lie in some identification region, determined by the available evidence and the assumptions that he finds credible. Hence, the clinician does not know his welfare function. He knows only that it is one of a set of functions, each corresponding to a feasible value for the unknown quantities.

The first step in application of decision theory is to specify the set of feasible welfare functions. Let Γ index the feasible values for the quantities that determine welfare. Thus, let $\{E_\gamma[y(0, t) |x], E_\gamma[y (1, t) |x, r], f_\gamma(r|x), t \in \{A, B\}, r \in \{n, p\}, x \in X\}, \gamma \in \Gamma$ be the joint identification region for the vector. Then the feasible welfare functions are $[W(*, \gamma), \gamma \in \Gamma]$, where

$$
\begin{aligned}
W(\delta, \gamma) = \sum_{x \in X} P(x) \big[&[1 - \delta_S(x)][1 - \delta_{T0}(x)]E_\gamma[y(0, A)|x] \\
&+ [1 - \delta_S(x)]\delta_{T0}(x)E_\gamma[y(0, B)|x] \\
&+ \sum_{r \in \{p, n\}} f_\gamma(r|x)\big\{\delta_S(x)[1 - \delta_{T1}(x, r)]E_\gamma[y(1, A)|x, r] \\
&+ \delta_S(x)\delta_{T1}(x, r)E_\gamma[y(1, B)|x, r]\big\}\big].
\end{aligned}
$$

[13]

Eq. 13 gives all feasible versions of the welfare function in Eq. **1**, each version indexed by γ. In decision theory, γ is called a state of nature and Γ is called the state space.

A basic principle of decision theory is that a decision maker should not choose a dominated action. In the present context, δ is dominated if there exists another allocation that is certain to yield at least the same welfare and may yield more. Thus, δ is dominated if there exists a δ' such that $W(\delta, \gamma) \leq W(\delta', \gamma)$ for all $\gamma \in \Gamma$ and $W(\delta, \gamma) < W(\delta', \gamma)$ for some $\gamma \in \Gamma$. *SI Text,* section C illustrates this.

Decision Criteria. The decision-theoretic prescription that a clinician should not choose a dominated allocation is compelling. However, how should he choose among undominated allocations?

Let δ and δ' be two undominated allocations. Then either $[W(\delta, \gamma) = W(\delta', \gamma), \gamma \in \Gamma]$ or there exist $\gamma_1 \in \Gamma$ and $\gamma_2 \in \Gamma$ such that $W(\delta, \gamma_1) > W(\delta', \gamma_1)$ and $W(\delta, \gamma_2) < W(\delta', \gamma_2)$. In the former case, the clinician is indifferent between δ and δ'. In the latter case, he cannot order the two allocations: δ may yield a better or a worse outcome than δ'. Thus, the question "How should the clinician choose?" has no unambiguously correct answer.

The fact that there is no one correct choice among undominated allocations explains why, in the introductory section, I criticized the supposition in the 2011 IOM report that CPGs make

recommendations intended to optimize patient care. Optimization may be infeasible when a clinician has only partial knowledge.

When optimization is infeasible, clinicians must still choose patient care. Hence, it is important to ask what guidance decision theory offers on choice among undominated actions. Many perspectives have been expressed and various decision criteria studied. I describe some prominent criteria in *SI Text,* section D.

Developing Clinical Guidelines under Ambiguity

I observed in the introductory section that rigorous medical decision making requires specification of the objective, knowledge, and the decision criterion. I then studied a class of testing and treatment decisions. To conclude, I return to the development of CPGs. I think it is important to separate two tasks for CPGs. One is to characterize medical knowledge. The other is to make recommendations for patient care.

Accomplishment of the first task has substantial potential to improve clinical practice. Characterization of available knowledge should draw on all available evidence, experimental and observational. It should maintain assumptions that are sufficiently credible to be taken seriously. It should combine the evidence and assumptions to draw logically valid conclusions.

I am skeptical whether CPGs should undertake the second task. Making rigorous recommendations for patient care asks the developers of CPGs to aggregate the benefits and harms of care into a scalar measure of welfare. And it requires them to specify a decision criterion to cope with partial knowledge. These activities might be uncontroversial if there were consensus about how welfare should be measured and what decision criterion should be used. However, care recommendations may be contentious if perspectives vary across clinicians, patients, and other relevant parties. Then the recommendations made by the developers of CPGs may not embody the considerations that motivate actual care decisions.

An alternative to having CPGs make care recommendations would be to bring specialists in decision analysis into the clinical team. Modern clinical practice often has a group of professionals jointly contribute to patient care. Surgeons and internists may work together and in conjunction with nurses and technical personnel. However, existing patient-care teams do not ordinarily draw on professionals having specific expertise in the framing and analysis of complex decision problems. It may be that adding such professionals to clinical teams would be more beneficial to patient care than asking physicians to adhere to care recommendations made by distant organizations.

1. Institute of Medicine (2011) *Clinical Practice Guidelines We Can Trust, Committee on Standards for Developing Trustworthy Clinical Practice Guidelines, Board on Health Care Services,* eds Graham R, et al. (National Academy Press, Washington, DC).
2. Phelps CE, Mushlin AI (1988) Focusing technology assessment using medical decision theory. *Med Decis Making* 8(4):279–289.
3. Meltzer D (2001) Addressing uncertainty in medical cost-effectiveness: analysis implications of expected utility maximization for methods to perform sensitivity analysis and the use of cost-effectiveness analysis to set priorities for medical research. *J Health Econ* 20(1):109–129.
4. Ellsberg D (1961) Risk, ambiguity, and the Savage axioms. *Q J Econ* 75:643–669.
5. Manski C (2000) Identification problems and decisions under ambiguity: Empirical analysis of treatment response and normative analysis of treatment choice. *J Econom* 95:415–442.
6. Manski C (2007) *Identification for Prediction and Decision* (Harvard Univ Press, Cambridge, MA).
7. Manski C (2009) Diversified treatment under ambiguity. *Int Econ Rev* 50:1013–1041.
8. Manski CF (2010) Vaccination with partial knowledge of external effectiveness. *Proc Natl Acad Sci USA* 107(9):3953–3960.
9. Manski C (2004) Statistical treatment rules for heterogeneous populations. *Econometrica* 72:221–246.
10. Hirano K, Porter J (2009) Asymptotics for statistical treatment rules. *Econometrica* 77:1683–1701.
11. Stoye J (2012) New perspectives on statistical decisions under ambiguity. *Annu Rev Econ* 4:257–282.
12. Basu A, Meltzer D (2007) Value of information on preference heterogeneity and individualized care. *Med Decis Making* 27(2):112–127.
13. Murphy S (2003) Optimal dynamic treatment regimes. *J R Stat Soc B* 65:331–366.
14. Manski C, Pepper J (2000) Monotone instrumental variables: With an application to the returns to schooling. *Econometrica* 68:997–1010.
15. Manski C (1997) Monotone treatment response. *Econometrica* 65:1311–1334.

ECONOMIC SCIENCES

MEDICAL SCIENCES