CREDIBLE ECOLOGICAL INFERENCE FOR PERSONALIZED MEDICINE:
Formalizing Clinical Judgment

Charles F. Manski

Department of Economics and Institute for Policy Research
Northwestern University

October 2016

Abstract

This paper studies an identification problem that arises when clinicians seek to personalize patient care by making health risk assessments conditional on observed patient attributes. Let y be a patient outcome of interest and let (x = k, w = j) be patient attributes that a clinician observes. The clinician may want to choose a care option that maximizes the patient's expected utility conditional on the observed attributes. To accomplish this, the clinician needs to know the conditional probability distribution P(y|x = k, w = j). It is common to have a trustworthy risk assessment that predicts y conditional on a subset of the observed attributes, say x, but not conditional on (x, w). Then the clinician knows P(y|x = k) but not P(y|x = k, w = j). Partial conclusions about P(y | x = k, w = j) may be drawn if the clinician also knows P(w = j|x = k). Tighter conclusions may be possible if he combines knowledge of P(y|x) and P(w|x) with credible structural assumptions embodying some a priori knowledge of P(y|x, w). This is the *ecological inference* problem studied here. A substantial psychological literature comparing actuarial predictions and informal clinical judgments has concluded that clinicians should not attempt to subjectively predict patient outcomes conditional on attributes such as w that are not utilized in evidence-based risk assessments. The analysis in this paper suggests that formalizing clinical judgment through analysis of the inferential problem may enable clinicians to make more informative personalized risk assessments.

I have benefitted from the opportunity to present this work in a seminar at the Federal Reserve Bank of Cleveland and from the comments of Pamela Giustinelli and Max Tabord-Meehan.

1. Introduction

This paper studies an identification problem that arises when clinicians seek to personalize patient care by making health risk assessments or predictions of treatment response conditional on observed patient attributes. Let y be a patient outcome of interest. In risk assessments, y commonly is a binary variable indicating whether the patient will develop a specified disease or a real variable denoting remaining life span. When predicting treatment response, y is an outcome assuming patient receipt of a potential treatment. Let (x = k, w = j) be patient attributes that a clinician observes and wants to condition on when choosing a care option. These attributes may include demographic traits, documentation of medical history, and the results of screening and diagnostic tests. Suppose that the clinician wants to choose a care option that maximizes the patient's expected utility conditional on the observed attributes. To accomplish this, the clinician wants to know the *long* conditional probability distribution P(y|x = k, w = j).

In practice, it is common to have a trustworthy risk assessment that predicts y conditional on a subset of the observed patient attributes, say x, but not conditional on (x, w). Thus, the clinician may know the *short* conditional distribution P(y|x = k) but not P(y|x = k, w = j). For example, readily available life tables enable prediction of remaining life span conditional on basic demographic traits (age, sex, race) but not conditional on attributes characterizing a patient's health status. Available tools assessing the risk of developing specific diseases condition their predictions on some patient attributes that clinicians observe but not on others.

The Law of Total Probability gives the algebraic relationship between the short and long predictive distributions for the patient under care, namely

$$(1)\quad P(y|x = k) = P(w = j|x = k)P(y|x = k, w = j) + P(w \neq j|x = k)P(y|x = k, w \neq j).$$

Equation (1) shows that knowledge of P(y|x = k) per se reveals nothing about P(y|x = k, w = j). Any distribution P(y|x = k, w = j) satisfies the equation if P(w = j|x = k) = 0. On the other hand, partial

conclusions about P(y | x = k, w = j) may be drawn if one knows both P(y|x = k) and P(w = j|x = k), provided that the latter is positive. Tighter conclusions may be possible if one combines knowledge of P(y|x) and P(w|x) with credible structural assumptions; that is, assumptions embodying some a priori knowledge of P(y|x, w). This is the identification problem studied here.

Aspects of the problem have been studied in several literatures with varying substantive concerns and terminology, including those on ecological inference, contaminated sampling, and Simpson's paradox. Manski (2007, Chapter 5) gives a textbook exposition. The term *ecological inference* is not particularly evocative, but it is prominent and I use it here.

The analysis in the paper applies both to risk assessment and to prediction of treatment response. In both cases, the objective is to predict patient outcomes conditional on observed attributes. Whereas risk assessments explicitly or implicitly assume clinical use of some conventional patient care strategy, predictions of treatment response forecast the outcomes that would occur if patients were to receive alternative forms of care. For simplicity, I focus on risk assessment.

Section 2 provides background. I explain the term *personalized medicine* and give a prominent illustration of risk assessment that conditions on a subset of the patient attributes that a clinician typically observes, this being prediction of risk of breast cancer. I discuss the psychological literature comparing actuarial predictions and informal clinical judgments. This literature has concluded that clinicians should not attempt to subjectively predict patient outcomes conditional on attributes not utilized in evidence-based risk assessments. I suggest that formalizing clinical judgment through analysis of the inferential problem may enable clinicians to make more informative personalized risk assessments.

Section 3 presents the identification analysis, bringing to bear aspects of my previous research. I first summarize findings that hold when one has no information beyond knowledge of P(y|x) and P(w|x). Depending on the values of P(y|x) and P(w|x), the identification region for P(y|x = k, w = j) may be sufficiently small for a clinician to make useful outcome predictions conditional on (x, w), but it may be so large that predictions conditional on these attributes are uninformative.

Tighter predictions become possible if one imposes structural assumptions. Structural assumptions express judgment in a formal manner that contrasts with informal clinical judgment. Point identification of P(y|x, w) can be achieved if sufficiently strong structural assumptions are imposed. Research on ecological inference has considered various such assumptions. A central issue is how to resolve the tension between the strength and credibility of maintained assumptions. Strong structural assumptions may have substantial identifying power but little credibility. I show how to predict outcomes using weaker *bounded-variation* assumptions, defined in Section 3.3, that have less power but greater credibility.

Section 4 considers the use of risk assessment in medical decision making when the structural assumptions (if any) that a clinician deems credible are not strong enough to point identify P(y|x = k, w = j). In some situations the clinician may be able to make a decision that is sure to be optimal for patients with attributes (k, j), even though knowledge of P(y|x = k, w = j) is incomplete. In other situations the available information is too limited to enable optimization. Then patient care is a problem of decision making under ambiguity. I explain the general problem of patient care under ambiguity and study it in a relatively simple setting where the clinician chooses between active surveillance of a patient and prophylactic treatment. This part of the paper adds new findings to my program of work on medical decision making under ambiguity.

Although this paper reports simple new findings on partial identification and decision making under ambiguity in Sections 3 and 4, the primary contribution is not to develop new methodology. It is rather to show how simple applications and extensions of existing methodology may potentially be used to improve patient care, a substantive matter of considerable importance. Although research on medical decision making has long distinguished between actuarial prediction and informal clinical judgment, to the best of my knowledge it has not until now formally studied personalized risk assessment as an identification problem.

Readers who are not specifically concerned with the practice of personalized medicine may nonetheless find aspects of the paper of interest. Many decision makers in realms other than health care face identification problems similar to that studied here as they attempt to make personalized risk assessments.

2. Background

2.1. Personalized Medicine

The term *personalized medicine* is sometimes defined to mean health care that is literally specific to the individual, as in this definition by Ginsburg and Willard (2009, p. 278), which was adopted by American Medical Association (2010): "Personalized medicine is . . . . . health care that is informed by each person's unique clinical, genetic, genomic, and environmental information." However, evidence to support complete personalization is rarely available. Hence, the term is commonly used to mean care that varies with some individual characteristics. President's Council of Advisors on Science and Technology (2008, p. 7) states:

> " 'Personalized medicine' refers to the tailoring of medical treatment to the specific characteristics of each patient. In an operational sense, however, personalized medicine does not literally mean the creation of drugs or medical devices that are unique to a patient. Rather, it involves the ability to classify individuals into subpopulations that are uniquely or disproportionately susceptible to a particular disease or responsive to a specific treatment."

Similarly, Academy of Medical Sciences (2015, p. 4) states "The terms 'stratified', 'personalised' or 'precision' medicine all refer to the grouping of patients based on risk of disease, or response to therapy, using diagnostic tests or techniques."

Thus, personalized medicine is a matter of degree rather than an all-or-nothing proposition. Clinicians classify patients into groups based on information about medical history and findings obtained through screening and diagnostic tests. Clinical practice guidelines (CPGs) recommend classifications that aim to be well-grounded in evidence on risk of disease and treatment response. The classification rules used depend on the evidence on group outcomes that is available.

Normative studies of personalized medicine have commonly assumed that the clinician has evidence enabling accurate probabilistic risk assessments and predictions of treatment response conditional on observed patient attributes. For example, Phelps and Mushlin (1988) studied optimal diagnostic testing as a prelude

to treatment. They assumed that the clinician knows the actual probability distributions of test results and of patient outcomes under alternative treatments, conditional on observed patient attributes. They assumed that the objective is to maximize the patient's expected utility. In this context, the usefulness of performing tests or making other efforts to learn patient attributes is expressed by the *expected value of information*, defined succinctly by Meltzer (2001) as (p. 119) "the change in expected utility with the collection of information."

In practice, clinicians commonly observe patient attributes other than those used as predictors in evidence-based risk assessments and studies of treatment response. They often use informal clinical judgment to predict how patient outcomes vary with these attributes. There has long been concern that exercise of clinical judgment may reduce rather than improve the quality of medical decision making (e.g., Dawes, Faust, and Meehl, 1989).

Section 2.2 gives a prominent illustration of evidence-based risk assessment that utilizes only a subset of the patient attributes that clinicians typically observe. I next discuss the longstanding concern with use of informal clinical judgment to predict patient outcomes (Section 2.3). I then introduce the analysis undertaken here (Section 2.4).

2.2. Illustration: Predicting Risk of Breast Cancer

An apt illustration of how available evidence affects risk assessment is the Breast Cancer Risk Assessment (BCRA) Tool of the National Cancer Institute (2016). The risk assessment yielded by this tool has become widely used in clinical practice (see Susan G. Komen, 2016) and is an important input to the CPG for breast cancer screening issued by the National Comprehensive Cancer Network (2016).

The BCRA Tool gives a predicted probability that a woman will develop invasive breast cancer conditional on eight personal attributes: (1) history of breast cancer or chest radiation therapy for Hodgkin Lymphoma (yes/no); (2) presence of a BRCA mutation or diagnosis of a genetic syndrome associated with

risk of breast cancer (yes/no/unknown); (3) current age, in years; (4) age of first menstrual period (7-11, 12-13, ≥ 14, unknown); (5) age of first live birth of a child (no births, < 20, 20-24, 25-29, ≥30, unknown); (6) number of first-degree female relatives with breast cancer (0, 1, >1, unknown); (7) number of breast biopsies (0, 1, > 1, unknown); and (8) race/ethnicity (White, African American, Hispanic, Asian American, American Indian or Alaskan Native, unknown).

In terms of the notation introduced in Section 1, y denotes whether a patient will develop invasive breast cancer, x are the attributes that the BCRA tool uses to predict y, and w are additional patient attributes that the clinician observes but that are not used by the tool. The BRCA tool yields an evidence-based estimate of the probability $P(y = 1|x)$ that a woman with attributes x will develop invasive cancer. The reason that the tool assesses risk conditional on eight attributes and not others is that the tool uses a modified version of the "Gail Model," based on the empirical research of Gail *et al.* (1989). The Gail *et al.* article reported predicted probabilities of breast cancer for white women who have annual breast examinations, conditional on attributes (1) through (7). Scientists at the National Cancer Institute later modified the model to predict invasive cancer within a wider population of women.

The BCRA Tool personalizes predicted risk of breast cancer in multiple respects, but it does not condition on further personal attributes that a clinician can observe and that may be associated with risk of cancer. For example, when considering the number of first-degree relatives with breast cancer (item 6), the BCRA Tool does not take into account the number and ages of a woman's first-degree relatives, which logically should matter when interpreting the response to the item. Nor does it condition on the prevalence of breast cancer among second-degree relatives, a consideration that figures prominently in another risk assessment model due to Claus, Risch, and Thompson (1994). When considering race/ethnicity (item 8), the BCRA Tool groups all white woman together and does not distinguish ethnic subgroups such as Ashkenazi Jews, who are thought to have considerably higher risk of a BRCA mutation than other white subgroups, a potentially important matter when the answer to item (2) is "unknown." Moreover, the BCRA Tool does not condition on behavioral attributes such as excessive drinking of alcohol, which has been associated with

substantial increased risk of breast cancer (Singletary and Gapstur, 2001).

2.3. Evidence-Based Care and Informal Clinical Judgment

The BCRA Tool exemplifies a common question in patient care. Evidence from medical research enables one to assess risk of disease conditional on certain patient attributes. A clinician observes these attributes and also observes additional attributes that may be informative predictors of future disease. However, the available evidence does not show how patient outcomes vary with these additional attributes. How should the clinician assess risk?

A similar question arises when choosing treatments. Research articles reporting on clinical trials or epidemiological studies provide evidence on treatment response for groups of patients who share certain attributes. Clinicians commonly observe not only these attributes but also others that may predict treatment response. In the absence of evidence of how treatment response varies with these additional attributes, how should clinicians make treatment decisions?

One option is to ignore the additional attributes w and base care only on the attributes x used in available evidence-based assessment tools or reports on treatment response. Thus, a clinician recommending a breast cancer screening plan to a woman might compute her BCRA risk assessment, disregarding attributes that he observes but that the BCRA Tool does not utilize. Another option is to make predictions conditional on (x, w). Such predictions are typically made informally, by a process called clinical judgment.

A substantial body of empirical psychological research comparing evidence-based statistical predictions with ones made by clinical judgment has concluded that the former consistently outperforms the latter when the predictions are made using the same patient attributes. Moreover, the gap in performance appears to persist even when clinical judgment uses additional attributes as predictors. This research began in mid-twentieth century, early contributions including Sarbin (1943, 1944) and Meehl (1954). To describe the conclusions of the literature, I rely mainly on the influential review article of Dawes, Faust, and Meehl

(1989). See also Camerer and Johnson (1997).

Dawes *et al.* distinguish actuarial prediction and clinical judgment as follows (p. 1668):

> "In the clinical method the decision-maker combines or processes information in her or her head. In the actuarial or statistical method the human judge is eliminated and conclusions rest solely on empirically established relations between data and the condition or event of interest."

Comparing the two in circumstances where a clinician observes patient attributes that are not utilized in available actuarial prediction, they state (p. 1670):

> "Might the clinician attain superiority if given an informational edge? For example, suppose the clinician lacks an actuarial formula for interpreting certain interview results and must choose between an impression based on both interview and test scores and a contrary actuarial interpretation based on only the test scores. The research addressing this question has yielded consistent results . . . . Even when given an information edge, the clinical judge still fails to surpass the actuarial method; in fact, access to additional information often does nothing to close the gap between the two methods."

Seeking to explain this empirical finding, they discuss an example in which the additional observed attribute is that a patient has a broken leg and then write (p. 1670-1671):

> "The broken leg possibility is easily studied by providing clinicians with both the available data and the actuarial conclusion and allowing them to use or countervail the latter at their discretion. The limited research examining this possibility, however, all shows that greater overall accuracy is achieved when clinicians rely uniformly on actuarial conclusions and avoid discretionary judgments . . . . When operating freely, clinicians apparently identify too many 'exceptions,' that is, the actuarial conclusions correctly modified are outnumbered by those incorrectly modified. If clinicians were more conservative in overriding actuarial conclusions they might gain an advantage, but this conjecture remains to be studied adequately."

Here and elsewhere, Dawes, Faust, and Meehl caution against use of clinical judgment to informally predict disease risk or treatment response conditional on patient attributes that are not utilized in evidence-based assessment tools or research reports. They attribute the weak performance of informal clinical judgment to clinician failure to adequately grasp the logic of the prediction problem and to their use of decision rules that place too much emphasis on exceptions such as broken legs.

2.4. Prediction Formalizing Clinical Judgment

Suppose that Dawes, Faust, and Meehl are correct to advise against use of informal clinical judgment to predict patient outcomes. This does not foreclose the possibility of making well-grounded predictions that combine evidence with judgment. The authors suggest this when they conjecture (p. 1671) that clinicians might gain an advantage if they were more conservative in overriding actuarial conclusions. They also conjecture (p. 1670) that theory-mediated judgments may potentially be superior to conclusions reached solely on the basis of empirical frequencies. However, they mention these ideas only briefly and do not propose specific approaches. Other authors have offered broad qualitative suggestions for integration of actuarial prediction and clinical judgment (e.g., Shlonsky and Wagner, 2005). Yet, as far as I am aware, psychologists have not studied the matter formally as a mathematical problem of conditional prediction.

This paper studies prediction of patient outcomes that coherently combines evidence and judgment. As shown in Section 1, knowledge of $P(y|x)$ by itself reveals nothing about $P(y|x = k, w = j)$. To draw conclusions about the long predictive distribution requires further information. Throughout the paper I suppose that the available other information includes at least knowledge of $P(w|x)$, the w-composition of the group of patients with attribute value x. One may also pose structural assumptions that restrict the form of $P(y|x, w)$.

Evidence on group composition is often available in practice, so analysis of its implications should be broadly useful to patient care. For example, in the breast cancer context, w might measure the number of first-degree relatives that a woman has or whether she is an excessive drinker of alcohol. In these cases, knowing $P(w|x)$ means that one knows the distribution of number of first-degree relatives or the prevalence of excessive drinking among women with attributes x. The availability of credible structural assumptions varies across settings. Illustrations are given in Section 3.

3. Risk Assessment Using Evidence and Structural Assumptions

I analyze personalized risk assessment in stages. Section 3.1 considers prediction of patient outcomes using evidence on group evidence and group composition, but without structural assumptions. Section 3.2 considers two types of strong structural assumptions, using instrumental variables and parametric models, that can point-identify the long predictive distribution. Section 3.3 studies weaker but potentially more credible bounded-variation assumptions.

3.1. Assessment with Evidence on Group Outcomes and Group Composition

As previously stated, let each member of a population of patients be characterized by a triple (y, x, w) and let P(y, x, w) denote the distribution of (y, x, w). Here y measures a patient outcome of interest, taking values in an outcome space Y. Patient attributes x, which take values in a finite space X, are used to predict y in an available evidence-based risk assessment tool. Attributes w, which take values in a finite set W, are not used in the risk assessment tool.

I assume that the evidence-based assessment tool is accurate, in the sense that it correctly reveals the short predictive distribution P(y|x). This assumption simplifies analysis and clinicians often maintain it in practice. Nevertheless, one should keep in mind that actual assessment tools may not be fully accurate. For example, the Gail Model underlying the BCRA Tool maintains various structural assumptions and was estimated using particular (outcome, attribute) data. The predictions made by the BCRA Tool may be suspect if the assumptions of the Gail model were not realistic, if the data used to estimate the model suffered from measurement problems, or if the predictive distribution that prevailed when the model was estimated does not accurately describe the risk of breast cancer today. The parameter estimates of the Gail model are also subject to ordinary finite-sample imprecision.

Consider a patient with attributes (x = k, w = j). Equation (1) gave the algebraic relationship between

the short and long predictive distributions for this patient. Abstractly, the joint identification region for $P(y|x = k, w = j)$ and $P(y|x = k, w \neq j)$ given knowledge of $P(y|x = k)$ and $P(w = j|x = k)$ is the set of pairs of long outcome distributions that satisfy (1). The equation holds if both long distributions equal the short distribution $P(y|x = k)$. This is the only feasible pair of equal long distributions. All other feasible pairs have $P(y|x = k, w = j) \neq P(y|x = k, w \neq j)$.

Social scientists have used the term *ecological inference* to describe the problem of identification of $P(y|x, w)$ given knowledge of $P(y|x)$ and $P(w|x)$. Important early contributions include Robinson (1950), Duncan and Davis (1953), and Goodman (1953). A prominent instance arises in analysis of the geographic and demographic variation in voting across the population. Surveys yielding information on individual attributes and voting behavior may not be available and, when they are, the credibility of self-reports of voting behavior may be suspect. Hence, social scientists have often sought to infer voting patterns from two data sources that are readily available and credible: (a) administrative records on voting by electoral district and (b) census data on the attributes of persons in each district.

Formally, let y denote the voting outcome, let x denote an electoral district, and let w denote personal attributes thought to be associated with voting behavior. The objective is to learn $P(y|x, w)$, the distribution of voting outcomes among persons in district x with attributes w. Voting records may reveal $P(y|x)$ and census data may reveal $P(w|x)$. The problem is to use knowledge of $P(y|x)$ and $P(w|x)$ to learn $P(y|x, w)$.

Inference on $P(y|x, w)$ given knowledge of $P(y|x)$ and $P(w|x)$ has also been studied in research on estimation with contaminated sampling, dating from Huber (1964). Here the object of interest is $P(y|x, w = j)$ for a specified value of j. Values of (y, x, w) with $w = j$ are said to be error-free and those with $w \neq j$ are said to be erroneous. The researcher only observes (y, x) pairs, not (y, x, w) triples, and so does not know which observations are error-free. The researcher is, however, assumed to know the conditional probability $P(w = j|x)$ that an observation is error-free, or at least to know a lower bound on this probability.

Whether the terminology be that of ecological inference or contaminated sampling, the mathematical problem is to characterize the restrictions on $P(y|x, w)$ implied by knowledge of $P(y|x)$ and $P(w|x)$, via the

Law of Total Probability (1). I first review the simple case in which y is a binary outcome and then the case of a general real-valued outcome.

3.1.1. Predicting Binary Outcomes

When outcome y is binary, the identification region for $P(y = 1|x = k, w = j)$ is an easily computed interval, namely

$$(2)\quad P(y = 1|x = k, w = j) \in [0, 1] \cap \left[\frac{P(y = 1|x = k) - P(w \neq j|x = k)}{P(w = j|x = k)}, \frac{P(y = 1|x = k)}{P(w = j|x = k)}\right].$$

This result was sketched by Duncan and Davis (1953) in their concise seminal study of ecological inference. They attributed it to the early statistician Yule. The first formal proof appears to be in Horowitz and Manski (1995, Corollary 1.2), in their study of identification under contaminated sampling.

Proof of (2) is easy. Solving the linear equation (1) for $P(y = 1 \,|\, x = k, w = j)$ yields

$$(3)\qquad P(y = 1 \,|\, x = k, w = j) = [P(y = 1|x = k) - P(w \neq j|x = k) \cdot P(y = 1 \,|\, x = k, w \neq j)]/P(w = j|x = k).$$

Letting $P(y = 1 \,|\, x = k, w \neq j)$ take all values in the interval [0, 1] yields a tentative identification region for $P(y = 1 \,|\, x = k, w = j)$, this being the interval

$$\left[\frac{P(y = 1|x = k) - P(w \neq j|x = k)}{P(w = j|x = k)}, \frac{P(y = 1|x = k)}{P(w = j|x = k)}\right].$$

However, not all values of $P(y = 1 \,|\, x = k, w \neq j)$ may be feasible; some values may make the lower bound of the above interval less than zero or the upper bound greater than one. Shrinking the interval to include only

proper probabilities yields the interval on the right-hand side of (2).

Observe that the lower bound in (2) is greater than zero, hence informative, when $P(y = 1|x = k) > P(w \neq j|x = k)$. The upper bound is less than one, hence informative, when $P(y = 1|x = k) < P(w = j|x = k)$. When both conditions hold, the interval has width $P(w \neq j|x = k)/P(w = j|x = k)$. A necessary but not sufficient condition for both bounds to be informative is that $P(w = j|x = k) > ½$.

Illustration: Consider application of the BCRA Tool to a woman with these attributes (x = k): (1) no history of breast cancer or chest radiation therapy; (2) unknown presence of a BRCA mutation; (3) 40 years old; (4) age of first menstrual period in interval 12-13; (5) age of first live birth in interval 20-24; (6) 0 first-degree female relatives with breast cancer; (7) 0 breast biopsies; (8) white race/ethnicity. The BCRA predicted lifetime risk that such a woman will develop invasive breast cancer is $P(y = 1|x = k) = 0.090$.

Suppose that the clinician asks the woman about her alcohol consumption, specifically whether or not she is a heavy drinker, defined as drinking five or more drinks on the same occasion on each of five or more days in the past thirty days. Let w = 1 if the patient is a heavy drinker and w = 0 otherwise. Data collected in the 2014 National Survey on Drug Use and Health (NSDUH) shows that the fraction of adult women who are heavy drinkers by this definition is 0.034 (Substance Abuse and Mental Health Services Administration, 2014). Thus, $P(w = 1) = 0.034$.

Suppose the clinician assumes that $P(w = 1|x = k) = P(w = 1)$. Given the specified values for P(y|x) and P(w|x), (1) yields these findings for P(y|x, w): $P(y = 1|x = k, w = 0) \in [0.058, 0.093]$ and $P(y = 1|x = k, w = 1) \in [0, 1]$. Thus, combining evidence on group outcomes and on group composition yields a tight bound on $P(y = 1|x = k, w = 0)$ but reveals nothing about $P(y = 1|x = k, w = 1)$. The source of this extreme difference in informativeness is that the fraction of women who are heavy drinkers is so small (0.034). □

3.1.2. Predicting Mean and Quantile Outcomes

When y is a real-valued outcome taking more than two values, there is no characterization of the

identification region for P(y|x, w) of simplicity comparable to (2). However, Horowitz and Manski (1995) derive relatively simple expressions for the identification regions of the mean and quantiles of P(y|x, w). Thus, consider the long conditional mean $E(y|x = k, w = j)$ or the α-quantile $Q_\alpha(y|x = k, w = j)$, where $\alpha \in (0, 1)$. Identification of these parameters can be studied directly, but it is easier to prove a general result for the class of parameters that respect stochastic dominance and then apply this result to the mean and quantiles.

To simplify notation, let $p \equiv P(w \neq j|x = k)$. It can be shown that the identification region for $P(y|x = k, w = j)$ contains a "smallest" member L that is stochastically dominated by all feasible values of $P(y|x = k, w = j)$ and a "largest" member U that stochastically dominates all feasible values of $P(y|x = k, w = j)$. These distributions are truncated versions of P(y|x = k): L right-truncates P(y|x = k) at its $(1 - p)$–quantile and U left-truncates P(y|x = k) at its p–quantile. Formally, L and U are defined as follows:

$$\text{(4a)} \qquad L[-\infty, t] \equiv P(y \leq t|x = k)/(1 - p) \quad \text{for } t < Q_{(1-p)}(y|x = k),$$

$$\equiv 1 \quad \text{for } t \geq Q_{(1-p)}(y|x = k).$$

$$\text{(4b)} \qquad U[-\infty, t] \equiv 0 \quad \text{for } t < Q_p(y|x = k),$$

$$\equiv [P(y \leq t|x = k) - p]/(1 - p) \quad \text{for } t \geq Q_p(y|x = k).$$

With this background, it follows immediately that if $D(\cdot)$ is a parameter that respects stochastic dominance, the smallest feasible value of $D[P(y|x = k, w = j)]$ is D(L) and the largest feasible value is D(U). Hence, sharp lower and upper bounds on $E(y|x = k, w = j)$ are the means of L and U. Similarly, sharp bounds on $Q_\alpha(y|x = k, w = j)$ are the α–quantiles of L and U.

3.2. Risk Assessment with Strong Structural Assumptions

The above summarizes findings on inference on $P(y|x, w)$ using only knowledge of $P(y|x)$ and $P(w|x)$. Tighter inferences may be feasible with structural assumptions. The literature has developed two

approaches imposing assumptions strong enough to point-identify P(y|x, w). I review these here.

3.2.1. Instrumental Variables

Goodman (1953) took the objective to be inference on E(y|x, w) when y is real-valued. This objective is the same as inference on P(y|x, w) when y is binary, as E(y|x, w) = P(y = 1|x, w) in that case. He used x as an instrumental variable, asserting that y is mean-independent of x, conditional on w. That is,

$$(5) \qquad E(y|x = k, w = j) = E(y|w = j), \quad \text{all } (k, j) \in X \times W.$$

In the classical voting application studied by social scientists, (5) assumes that persons who have the same demographic attributes but who reside in different districts vote the same way, on average.

To begin, Goodman observed that the Law of Iterated Expectations gives

$$(6) \qquad E(y|x = k) = \sum_{j \in W} E(y|x = k, w = j)P(w = j|x = k), \quad k \in X.$$

For each $k \in X$, the data reveal $E(y|x = k)$ and $[P(w = j|x = k), j \in W]$, but not $[E(y|x = k, w = j), j \in W]$. So (6) is a system of $|X|$ linear equations with the $|X| \times |W|$ unknowns $E(y|x = k, w = j)$, $(k, j) \in X \times W$.

Now impose assumption (5). Then (6) becomes

$$(7) \qquad E(y|x = k) = \sum_{j \in W} E(y|w = j)P(w = j|x = k), \quad k \in X.$$

This is a system of $|X|$ linear equations with the $|W|$ unknowns $E(y|w = j)$, $j \in W$. Goodman observed that the equations have a unique solution if $|X| \geq |W|$ and if the $|X| \times |W|$ dimensional matrix $[P(w = j|x = k), (k, j) \in X \times W]$ has full rank $|W|$. Then assumption (5) point-identifies $E(y|w = j)$, $j \in W$.

Goodman also observed that assumption (5) is refutable. Equation system (7) may have no solution,

or its solution may lie outside the logical range of y. In both cases, it follows that (5) is incorrect.

Goodman's remarkably simple analysis fully resolves the identification problem when equation system (7) has one solution. It does not show how to use assumption (5) when (7) has multiple solutions, as is generically the case when $|X| < |W|$. Extending Goodman's analysis to cover this case, Cross and Manski (2002) characterize the identification region for $[E(y|w = j), j \in W]$.

Although Goodman (1953) demonstrated the identifying power of assumption (5), he did not advocate its regular use in practice. He cautioned that the assumption holds (p. 663) "in very special circumstances." Goodman had in mind applications to social science rather than to medicine. I find it difficult to conjecture instances in assessment of health risk where assumption (5) may be credible. In risk assessment, x are attributes used to predict outcomes by evidence-based assessment tools and w are clinician-observed attributes that are not used by these tools. If (5) holds, the attributes x used by assessment tools have no predictive power once one conditions prediction on the additional attributes w. It is mathematically possible for this to occur, but it seems unlikely to occur in practice.

3.2.2. Parametric Models

The second approach used to point-identify $P(y|x, w)$ is to assert a parametric model that places these conditional distributions in a finite-dimensional family. Thus, let Θ be a specified subset of L-dimensional real space, let $F(\cdot, \cdot, \cdot)$ be a specified function mapping $X \times W \times \Theta$ into probability distributions on the outcome space Y, and assume that there exists a $\theta \in \Theta$ such that

(8) $\quad P(y|x = k, w = j) = F(k, j, \theta), \quad$ all $(k, j) \in X \times W$.

Combining the Law of Total Probability with assumption (8) yields

$$(9) \qquad P(y|x = k) = \sum_{j \in W} F(k, j, \theta)P(w = j|x = k), \quad k \in X.$$

For each $k \in X$, the data reveal $P(y|x = k)$ and $[P(w = j|x = k), j \in W]$. Hence, (9) is a system of $|X|$ distributional equations restricting the L-dimensional parameter θ.

Analysis of distributional equations is difficult, but progress can be made by considering the implications for prediction of mean outcomes. Let $e(k, j, \theta)$ denote the mean of the random variable with distribution $F(k, j, \theta)$. Insertion of $e(k, j, \theta)$ into the Law of Iterated Expectations (6) yields

$$(10) \qquad E(y|x = k) = \sum_{j \in W} e(k, j, \theta)P(w = j|x = k), \quad k \in X.$$

The system (10) of equations is similar to Goodman's system (7) except that $e(k, j, \theta)$ generally varies nonlinearly with θ. Nonlinearity in θ implies that solution of (10) is more complex than was the case with Goodman's instrumental variable assumption. Nevertheless, the equations have a unique solution if $|X| \geq L$ and if sufficient regularity conditions hold. Then assumption (8) point-identifies $P(y|x, w)$.

There are innumerable alternative parametric models for $P(y|x, w)$ and, hence, innumerable potential implementations of this approach to inference. Twenty years ago, a particular model was proposed with enthusiasm by King (1997), who asserted that he had achieved "a solution to the ecological inference problem" in a book of that name. However, his assumptions immediately drew criticism, as evidenced in a dispute played out in the *Journal of the American Statistical Association* (Freedman, Klein, Ostland, and Roberts, 1998, 1999; King, 1999) and elsewhere (Cho, 1998, and Cho and Gaines, 2004). Wakefield (2008) cautions against application of the King model or other parametric models to public health research.

Illustration: A common problem in risk assessment is to predict a patient's remaining life span conditional on observed attributes. Let y denote remaining life span. Life tables provide actuarial predictions of mean life span conditional on age, race, and sex; see, for example, Centers for Disease Control and Prevention

(2015). However, existing life tables do not predict mean life span conditional on the patient attributes that clinicians commonly observe. The administrative data contained in death records may state a person's immediate "cause of death," but they do not document the person's medical history. Population surveys may provide richer personal data but most do not follow sample members until death.

Organizations developing clinical practice guidelines for breast cancer screening have had to acknowledge the absence of evidence-based predictions of life spans under alternative screening regimes. For example, in a review of knowledge of the benefits and harms of screening commissioned by the American Cancer Society, Myers *et al.* (2015) state (p. 1616):

> "We did not identify any direct evidence on the association between mammographic screening and life expectancy, which would require following up all participants in an RCT or cohort study until death from any cause. . . . . Because estimates of life expectancy gains from screening are by definition indirect and there is considerable uncertainty about the value of several parameters important for estimating these gains (in particular the magnitude of mortality reduction associated with screening at different ages and different intervals), we judged the quality of evidence for the magnitude of the association between screening and life expectancy to be LOW."

Myers *et al.* summarize model-based predictions of life expectancy reported by Mandelblatt *et al.* (2009). The latter authors describe six parametric models developed by different research teams. Each team had access to a shared database that characterized the life trajectories from age 25 to age 40 of American women born in 1960. The six teams made varying predictions of life expectancy after age 40 by combining these shared data with auxiliary data and with alternative structural assumptions strong enough to yield point identification. □

3.3. Bounded-Variation Assumptions

A clinician contemplating risk assessment conditioning on patient attributes not used in evidence-based assessment tools may be discomforted by the analysis in Sections 3.1 and 3.2. Without structural

assumptions, drawing informative conclusions about P(y|x = k, w= j) requires a relatively high prevalence of attribute w = j in the group with attributes x = k. Strong assumptions may point-identify P(y|x = k, w= j), but the conclusion drawn may have low credibility.

There is a substantial middle ground between the polar cases of no assumptions and assumptions strong enough to yield point identification. This section suggest a class of *bounded-variation* assumptions that clinicians should find easy to contemplate and apply. These assumptions flexibly restrict the magnitudes of risk assessments and the degree to which they vary with patient attributes, enabling clinicians to express quantitative judgments in a structured way. Bounded-variation assumptions have previously been used to provide identifying power in other settings. See Manski and Pepper (2000, 2013, 2016).

Section 3.3.1 analyzes some simple bounded-variation assumptions when outcome y is binary. Section 3.3.2 considers assumptions that place more general bounds on mean outcomes.

3.3.1. Binary Outcomes

Recall the derivation of identification region (2) for P(y = 1|x = k, w = j) in the absence of structural assumptions. Applying the Law of Total Probability (1) to P(y = 1|x = k) and solving this linear equation for P(y = 1 | x = k, w = j) yielded equation (3). Result (2) emerged by imposing only the logical constraints that P(y = 1 | x = k, w ≠ j) and P(y = 1 | x = k, w = j) both lie in the unit interval.

A clinician may find it credible to assume that these long conditional probabilities lie within specified bounds within the unit interval, say [a(k, ≠j), b(k, ≠j)] and [a(k, j), b(k, j)]. Thus, let the clinician assume that

$$(11a) \qquad a(k, \neq j) \leq P(y = 1 \,|\, x = k, w \neq j) \leq b(k, \neq j),$$

$$(11b) \qquad a(k, j) \leq P(y = 1 \,|\, x = k, w = j) \leq b(k, j).$$

Using these bounds in (3) yields a bounded-variation identification region for P(y = 1 | x = k, w = j), namely

(12) $P(y = 1 \mid x = k, w = j) \in$

$$[a(k, j), b(k, j)] \cap \left[\frac{P(y = 1|x = k) - b(k, \neq j) \cdot P(w \neq j|x = k)}{P(w = j|x = k)}, \frac{P(y = 1|x = k) - a(k, \neq j) \cdot P(w \neq j|x = k)}{P(w = j|x = k)} \right].$$

This interval has a simple and flexible form. It shrinks to a point as the width of either bound (11a) or (11b) approaches zero. It widens to interval (2) as the widths of bounds (11a) and (11b) both approach one. An important feature of assumptions (11a) and (11b) is that each bound helps to identify both long predictive probabilities. That is, bound (11a) on $P(y = 1 \mid x = k, w \neq j)$ helps to identify $P(y = 1 \mid x = k, w = j)$ and vice versa. This occurs because equation (3) connects the two probabilities to one another. Assumptions that restrict one imply restrictions on the other.

Illustration: It often is credible to assume that risk of illness varies monotonically with the value of a patient attribute. For example, consider the earlier illustration of risk of breast cancer in which $w = 1$ denotes a heavy drinker and $w = 0$ a non-heavy drinker. Epidemiological research indicates that risk of breast cancer increases with alcohol consumption (e.g., Singletary and Gapstur, 2001). Under this assumption, the bounds (11) are

(13a) $$0 \leq P(y = 1 \mid x = k, w = 0) \leq P(y = 1 \mid x = k),$$

(13b) $$P(y = 1 \mid x = k) \leq P(y = 1 \mid x = k, w = 1) \leq 1.$$

The resulting identification regions for $P(y = 1 \mid x = k, w = 0)$ and $P(y = 1 \mid x = k, w = 1)$ are

(14a) $$P(y = 1 \mid x = k, w = 0) \in [0, P(y = 1 \mid x = k)] \cap \left[\frac{P(y = 1|x = k) - P(w = 1|x = k)}{P(w = 0|x = k)}, P(y = 1|x = k) \right],$$

(14b) $P(y = 1 \mid x = k, w = 1) \in [P(y = 1 \mid x = k), 1] \cap [P(y = 1 \mid x = k), \dfrac{P(y = 1 \mid x = k)}{P(w = 1 \mid x = k)}]$.

Recall from the earlier illustration that $P(y = 1|x = k) = 0.090$, $P(w = 1|x = k) = 0.034$, and $P(w = 0|x = k) = 0.966$. Inserting these values into (14) yields $P(y = 1|x = k, w = 0) \in [0.058, 0.09]$ and $P(y = 1|x = k, w = 1) \in [0.09, 1]$. These identification regions modestly tighten the ones obtained without any assumption.

Stronger findings emerge if a clinician assumes more than monotonicity. One might, for example, believe that not being a heavy drinker can at most reduce the risk of breast cancer to 0.08. The identification regions combining assumption (13) with this lower bound on $P(y = 1|x = k, w = 0)$ are $P(y = 1|x = k, w = 0) \in [0.08, 0.09]$ and $P(y = 1|x = k, w = 1) \in [0.09, 0.37]$. Thus, asserting a lower bound on the cancer risk of women who are not heavy drinkers tightens the upper bound on the risk for heavy drinkers. □

3.3.2. Bounds on Mean Outcomes

Recall the Goodman (1953) use of instrumental variables to identify $E(y|x, w)$. The derivation began with the Law of Iterated Expectations (6). Goodman studied solution of (6) under an invariance assumption, namely that $E(y \mid x = k, w = j)$ does not vary with x, conditional on w.

In place of Goodman's invariance assumption, one might assert a set of bounded-variation assumptions that impose $M > 0$ linear inequalities restricting $E(y \mid x = k, w = j)$, $(k, j) \in X \times W$. In abstraction, these inequalities have the form

$$(15) \qquad a(m) \leq \sum_{(k, j) \in X \times W} c(m, k, j) \cdot E(y \mid x = k, w = j) \leq b(m), \quad m = 1, \ldots, M,$$

where $[a(m), b(m), c(m, k, j), m = 1, \ldots, M; (k, j) \in X \times W]$ are specified constants.

The identification region for $E(y \mid x = k, w = j)$ is the interval whose lower (upper) bound minimizes (maximizes) $E(y \mid x = k, w = j)$ subject to (6) and (15). These lower and upper bounds solve linear

programming problems. Hence, they should be easy to compute with modern algorithms. It should be straightforward to develop a prediction support tool that queries the clinician to input values for $E(y|x)$, $P(w|x)$, and $[a(m), b(m), c(m, k, j), m = 1, \ldots, M; (k, j) \in X \times W]$. This done, the tool would compute the lower and upper bounds on $E(y \mid x = k, w = j)$ for any specified value of (k, j).

Three special cases of the inequalities (15) may be particularly useful in clinical practice. First, one may impose inequalities of the form

$$(16) \qquad a(m) \leq E(y \mid x = k, w = j) \leq b(m).$$

Such inequalities bound the magnitudes of mean risk assessments. They generalize the bounds (11) on magnitudes posed earlier in the context of binary outcomes to settings with real-valued outcomes.

Second, one may impose inequalities of the form

$$(17) \qquad a(m) \leq E(y \mid x = k, w = j) - E(y \mid x = k', w = j) \leq b(m),$$

where k and k′ are two values of attribute x. Such inequalities bound the variation of mean risk assessments with x, holding w fixed. Goodman's invariance assumption is the special case in which $a(m) = b(m) = 0$.

Third, one may impose inequalities of the form

$$(18) \qquad a(m) \leq E(y \mid x = k, w = j) - E(y \mid x = k, w = j') \leq b(m),$$

where j and j′ are two values of attribute j. Such inequalities bound the variation of mean risk assessments across values of w rather than across values of x.

4. Patient Care with Partial Personalized Risk Assessment

4.1. Optimal and Reasonable Care

Section 3 characterized risk assessment that combines evidence on group outcomes and composition with structural assumptions. The basic lesson was that one may often draw some credible conclusions about the long predictive distribution $P(y|x = k, w = j)$ but one can rarely learn it precisely. Thus, partial personalized risk assessment would seem the norm in clinical practice.

This section considers medical decision making. As mentioned in Section 2, normative studies such as Phelps and Mushlin (1988) have assumed that clinicians maximize expected utility with accurate probabilistic risk assessments conditional on observed patient attributes. Our concern is decision making with less information.

A clinician with partial knowledge may have sufficient information to choose a care option that maximizes expected utility. Consider choice of a breast-cancer care strategy for a woman who has not been diagnosed with the disease. Some CPGs specify that the optimal strategy is (A) periodic screening if the risk of developing the disease is below a certain threshold or (B) prophylactic treatment if risk is above the threshold (e. g., National Comprehensive Cancer Network, 2016). If so, determination of the optimal strategy does not require precise knowledge of risk. It suffices to know whether risk is below or above the threshold.

How might a clinician choose patient care when credible risk assessment is not sufficiently informative to maximize expected utility? Bayesian decision theorists, citing axioms for decision making under uncertainty proposed and studied by Savage (1954), suggest maximization of subjective expected utility, using clinical judgment to make a subjective probabilistic risk assessment. Bayesian patient care may be attractive if subjective risk assessment has a credible foundation, but it may be harmful otherwise.

A clinician who acts without making a subjective probabilistic risk assessment faces a problem of decision making under *ambiguity* (Ellsberg, 1961). There exists no optimal strategy for patient care under

ambiguity. Nevertheless, one can usefully pose alternative decision criteria and compare their properties, the aim being to provide options that a decision maker may view as reasonable.

A broadly reasonable idea is to use a criterion that achieves uniformly satisfactory results, whatever the truth may be. There are multiple ways to formalize the idea of uniformly satisfactory results. Two that have long been prominent are the maximin (von Neumann and Morgenstern, 1944; Wald, 1950) and minimax-regret (Savage, 1951) criteria. It appears that these criteria have not been applied to medical decision making until recently. Manski (2009) studied maximin and minimax-regret treatment choice in an abstract setting. Manski (2010, 2016) used these criteria to consider how society might reasonably choose a vaccination policy under ambiguity. Manski (2013) considered the decision to perform diagnostic testing as a prelude to treatment.

Here I add to this small recent literature by considering choice under ambiguity between active surveillance of a patient (aka watchful waiting) and prophylactic treatment. An apt example is choice between periodic screening for breast cancer and prophylactic treatment. Similar choices are made regularly by clinicians who care for patients at risk of aggressive prostate cancer, heart disease, and many other illnesses. I pose a relatively simple version of the decision problem, for which it is easy to determine the maximin and minimax-regret choices.

The present analysis differs from my earlier work on medical decision making under ambiguity in two important respects. First, this paper maintains a patient-centric perspective in which a clinician wants to care as well as possible for a specific patient. In contrast, my earlier work presumed a public-health perspective in which a health planner wants to maximize a social welfare function that aggregates outcomes across a population of patients. Second, the ecological inference problem studied here has a different structure than the identification problems that I have considered earlier. The aspect of my earlier work most closely related to the present analysis is a short examination in Manski (2000) of treatment choice by a social planner who observes the aggregate outcomes of a classical randomized trial but who does not observe outcomes within sub-populations of subjects.

4.2. Choice under Ambiguity Between Active Surveillance and Prophylactic Treatment

Let y = 1 if a patient will develop a specified disease and y = 0 if not. Let $P_{jk} \equiv P(y = 1|x = k, w = j)$ denote the objective probability that the patient will develop the disease, conditional on the observed attributes (x = k, w = j). Using the available evidence and credible structural assumptions, suppose that the clinician treating the patient concludes that P_{jk} lies in some interval $[P_L, P_H]$; thus, P_L and P_H are the lowest and highest feasible values of the patient's risk of illness.

Suppose that the clinician chooses between two care options, c = A denoting active surveillance and c = B denoting prophylactic treatment. The utility of each option depends on whether the patient will or will not develop the disease. Let U(c, y) denote the utility of option c in the presence of illness outcome y. The utility function U(·, ·) expresses patient preferences and may be specific to the patient under consideration.

Choice between c = A and c = B is a non-trivial problem if the merits of surveillance and treatment vary with the illness outcome. It often is reasonable to suppose that prophylactic treatment is the better option if the patient will develop the disease and that surveillance is the better option otherwise. That is,

(19) $U(B, 1) > U(A, 1)$ and $U(A, 0) > U(B, 0)$.

It is also often reasonable to suppose that, whatever option is used, it is better to be healthy than ill. That is,

(20) $U(A, 0) > U(A, 1)$ and $U(B, 0) > U(B, 1)$.

I assume that these inequalities hold in parts of the analysis below.

4.2.1. Care Maximizing Objective or Subjective Expected Utility

The clinician chooses a care option without knowing the illness outcome. The normative literature

on medical decision making has supposed that the clinician knows the utility function $U(\cdot, \cdot)$ and the personalized illness probability P_{jk}, and that he chooses a care option that maximizes objective expected utility. Thus, the clinician acts as follows:

(21a) Choose c = A if $P_{jk}\cdot U(A, 1) + (1 - P_{jk})\cdot U(A, 0) \geq P_{jk}\cdot U(B, 1) + (1 - P_{jk})\cdot U(B, 0)$,

(21b) Choose c = B if $P_{jk}\cdot U(B, 1) + (1 - P_{jk})\cdot U(B, 0) \geq P_{jk}\cdot U(A, 1) + (1 - P_{jk})\cdot U(A, 0)$.

The solution to (21) is easy to characterize when inequalities (19) hold. Let P^* denote the threshold value of P_{jk} that makes options A and B have the same expected utility. This value is

$$(22) \qquad P^* = \frac{U(A, 0) - U(B, 0)}{[U(A, 0) - U(B, 0)] + [U(B, 1) - U(A, 1)]}.$$

Option A is optimal if $P_{jk} \leq P^*$ and option B if $P_{jk} \geq P^*$.

Our concern is decision making when the clinician does not know P_{jk}. He only knows that $P_{jk} \in [P_L, P_H]$. To focus on the implications of partial personalized risk assessment, I maintain the traditional normative assumption that the clinician knows $U(\cdot, \cdot)$.

A clinician with partial knowledge of P_{jk} can maximize objective expected utility if the threshold probability P^* is not interior to the interval $[P_L, P_H]$. Option A is sure to be optimal if $P_H \leq P^*$ and B is sure to be optimal if $P^* \leq P_L$. The clinician cannot maximize objective expected utility if P^* is interior to $[P_L, P_H]$. Then there exist feasible values of P_{jk} that make only A optimal and other values that make only B optimal.

The Bayesian prescription is to place a subjective distribution on P_{jk} and to maximize subjective expected utility. The Bayesian prescription is easy to characterize in the present decision problem because objective expected utility is linear in P_{jk}. Let π_{jk} denote the subjective mean that a Bayesian clinician holds for P_{jk}. The Bayesian clinician acts as if $P_{jk} = \pi_{jk}$. Thus, option A maximizes subjective expected utility if

$\pi_{jk} \leq P^*$ and B if $\pi_{jk} \geq P^*$.

I henceforth suppose that the clinician does not place a subjective distribution on P_{jk}. Sections 4.2.2 and 4.2.3 study maximin and minimax-regret care respectively.

4.2.2. Maximin Care

The maximin criterion evaluates each action by the worst welfare that it may yield and it chooses an action with the least-bad worst welfare. In the present setting, there are two ways that one might reasonably define the worst welfare of a care option. Hence, there are two ways to implement the maximin criterion.

One approach considers the two possible illness outcomes, y = 0 and y = 1. Then the worst welfare under option A is min[U(A, 0), U(A, 1)] and the worst under B is min[U(B,0), U(B, 1)]. With this definition of worst welfare, option A is a maximin choice if min [U(A, 0), U(A, 1)] ≥ min [U(B, 0), U(B, 1)] and option B if min [U(B, 0), U(B, 1)] ≥ min [U(A, 0), U(A, 1)]. When inequalities (19) and (20) hold, option B is the maximin choice.

The other approach considers the possible values for the objective expected utility of each option. When inequalities (20) hold, the worst feasible expected utility under options A and B both occur when P_{jk} equals its upper bound P_H. Then A and B have objective expected utilities $P_H \cdot U(A, 1) + (1 - P_H) \cdot U(A, 0)$ and $P_H \cdot U(B, 1) + (1 - P_H) \cdot U(B, 0)$. A clinician using this version of the maximin criterion acts as follows:

(23a) Choose c = A if $P_H \cdot U(A, 1) + (1 - P_H) \cdot U(A, 0) \geq P_H \cdot U(B, 1) + (1 - P_H) \cdot U(B, 0)$,

(23b) Choose c = B if $P_H \cdot U(B, 1) + (1 - P_H) \cdot U(B, 0) \geq P_H \cdot U(A, 1) + (1 - P_H) \cdot U(A, 0)$.

Thus, this maximin choice is option A if $P_H \leq P^*$ and B if $P_H \geq P^*$.

The maximin criterion has a deserved reputation for conservatism. Among the two versions of maximin considered here, the former is more conservative than the latter. A clinician using the former version acts as if the patient will become ill for sure. A clinician using the latter version acts as if the patient

will become ill with probability P_H, the upper bound on his risk assessment.

4.2.3. Minimax-Regret Care

The minimax-regret (MMR) criterion evaluates each action by the worst reduction in welfare that it may yield relative to the highest welfare achievable. The term *regret* connotes reduction in welfare relative to the highest achievable. *Maximum regret* is the worst reduction possible, considering all feasible risk assessments. The criterion chooses an action that minimizes maximum regret.

As with maximin, there are two ways that one might reasonably define maximum regret when considering patient care. Hence, there are two ways to implement the criterion. In what follows, I assume that inequalities (19) and (20) hold.

Again, one approach considers the two possible illness outcomes, $y = 0$ and $y = 1$. Inequalities (19) state that prophylactic treatment is the better option when illness occurs and surveillance is better otherwise. Hence, option A has zero regret when $y = 0$ and positive regret $U(B, 1) - U(A, 1)$ when $y = 1$; thus, maximum regret is $U(B, 1) - U(A, 1)$. Symmetrically, option B has zero regret when $y = 1$ and positive regret $U(A, 0) - U(B, 0)$ when $y = 0$; thus, its maximum regret is $U(A, 0) - U(B, 0)$. It follows that option A is a minimax-regret choice if $U(B, 1) - U(A, 1) \leq U(A, 0) - U(B, 0)$ and B if $U(B, 1) - U(A, 1) \geq U(A, 0) - U(B, 0)$. The MMR choice is the same as a clinician maximizing objective expected utility would make if he were to know that the probability of illness is $P_{jk} = ½$.

The other approach considers the possible values for the objective expected utility of each option. Given inequalities (19), maximum regret under option A occurs when P_{jk} equals its upper bound P_H and equals the expected utility difference

$$(24) \qquad [P_H \cdot U(B, 1) + (1 - P_H) \cdot U(B, 0)] - [P_H \cdot U(A, 1) + (1 - P_H) \cdot U(A, 0)].$$

Symmetrically, maximum regret under option B occurs when P_{jk} equals its lower bound P_L and equals

(25) $[P_L \cdot U(A, 1) + (1 - P_L) \cdot U(A, 0)] - [P_L \cdot U(B, 1) + (1 - P_L) \cdot U(B, 0)]$.

Option A is an MMR choice if the difference in expected utilities in (24) is less than or equal to that in (25).

A more transparent representation of this finding emerges if we define P_M to be the midpoint of interval $[P_L, P_H]$. Then rearrangement of terms in (24) and (25) shows that a clinician using this version of the MMR criterion acts as follows:

(26a) Choose c = A if $P_M \cdot U(A, 1) + (1 - P_M) \cdot U(A, 0) \geq P_M \cdot U(B, 1) + (1 - P_M) \cdot U(B, 0)$,

(26b) Choose c = B if $P_M \cdot U(B, 1) + (1 - P_M) \cdot U(B, 0) \geq P_M \cdot U(A, 1) + (1 - P_M) \cdot U(A, 0)$.

Thus, the MMR choice is the same as a clinician maximizing objective expected utility would make if he were to know that the probability of illness is $P_{jk} = P_M$.

The maximin and MMR criteria are sometimes confused with one another. The above derivations and findings make clear that they differ. A clinician using the maximin criterion chooses a care option that maximizes the minimum utility or expected utility that might possibly occur. A clinician using the MMR criterion chooses an option that minimizes the maximum reduction in utility or expected utility that can possibly result from incomplete knowledge.

We have found that the care choices yielded by the two maximin criteria are those that a clinician maximizing expected utility would make if he were to adopt the pessimistic perspective that illness will occur for sure or with probability P_H. In contrast, the choices yielded by the two MMR criteria are those that a clinician maximizing expected utility would choose if he were to adopt the middle-ground perspective that illness will occur with probability ½ or P_M.

5. Conclusion: Rethinking Care with Actuarial Prediction

The psychological literature on clinical judgment exemplified by Dawes, Faust, and Meehl (1989) does not recommend clinical use of any of the decision criteria discussed in Section 4—not maximization of subjective expected utility, nor maximin, nor MMR. Instead it recommends that the clinician suppress his knowledge of patient attributes $w = j$ and act as if $P_{jk} = P(y = 1|x = k)$.

Acting as if $P_{jk} = P(y = 1|x = k)$ is clearly inappropriate if the value of this short probability lies outside the interval $[P_L, P_H]$ of credible potential values of P_{jk}. One might rationalize behaving in this manner if $P(y = 1|x = k) \in [P_L, P_H]$ by asserting that the short probability is a possible value of P_{jk}. However, the same assertion can be made for any element of $[P_L, P_H]$. I am unaware of any formal argument that justifies singling out $P(y = 1|x = k)$.

Section 4.2 showed that decision making with the second version of the maximin or MMR criterion is equivalent to acting as if P_{jk} takes particular values in $[P_L, P_H]$, namely P_H for maximin and P_M for MMR. Singling out these values has a firmer justification because they yield care choices that are uniformly satisfactory in the maximin or MMR sense.

The negative conclusion reached here regarding acting as if $P_{jk} = P(y = 1|x = k)$ does not contradict the longstanding conclusion of psychological research that actuarial prediction outperforms informal clinical judgment. Psychologists may be correct that clinician failure to adequately grasp the logic of the prediction problem generates a broad empirical finding in favor of actuarial prediction. Coherent combination of evidence and judgment is a subtle matter. It may be unrealistic to expect clinicians to understand the mathematics of ecological inference or to accurately perform it in their heads, without the assistance of decision support tools.

What my analysis does suggest is that it may be possible to improve on both actuarial prediction and informal clinical judgment by formalizing clinical judgment.

References

Academy of Medical Sciences (2015), *Stratified, Personalised or P4 Medicine: a New Direction for Placing the Patient at the Centre of Healthcare and Health Education*, www.acmedsci.ac.uk/viewFile/56e6d483e1d21.pdf, accessed July 4, 2016.

American Medical Association (2010), *Genomic-based Personalized Medicine*, Report 4 of the Council on Science and Public Health, download.ama-assn.org/resources/doc/csaph/x-pub/a10csaph4.pdf, accessed July 4, 2016.

Camerer, C. and E. Johnson (1997), "The Process-Performance Paradox in Expert Judgment: How Can Experts Know so Much and Predict so Badly," in *Research on Judgment and Decision Making*, W. Goldstein and R. Hogarth (editors), Cambridge: Cambridge University Press.

Centers for Disease Control and Prevention (2015), *United States Life Tables*, 2011, www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_11.pdf, accessed August 18, 2016.

Cho, W. (1998), "Iff the Assumption Fits . . . : a Comment on the King Ecological Inference Solution," *Political Analysis*, 7, 143-163.

Cho, W. and B. Gaines (2004), "The Limits of Ecological Inference: the Case of Split-ticket Voting," *American Journal of Political Science*, 48, 52–71.

Claus, E, N. Risch, and W. Thompson (1994), "Autosomal Dominant Inheritance of Early-onset Breast Cancer. Implications for Risk Prediction," *Cancer*, 73, 643-651.

Cross, P. and C. Manski (2002), "Regressions, Short and Long," *Econometrica*, 70, 357-368.

Dawes, R., R. Faust, and P. Meehl (1989), "Clinical Versus Actuarial Judgment," *Science*, 243, 1668-1674.

Duncan, O. and B. Davis (1953), "An Alternative to Ecological Correlation," *American Sociological Review*, 18, 665-666.

Ellsberg D. (1961), "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics*, 75, 643-669.

Freedman, D., S. Klein, M. Ostland, and M. Roberts (1998), "Review of *A Solution to the Ecological Inference Problem*, by G. King," *Journal of the American Statistical Association*, 93, 1518-1522.

Freedman, D., S. Klein, M. Ostland, and M. Roberts (1999), "Response to King's Comment," *Journal of the American Statistical Association*, 94, 355-357.

Gail, M, L. Brinton, D. Byar, D. Corle, S. Green, C. Shairer, and J. Mulvihill (1989), "Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually," *Journal of the National Cancer Institute*, 81,1879-86.

Ginsburg, G. and H. Willard (2009), "Genomic and Personalized Medicine: Foundations and Applications," *Translational Research*, 154, 277-287.

Goodman, L. (1953), "Ecological Regressions and Behavior of Individuals," *American Sociological Review*, 18, 663-664.

Horowitz, J. and C. Manski (1995), "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63, 281-302.

Huber, P. (1964), "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35, 73-101.

King, G. (1997), *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*, Princeton: Princeton University Press.

King, G. (1999), "The Future of Ecological Inference Research: A Comment on Freedman et al.," *Journal of the American Statistical Association*, 94, 352-355.

Mandelblatt J., K. Cronin, S. Bailey *et al.,* Breast Cancer Working Group of the Cancer Intervention and Surveillance Modeling Network (2009), "Effects of Mammography Screening under Different Screening Schedules: Model Estimates of Potential Benefits and Harms," *Annals of Internal Medicine*, 151, 738-747.

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. (2007), *Identification for Prediction and Decision*, Cambridge, Harvard University Press.

Manski C. (2009), "Diversified Treatment under Ambiguity," *International Economic Review*, 50,1013-1041.

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

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. (2016), "Mandating Vaccination with Unknown Indirect Effects," *Journal of Public Economic Theory*, forthcoming.

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

Manski, C. and J. Pepper (2013), "Deterrence and the Death Penalty: Partial Identification Analysis Using Repeated Cross Sections," *Journal of Quantitative Criminology*, 29, 123-141.

Manski, C. and J. Pepper (2016), "How Do Right-to-Carry Laws Affect Crime Rates? Coping with Ambiguity Using Bounded-Variation Assumptions," Department of Economics, Northwestern University.

Meehl, P. (1954), *Clinical Versus Statistical Prediction: a Theoretical Analysis and a Review of the Evidence*, Minneapolis: University of Minnesota Press.

Meltzer, D. (2001), "Addressing Uncertainty in Medical Cost-effectiveness: Implications of Expected Utility Maximization for Methods to Perform Sensitivity Analysis and the Use of Cost-effectiveness Analysis to Set

Priorities for Medical Research," *Journal of Health Economics,* 20, 109-129.

Myers, E, P. Moorman, J. Gierisch, L. Havrilesky, L. Grimm, S. Ghate, B. Davidson, R. Mongtomery, M. Crowley, D. McCrory, A. And G. Sanders (2015), "Benefits and Harms of Breast Cancer Screening: A Systematic Review," *Journa of the American Medical Association,* 314, 1615-1634.

National Cancer Institute (2016), *Breast Cancer Risk Assessment Tool*, http://www.cancer.gov/bcrisktool/, accessed July 4, 2016.

National Comprehensive Cancer Network (2016), *Breast Cancer Screening and Diagnosis*, Version 1.2016, www.nccn.org/professionals/physician_gls/pdf/breast-screening.pdf, login required, accessed August 14, 2016.

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

President's Council of Advisors on Science and Technology (2008), "Priorities for Personalized Medicine," www.whitehouse.gov/files/documents/ostp/PCAST/pcast_report_v2.pdf, accessed July 3, 2016.

Robinson, W. (1950), "Ecological Correlation and the Behavior of Individuals," *American Sociological Review*, 15, 351-357.

Sarbin, T. (1943), "A Contribution to the Study of Actuarial and Individual Methods of Prediction," *American Journal of Sociology*, 48, 593– 602.

Sarbin, T. (1944), "The Logic of Prediction in Psychology," *Psychological Review*, 51, 210-228.

Savage, L. (1951), "The Theory of Statistical Decision," *Journal of the American Statistical Association*, 46, 55‐67.

Savage, L. (1954), *The Foundations of Statistics*, New York: Wiley.

Shlonsky, A. and D. Wagner (2005), "The next Step: Integrating Actuarial Risk Assessment and Clinical Judgment into an Evidence-based Practice Framework in CPS Case Management," *Children and Youth Services Review*, 27, 409-427.

Singletary, K. and S. Gapstur (2001), "Alcohol and Breast Cancer: Review of Epidemiologic and Experimental Evidence and Potential Mechanisms," *Journal of the American Medical Association*, 286, 2143-2151.

Substance Abuse and Mental Health Services Administration (2014), *2014 National Survey on Drug Use and Health (NSDUH). Table 2.46B—Alcohol use, binge alcohol use, and heavy alcohol use in the past month among persons aged 18 or older, by demographic characteristics: Percentages, 2013 and 2014,* www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs2014/NSDUH-DetTabs2014.htm#tab2-46b, accessed July 17, 2016.

Susan G. Komen (2016), *Estimating Breast Cancer Risk*, ww5.komen.org/BreastCancer/GailAssessmentModel.html, accessed July 9, 2016.

Von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior*, Princeton: Princeton University Press.

Wakefield, J. (2008), "Ecological Studies Revisited," *Annual Review of Public Health*, 29, 75-90.

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