CREDIBLE ECOLOGICAL INFERENCE FOR MEDICAL DECISIONS WITH PERSONALIZED RISK ASSESSMENT

Charles F. Manski

Department of Economics and Institute for Policy Research
Northwestern University

first public version October 2016
this version September 2017

Abstract

This paper studies an identification problem that arises when clinicians seek to personalize patient care by predicting health outcomes conditional on observed patient covariates. Let y be an outcome of interest and let (x = k, w = j) be observed patient covariates. Suppose a clinician wants to choose a care option that maximizes a patient's expected utility conditional on the observed covariates. 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 evidence-based risk assessment that predicts y conditional on a subset of the observed covariates, say x, but not conditional on (x, w). Then the clinician knows P(y|x = k) but not P(y|x = k, w = j). Research on the ecological inference problem studies partial identification of P(y | x, w) given knowledge of P(y|x) and P(w|x). Combining this knowledge with structural assumptions yields tighter conclusions. A psychological literature comparing actuarial predictions and clinical judgments has concluded that clinicians should not attempt to subjectively predict patient outcomes conditional on covariates that are not utilized in evidence-based risk assessments. I argue that formalizing clinical judgment through analysis of the identification problem can improve risk assessments and care decisions.

I have benefitted from the opportunity to present this work in seminars at the Bonn Graduate School of Economics, the Federal Reserve Bank of Cleveland, Georgia State University, Hebrew University, McMaster University, Northwestern University, and Rice University. I have benefitted from the comments of Pamela Giustinelli, Max Tabord-Meehan, and three anonymous reviewers, as well as the research assistance of Shaun Shaikh. Earlier versions of this paper were circulated under the title "Credible Ecological Inference for Personalized Medicine: Formalizing Clinical Judgment."

1. Introduction

Let each member of a population be characterized by a triple (y, x, w) whose value lies in a set $Y \times X \times W$. Let P(y, x, w) denote the population distribution. A well-known identification problem is inference on the *long* predictive distribution $P(y|x, w)$ given knowledge of the *short* distributions $P(y|x)$ and $P(w|x)$. The basic problem supposes that no other information is available. Stronger conclusions may be drawn if one combines knowledge of P(y|x) and P(w|x) with structural assumptions embodying some a priori knowledge of P(y|x, w). The problem has been studied in several literatures with varying substantive concerns and terminology, including those on ecological inference and contaminated sampling. Cross and Manski (2002) reviews the literature and Manski (2007a, Chapter 5) provides a textbook exposition. I use the term *ecological inference* here.

This paper addresses a common occurrence of the ecological inference problem in decision making with personalized risk assessments. I study the situation of a clinician caring for patients.[1] The clinician may want to predict health outcomes such as disease development or life span conditional on a host of observed patient covariates—demographic traits, medical history, the results of screening and diagnostic tests, and current treatment status. However, an available evidence-based tool may predict outcomes conditional on just a subset of these covariates.

Many other decision makers face similar prediction problems. A judge deciding how to sentence convicted offenders may want to assess the probability of recidivism conditional on attributes including (gender, age, prior convictions, demeanor in court), but an available prediction tool may only condition on (gender, age, prior convictions). A lender deciding whether to approve loan applications may want to assess the probability of repayment conditional on (assets, debt history, family status, employment), but an available credit scoring tool may only reveal the probability of repayment conditional on assets and debt history. A

[1] Throughout the paper I refer to the decision maker as the 'clinician.' Some health care choices may be made by patients or jointly by clinicians and patients. Assessment of health outcomes is a common concern whoever the actual decision maker may be.

firm deciding whether to employ job applicants may want to assess the probability of good job performance conditional on (schooling, previous employment, quality of job interview), but an available tool may predict performance conditional only on schooling. A college deciding whether to admit student applicants may be in an analogous situation.

I focus on the clinician's problem, leaving other applications for future research. Medical decision making is an important societal matter, with considerable recent attention to the practice of personalized medicine.[2] Medicine offers multiple instances of the ecological problem that are ripe for study. There exist many publicly available, evidence-based tools that make probabilistic predictions P(y|x) of clinically important health outcomes conditional on some observable patient covariates but not others. There also are surveys and administrative data that reveal the distribution P(w|x) for clinically observable covariates (x, w).

A simple example of a publicly available evidence-based predictor of health outcomes is the set of life tables published by the U. S. Centers for Disease Control (CDC), which predict life span conditional on (age, sex, race) but not conditional on a person's health or treatment status (Arias, 2015). A more complex example is the Breast Cancer Risk Assessment (BCRA) Tool of the National Cancer Institute (2016). The BCRA Tool personalizes risk of breast cancer in multiple respects, but it does not condition on further observable patient covariates that may be associated with risk of cancer.[3] The BCRA Tool has become

---

[2] The term *personalized medicine* is sometimes taken to mean care that is literally specific to the individual, but patient data to support complete personalization is rarely available. Hence, the term is commonly used to mean care that varies with observed patient covariates.

[3] The BCRA Tool gives a predicted probability P(y = 1|x) that a woman will develop invasive breast cancer (y = 1) conditional on eight covariates: (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).

The reason that the tool assesses risk conditional on these covariates and not others is that it uses a modified version of the "Gail Model," based on the research of Gail *et al.* (1989). The Gail *et al.* article estimated probabilities of breast cancer for white women who have annual breast examinations, conditional on covariates (1) through (7). Scientists at the National Cancer Institute later modified the model to predict

widely used in clinical practice (Susan G. Komen, 2016) and is an important input to the clinical practice guideline (CPG) for breast cancer screening issued by National Comprehensive Cancer Network (2016). The CPG recommends that women whose risk assessment is below a specified threshold receive periodic screening. It recommends that women with risk assessment above the threshold receive more intense screening or some form of risk-reduction treatment.

To motivate the analysis of this paper, Section 1.1 discusses the two polar prediction options that have been available to date, evidence-based prediction conditional on x and subjective prediction conditional on (x, w). Section 1.2 explains how study of the ecological inference problem can potentially improve on both polar options and summarizes the contributions made here.

1.1. Evidence-Based Prediction and Clinical Judgment

The BCRA Tool exemplifies a common clinical quandary. Available evidence enables one to assess risk conditional on covariates x. One also observes additional covariates w that may be informative predictors of patient outcomes. How should a clinician assess risk?

One polar option has been to ignore w and base care only on x. Another has been to use clinical judgment to predict outcomes conditional on (x, w) and then base care on (x, w). A clinician using the latter

---

invasive cancer in a wider population of women. The outcome predicted by the Gail Model is future development of cancer unconditional on receipt of a future treatment that might reduce the risk of disease.

There are many potentially observable patient covariates that may predict risk of breast cancer but are not used by the BCRA Tool. For example, when considering the number of first-degree relatives with breast cancer (item 6), the Tool does not take into account the number and ages of a woman's first-degree relatives, which 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 Tool groups all white woman together and does not distinguish 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 Tool does not condition on behavioral covariates such as excessive drinking of alcohol, which has been associated with increased risk of breast cancer (Singletary and Gapstur, 2001).

option essentially acts as a subjective Bayesian. If the clinician has rational expectations, that is accurate knowledge of P(y|x, w), prediction using clinical judgment necessarily performs at least as well as evidence-based prediction.[4] However, clinical judgment may yield worse predictions if the clinician lacks rational expectations.[5]

Psychological research comparing evidence-based 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 covariates. The gap in performance persists even when clinical judgment uses additional covariates as predictors.

The influential review article of Dawes, Faust, and Meehl (1989) summarizes the literature well. They distinguish actuarial prediction and clinical judgment as follows:

> "In the clinical method the decision-maker combines or processes information in his 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 when a clinician observes patient covariates not utilized in actuarial prediction, they state:

> "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."

They attribute the weak performance of clinical judgment to clinician failure to adequately grasp the logic

---

[4] It has long been known that, when maximizing expected utility with rational expectations, prediction of outcomes using P(y|x, w) necessarily performs at least as well as prediction using P(y|x). This result is often stated without attribution and I do not know who first proved it. An early version is given in Good (1967). Phelps and Mushlin (1988), Manski (2007a, Sec. 11.4), and Kadane, Schervish, and Seidenfeld (2008) give later statements and proofs in different contexts.

[5] Economists often cite the axiomatic choice analysis of Savage (1954) as an argument that decision makers should act as Bayesians. However, the Savage analysis does not imply rational expectations. Berger (1985) calls pertinent attention to this, stating (page 121): "A Bayesian analysis may be 'rational' in the weak axiomatic sense, yet be terrible in a practical sense if an inappropriate prior distribution is used."

of the prediction problem and to their use of decision rules that place too much emphasis on w relative to x.

Dawes *et al.* caution against use of clinical judgment to predict patient outcomes conditional on patient covariates that are not utilized in evidence-based assessment tools or research reports. In the notation of the present paper, they find that prediction based on accurate evidence-based knowledge of P(y|x) tends to outperform prediction based on a possibly inaccurate subjective estimate of P(y|x, w).[6]

1.2. Prediction and Decision Making with Credible Ecological Inference

Suppose that Dawes *et al.* are correct to advise against subjective clinical judgment to predict patient outcomes. This does not foreclose the possibility of making well-grounded predictions that condition on (x, w) rather than x alone. 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).

As far as I am aware, psychologists and other researchers concerned with medical decision making have not studied personalized risk assessment as an identification problem. Section 2 of this paper does so, bringing to bear my previous research and adding to it. I first summarize established findings on partial identification of P(y|x, w) in the basic case where one only knows P(y|x) and P(w|x). I then consider the identifying power of various structural assumptions.

---

[6] Research published after Dawes, Faust, and Meehl (1989) has largely corroborated the conclusions reached there. For example, Groves *et al.* (2000) conclude their meta-analysis of the literature as follows (p. 25): "Even though outlier studies can be found, we identified no systematic exceptions to the general superiority (or at least material equivalence) of mechanical prediction. It holds in general medicine, in mental health, in personality, and in education and training settings. It holds for medically trained judges and for psychologists. It holds for inexperienced and seasoned judges."

It is well known that point identification of P(y|x, w) can be achieved if sufficiently strong structural assumptions are imposed. A central issue is how to resolve the tension between the strength and credibility of assumptions: stronger assumptions have more identifying power but less credibility. I propose use of credible *bounded-variation* assumptions. These weaken conventional clinical judgment by having the clinician place bounds on features of P(y|x, w) rather than having them conjecture the entire distribution. A prediction support tool could elicit bounded-variation assumptions from clinicians and combine these assumptions with available evidence on P(y|x) and P(w|x) to draw credible partial conclusions on P(y|x, w).

I use two notable prediction problems to demonstrate what may be learned about P(y|x, w) in the basic case and given bounded-variation assumptions. One is prediction of future development of breast cancer. Here the health outcome is binary (y =1 if disease develops, y = 0 otherwise), the BCRA tool provides an evidence-based assessment of P(y = 1|x), and I take the additional observed covariate w to be an indicator of heavy consumption of alcohol.[7] The other problem is prediction of mean life span. Here the outcome is integer-valued (y = number of years of remaining life), the CDC life tables provide an evidence-based assessment of E(y|x) for x = (age, race, sex), and I take w to be an indicator of hypertension status. In both prediction problems, the basic bounds turn out to be wide but imposition of arguably credible bounded-variation assumptions narrows them considerably.

In medical decision making, risk assessment is not an objective per se but rather a prelude to treatment choice. Normative studies of personalized medicine have commonly assumed that the clinician makes accurate probabilistic risk assessments conditional on all observed patient covariates. For example, Phelps and Mushlin (1988) studied optimal diagnostic testing as a prelude to treatment. They assumed that the clinician has rational expectations and that the objective is to maximize expected utility.

---

[7] A similarly interesting binary prediction problem, which I do not address in the paper, is prediction of development of cardiovascular disease (CVD). The American College of Cardiology makes available an online predictor at http://tools.acc.org/ASCVD-Risk-Estimator-Plus/#!/calculate/estimate/. The tool predicts ten-year and lifetime risk of CVD conditional on (age, sex, race, several cholesterol levels, systolic blood pressure, history of diabetes and smoking, current treatment status). Other clinically observable covariates that are not used but that may help predict CVD include obesity, job stress, and exercise.

Section 3 considers the use of risk assessment in medical decision making when a clinician treats a patient with observed covariates (x, w), but the structural assumptions deemed credible are not strong enough to point-identify P(y|x, w). The available knowledge of P(y|x, w), albeit incomplete, may suffice to determine the optimal treatment. If so, decision making using the partially identified P(y|x, w) necessarily yields at least as high welfare as does decision making using either of the two polar prediction options that have been available to date; that is, the point-identified P(y|x) or subjective assessment of P(y|x, w). The decision using the partially identified P(y|x, w) is strictly better when a polar option yields a sub-optimal choice.

If the available information is too limited to determine the optimal treatment, 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 common setting where the clinician chooses between surveillance of a patient and aggressive treatment. The optimal choice is surveillance if the probability of disease development conditional on (x, w) is below a certain personalized threshold value and is aggressive treatment otherwise. The clinician faces ambiguity if he does not know whether the probability of disease development is below or above the threshold.

I derive the maximin and minimax-regret decisions. Both criteria have appealing decision theoretic foundations, yielding decisions that are uniformly satisfactory across states of nature in specific senses. This part of the paper adds new findings to my program of work on medical decision making under ambiguity in two ways. First, it provides new analysis of the substantively important problem of choice between surveillance and aggressive treatment using partial personalized risk assessment. Second, it approaches the problem from a patient-centric perspective, whereas my earlier work took a public health perspective.

The concluding Section 4 comments on potential non-medical applications of the analysis in this paper and suggests directions for further methodological research.

2. Risk Assessment Using Evidence and Structural Assumptions

I analyze personalized risk assessment in stages. Section 2.1 considers the basic problem of inference on P(y|x, w) given knowledge of P(y|x) and P(w|x), without structural assumptions. Section 2.2 considers strong structural assumptions that can point-identify P(y|x, w). Section 2.3 studies weaker but potentially more credible bounded-variation assumptions.

The ecological inference problem may arise whatever the space $Y \times X \times W$ may be. The research to date has assumed that $X \times W$ is a finite space and that Y is either the binary set $\{0, 1\}$ or the real line. I maintain these restrictions. I also assume that the evidence-based assessment tool correctly reveals the short predictive distribution P(y|x). This assumption simplifies analysis and clinicians often maintain it in practice.[8]

The outcome of interest may measure future health status unconditional on future treatment status or it may be the potential outcome y(t) that would occur if the patient were to receive a specified future treatment t. For conciseness, I suppress the t notation when considering potential outcomes.

2.1. The Basic Problem

Consider a patient with covariates $(x = k, w = j)$. The Law of Total Probability relates the short and long predictive distributions:

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

[8] 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, covariate) 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 finite-sample imprecision.

Knowledge of P(y|x = k) alone reveals nothing about P(y|x = k, w = j); any distribution P(y|x = k, w = j) satisfies (1) if P(w = j|x = k) = 0. A partial conclusion may be drawn if one knows P(w = j|x = k), if it is positive. The same conclusion holds if one knows a lower bound on P(w = j|x = k). Then the identification region for P(y|x = k, w = j) is the same as the region obtained if one were to know that P(w = j|x = k) equals its lower bound.[9]

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

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

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

When y is a general real-valued outcome, there is no characterization of the identification region for P(y|x, w) of simplicity comparable to (2). However, Horowitz and Manski (1995) derive tractable sharp bounds on the mean and quantiles of P(y|x, w). Appendix A summarizes their work.

Sections 2.1.1 and 2.1.2 use specific problems of health risk assessment to demonstrate application of the results for binary outcomes and means.

---

[9] The present partial identification analysis, which shows that a person with accurate knowledge of P(y|x) may have only partial knowledge of P(y|x, w), provides specific demonstrations of an abstract probabilistic idea called *dilation*. As defined by Herron, Seidenfeld, and Wasserman (1997), dilation is the phenomenon that (p. 411): "Conditioning can make imprecise probabilities uniformly more imprecise." Grünwald and Halpern (2004) give an extreme medical example in which y is occurrence of a disease and w is a patient's address. In their example, the clinician has accurate knowledge of P(y|x) but no knowledge of P(y|x, w). As far as I am aware, the particular types of dilation that emerge with ecological inference have not been studied in the literature concerned with dilation.

2.1.1. Predicting Risk of Breast Cancer

Consider application of the BCRA Tool at National Cancer Institute (2016) to a woman with these covariates ($x = k$): (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 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 $P(w|x)$, expression (2) 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 heavy drinkers is so small (0.034).

2.1.2. Predicting Life Expectancy Conditional on Hypertension Status

A common problem in health risk assessment is to predict a patient's remaining life span conditional on observed covariates. Let y denote remaining life span. CDC life tables provide actuarial predictions of life span in the United States conditional on age, sex, and race (Arias, 2015). The life tables do not predict life span conditional on other patient covariates that clinicians commonly observe.

Consider prediction of remaining life span conditional on x = (age, sex, race) and w = hypertension

status. In particular, let w be a standard binary classification of persons into those who have or do not have high blood pressure (HBP).[10] For specificity, I consider prediction of remaining life span when x denotes a 50-year-old male in one of two race categories, non-Hispanic (NH) black or white. Tables 14 and 17 of Arias (2015) report that, among males of age 50, NH blacks have over three years lower life expectancy than NH whites, with E(y|age 50, NH black male) = 26.6 and E(y|age 50, NH white male) = 29.7. The tables also provide detailed year-to-year survivorship statistics.

The raw data in the periodic National Health and Nutrition Examination Survey of the National Center for Health Statistics provides a basis for estimation of P(w|x). More easily accessed tables in Go *et al.* (2013) aggregate all adults of age at least 20 years and report P(w|race, sex). I use the age-aggregated probabilities here. Table 9-1 of Go *et al.* (2013) reports that the prevalence of HBP among black males is higher than among white males, with P(HBP|NH black male) = 0.426 and P(HBP|NH white male) = 0.334.[11]

The detailed year-by-year survivorship statistics in Tables 14 and 17 of Arias (2015) provide the data needed to compute the means $[E(L_{jk}), E(U_{jk})]$ of the truncated distributions L_{jk} and U_{jk} shown by Horowitz and Manski (1995) to bound E(y|x =k, w = j). Appendix A explains the calculations. The bounds on E(y|age, race, sex, hypertension status) are

E(y|age 50, NH black male, not HBP) $\in$ [18.1, 35.4], E(y|age 50, NH black male, HBP) $\in$ [14.3, 38.5],
E(y|age 50, NH white male, not HBP) $\in$ [23.8, 36.4], E(y|age 50, NH white male, HBP) $\in$ [15.6, 42.0].

Two features of these bounds are noteworthy. First, the bound on the long conditional mean E(y|x

[10] A person is classified as having HBP if at least one of several conditions hold: systolic blood press is 140 mm Hg or greater, diastolic blood pressure is 90 mm Hg or greater, the person takes antihypertensive medicine, or the person has been told at least twice by a physician or other health professional that he or she has HBP. See Go *et al.* (2013).

[11] These age-aggregated probabilities should be reasonably close approximations to the corresponding probabilities for persons of age 50. Chart 9-1 of Go *et al.* (2013) gives the prevalence of HBP by (age, sex), aggregating across races. It shows that the fraction of males in the age range 45-54 who have HBP is 0.377.

= k, w = j) contains the observed value of the corresponding short conditional mean E(y|x = k). This holds algebraically because E(y|x = k) is necessarily a feasible value for E(y|x =k, w = j). Second, for each value of x, the bound for persons with w = HBP contains the one for persons with w = not HBP. This holds because, for each value of x, the fraction of persons with HBP is smaller than the fraction without HBP.

It is reasonable to assume that, holding x fixed, persons with HBP have lower mean life spans than those without HBP. The bounds given here do not impose this assumption. The identifying power of this and other structural assumptions will be examined in Section 2.3.

2.2. Risk Assessment with Strong Structural Assumptions

The above findings on identification of $P(y|x, w)$ only use knowledge of $P(y|x)$ and $P(w|x)$. The literature has developed two approaches that impose structural assumptions strong enough to yield point identification, one assuming the existence of an instrumental variable and the other assuming a parametric model for $P(y|x, w)$. I review the instrumental-variable analysis of Goodman (1953) here because I view my analysis of bounded-variation assumptions in Section 2.3 in part to be a descendant of his work that weakens the equalities he studied to inequalities. Appendix B summarizes the parametric modeling approach, which is not used in this paper.

Goodman took the objective to be inference on $E(y|x, w)$ when y is real-valued. He used x as an instrumental variable, asserting that y is mean-independent of x, conditional on w. That is,

$$(3) \qquad E(y|x = k, w = j) = E(y|w = j), \quad \text{all } (k, j) \in X \times W.$$

The Law of Iterated Expectations gives

$$(4) \qquad E(y|x=k) = \sum_{j \in W} E(y|x=k, w=j)P(w=j|x=k), \qquad k \in X.$$

For $k \in X$, the data reveal $E(y|x=k)$ and $[P(w=j|x=k), j \in W]$, but not $[E(y|x=k, w=j), j \in W]$. Without assumption (3), (4) is a system of $|X|$ linear equations in the $|X| \times |W|$ unknowns $E(y|x=k, w=j)$, $(k, j) \in X \times W$. With the assumption, (4) becomes

$$(5) \qquad E(y|x=k) = \sum_{j \in W} E(y|w=j)P(w=j|x=k), \qquad k \in X.$$

This is a system of $|X|$ linear equations in 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 (3) point-identifies $E(y|w=j)$, $j \in W$.[12]

Assumption (3) is refutable. Goodman pointed out that equation system (5) may have no solution, implying the assumption is incorrect. Cross and Manski (2002) observed that (3) is also refuted if system (5) has solutions, but none that are feasible given the restrictions implied by knowledge of P(y|x) and P(w|x).

2.2.1. Predicting Life Expectancy with an Instrumental Variable

Consider again prediction of life expectancy given hypertension status. Here assumption (3) states that mean remaining life span does not vary with race, conditional on (age, sex, hypertension status); that is,

[12] Although Goodman (1953) demonstrated the identifying power of assumption (3), he did not advocate its regular use in practice. He cautioned that the assumption holds (p. 663) "in very special circumstances."

It is difficult to conjecture cases of assessment of health risk where (3) may be credible. In risk assessment, x are covariates used to predict outcomes by evidence-based assessment tools and w are clinician-observed covariates not used by these tools. If (3) holds, the covariates x used by assessment tools have no predictive power once one conditions prediction on the additional covariates w. This seems unlikely to occur in general, but it may on occasion. For example, (3) may be credible when health risk is genetically determined, x measures a phenotype that is statistically associated with the underlying genetic determinant of risk, and w is the finding of a DNA test that directly measures the genetic determinant. Then x may have no predictive power for y when w is used as a predictor.

E(y|age 50, NH black male, not HBP) = E(y|age 50, NH white male, not HBP),

E(y|age 50, NH black male, HBP) = E(y|age 50, NH white male, HBP).

Then (5) is a system of two equations in two unknowns, whose unique solution is

$$E(y|\text{age 50, not HBP}) = 41.0, \quad E(y|\text{age 50, HBP}) = 7.3.$$

However, this solution is not feasible. The solution 41.0 for non-HBP persons is larger than the upper bounds for black and white persons derived in Section 2.1.2 using only knowledge of P(y|x) and P(w|x). The solution 7.3 for HBP persons is smaller than the lower bounds derived earlier. Thus, assumption (3) is refuted.

2.3. Bounded-Variation Assumptions

A clinician wanting to assess risk conditioning on patient covariates not used in evidence-based predictors may be discomforted by the analysis in Sections 2.1 and 2.2. Without structural assumptions, drawing informative conclusions about P(y|x = k, w = j) requires a relatively high prevalence of covariate w = j in the group with covariates 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 making no structural assumptions and ones that yield point identification. This section proposes a class of *bounded-variation* assumptions that clinicians may find easy to contemplate and apply. These assumptions flexibly restrict the magnitudes of risk assessments and the degree to which they vary with patient covariates, enabling clinicians to express quantitative judgments in a structured way. Formally, they place partial subjective distributions on P(y|x, w) rather than the precise subjective distributions examined in studies of clinical judgment. I have previously used bounded-variation

assumptions to provide identifying power in other settings (Manski and Pepper, 2000, 2013, 2017).

Section 3.3.1 analyzes some simple bounded-variation assumptions when outcome y is binary. Section 3.3.2 considers assumptions that bound the means of real-valued outcomes.

2.3.1. Binary Outcomes

Recall identification of $P(y = 1|x = k, w = j)$ without structural assumptions. Result (2) emerges by combining (1) with the logical constraints that $P(y = 1\,|\,x = k, w \neq j)$ and $P(y = 1\,|\,x = k, w = j)$ lie in the unit interval.

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

(6a) $$a(k, \neq j) \leq P(y = 1\,|\,x = k, w \neq j) \leq b(k, \neq j),$$

(6b) $$a(k, j) \leq P(y = 1\,|\,x = k, w = j) \leq b(k, j).$$

Combining (1) and (6) yields this bounded-variation identification region for $P(y = 1\,|\,x = k, w = j)$:

(7) $$[a(k, j), b(k, j)] \cap$$

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

It is easy to prove that (7) is the identification region. Solving (1) for $P(y = 1\,|\,x = k, w = j)$ yields

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

Letting $P(y = 1 \mid x = k, w \neq j)$ take all values consistent with (6a) yields a tentative identification region for $P(y = 1 \mid x = k, w = j)$, this being the interval

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

Values of $P(y = 1 \mid x = k, w = j)$ in this interval that are inconsistent with (6b) are infeasible. Shrinking the interval to include only feasible values yields (7). Every point in (7) is feasible because the pair [$P(y = 1 \mid x = k, w = j)$, $P(y = 1 \mid x = k, w \neq j)$] satisfies (6) and no further information is available.

Interval (7) has a simple form. It shrinks to a point as the width of either bound (6a) or (6b) approaches zero. It widens to interval (2) as the widths of bounds (6a) and (6b) both approach one. An important feature of assumptions (6a) and (6b) is that each bound helps to identify both long predictive probabilities. That is, bound (6a) 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 (1) connects the two probabilities to one another. Assumptions that restrict one imply restrictions on the other.

Predicting the Risk of Breast Cancer

It often is credible to assume that risk of illness varies monotonically with the value of a patient covariate. In the context of breast cancer, it is credible to assume that the risk of illness increases with several observable covariates not used by the BCRA Tool. It should increase with observable indicators of genetic susceptibility such as the prevalence of breast cancer among second-degree relatives and membership in ethnic subgroups such as Ashkenazi Jews. It should also increase with behavioral covariates such as excessive drinking of alcohol.

Consider the earlier application to 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). Given this assumption, the bounds (6) are

(8a) $0 \leq P(y = 1 \mid x = k, w = 0) \leq P(y = 1 \mid x = k)$,

(8b) $P(y = 1 \mid x = k) \leq P(y = 1 \mid x = k, w = 1) \leq 1$.

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

(9a) $$P(y = 1 \mid x = k, w = 0) \in [0, P(y = 1 \mid x = k)] \cap \left[\frac{P(y = 1|x = k) - P(w = 1|x = k)}{P(w = 0|x = k)}, P(y = 1|x = k)\right],$$

(9b) $$P(y = 1 \mid x = k, w = 1) \in [P(y = 1 \mid x = k), 1] \cap \left[P(y = 1|x = k), \frac{P(y = 1|x = k)}{P(w = 1|x = k)}\right].$$

Recall from the earlier application 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 (9) 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 (8) 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.

2.3.2. Bounds on Mean Outcomes

A clinician may find it credible to assume various bounds on $E(y|x, w) = [E(y \mid x = k', w = j'), (k', j') \in X \times W]$. For example, one might bound the magnitudes of mean risk assessments, assuming that

(10) $\quad a(k', j') \le E(y|x = k', w = j') \le b(k', j')$

for specified covariate values $(x = k', w = j')$ and constants $[a(k', j'), b(k', j')]$. Such inequalities generalize the bounds (6) on probabilities of binary outcomes to settings with real-valued outcomes.

One might fix $w = j$ and bound the variation of mean risk assessments with x, assuming that

(11) $\quad a(k', k'', j) \le E(y|x = k', w = j) - E(y|x = k'', w = j) \le b(k', k'', j),$

where k' and k'' are two values of x. Analogously one might fix $x = k$ and bound the variation of mean risk assessments across w, assuming that

(12) $\quad a(k, j', j'') \le E(y|x = k, w = j') - E(y|x = k, w = j'') \le b(k, j', j''),$

where j' and j'' are two values of w. Inequalities of form (11) weaken the Goodman mean-invariance assumption, which is the special case with $a(k', k'', j) = b(k', k'', j) = 0$.

Inequalities (10)–(12) are members of a class of bounded-variation assumptions that impose $M > 0$ linear inequalities restricting $E(y|x, w)$. In abstraction, these inequalities have the form

(13) $\quad a(m) \le \sum_{(k', j') \in X \times W} c(m, k', j') \cdot E(y|x = k', w = j') \le b(m), \quad m = 1, \ldots, M,$

where $[a(m), b(m), c(m, k', j'), m = 1, \ldots, M; (k', j') \in X \times W]$ are specified constants. In (10)–(12), each constant $c(m, k', j')$ takes one of the values (-1, 0, 1). The identification region for $E(y|x = k, w = j)$ is the interval whose lower (upper) bound minimizes (maximizes) $E(y|x = k, w = j)$ subject to the bounded-variation assumptions (13) and the restrictions on $E(y|x, w)$ implied by knowledge of $P(y|x)$ and $P(w|x)$.

Cross and Manski (2002) proves that, for each value of k', knowledge of $P(y|x)$ and $P(w|x)$ confines

the |W|-vector [E(y|x = k′, w] to a bounded convex set whose extreme points are the expectations of certain |W|-tuples of *stacked distributions*. Knowledge of P(y|x) and P(w|x) yields no cross-x restrictions. Hence, the identification region for E(y|x, w) without bounded-variation assumptions is the Cartesian Product of these convex sets across the values of x.

Cross and Manski (2002) did not study identification when knowledge of P(y|x) and P(w|x) is combined with bounded-variation assumptions, but they did study the case when knowledge of P(y|x) and P(w|x) is combined with the Goodman mean-invariance assumption. The analysis is complex, suggesting that computation of the identification region for E(y|x = k, w = j) is generally difficult when bounded-variation assumptions are imposed. Hence, it can be useful to obtain an informative bound on E(y|x = k, w = j) that may not be sharp but is easy to compute.

With this in mind, recall the finding of Horowitz and Manski (1995) that the identification region for E(y|x = k, w = j) given only knowledge of P(y|x) and P(w|x) is the interval $[E(L_{jk}), E(U_{jk})]$, where L_{jk} and U_{jk} are the truncated outcome distributions defined in Appendix A. It follows that the vector E(y|x, w) is contained in the |X|×|W|-rectangle $\times_{(k', j') \in X \times W} [E(L_{j'k'}), E(U_{j'k'})]$, which is easy to compute. This is the smallest rectangular set that encloses the convex identification region studied by Cross and Manski.

Now combine the bounded-variation assumptions (13) with knowledge that E(y|x, w) satisfies the Law of Iterated Expectations (4) and is contained in $\times_{(k', j') \in X \times W} [E(L_{j'k'}), E(U_{j'k'})]$. The result is a simple, albeit possibly non-sharp, bound on E(y|x = k, w = j). The lower (upper) bound minimizes (maximizes) E(y|x = k, w = j) subject to the linear inequalities in (13), the additional linear inequalities $E(L_{j'k'}) \leq E(y|x = k', w = j') \leq E(U_{j'k'})$, $(k', j') \in X \times W$, and the linear equation (4). The lower and upper bounds on E(y|x = k, w = j) solve linear programming problems and thus are typically easy to compute. It should be straightforward to develop a prediction support tool that queries a clinician to determine the bounded-variation assumptions he thinks credible and combines these assumptions with available evidence on P(y|x) and P(w|x) to compute the lower and upper bounds on E(y|x = k, w = j) for specified values of (k, j).

Predicting Life Expectancy Conditional on Hypertension Status

Consider again prediction of life expectancy conditional on hypertension status. It was shown in Section 2.2.1 that the mean-invariance assumption (3) is refuted. Bounded-variation assumptions may be credible, informative, and consistent with the available evidence.

One may, for example, find it reasonable to conjecture that, holding (age, sex, race) fixed, persons with HBP have lower life expectancy than those without HBP. This gives the bounded-variation assumptions

$$0 \leq E(y|\text{age 50, NH white male, not HBP}) - E(y|\text{age 50, NH white male, HBP}),$$

$$0 \leq E(y|\text{age 50, NH black male, not HBP}) - E(y|\text{age 50, NH black male, HBP}).$$

One might also conjecture that black males tend to face various health disadvantages relative to white males beyond high blood pressure and, hence, that black males have lower life expectancy than white males conditional on hypertension status. Going further, one may perhaps find it credible to conjecture that, at age 50 and conditional on hypertension status, the life expectancy of white males is between zero and 2.5 years greater than black males. This gives the bounded-variation assumptions[13]

$$0 \leq E(y|\text{age 50, NH white male, not HBP}) - E(y|\text{age 50, NH black male, not HBP}) \leq 2.5,$$

$$0 \leq E(y|\text{age 50, NH white male, HBP}) - E(y|\text{age 50, NH black male, HBP}) \leq 2.5.$$

Combining these assumptions with the bounds on E(y|x, w) derived in Section 2.1.2 using only

[13] One might conjecture a tighter upper bound than 2.5 years. However, imposing a bound much lower than 2.5 years turns out to imply that the linear program has no feasible solution. In particular, there is no solution when the upper bound is specified to be two years. On the other hand, conjecturing a higher upper bound than 2.5 weakens the derived bounds.

knowledge of P(y|x) and P(w|x) yields the bounds[14]

E(y|age 50, NH black male, not HBP) $\in$ [29.4, 35.4], E(y|age 50, NH black male, HBP) $\in$ [14.7, 22.9],
E(y|age 50, NH white male, not HBP) $\in$ [31.9, 36.4], E(y|age 50, NH white male, HBP) $\in$ [16.3, 25.4].

These bounds are highly informative. They reveal that the life expectancy of 50-year-old black males without HBP is at least 6.5 (i.e., 29.4 – 22.9) years higher than that of those with HBP. For 50-year-old white males, the corresponding disparity is also at least 6.5 (i.e., 31.9 – 25.4) years.

Recall from Section 2.1.2 that the life expectancies of 50-year-old black and white males unconditional on hypertension status are E(y|age 50, NH black male) = 26.6 and E(y|age 50, NH white male) = 29.7. These values lie well outside all of the bounds on life expectancy conditional on hypertension status.

The above example suffices to make the point, but it is evident that other bounded-variation assumptions may be credible as well. For example, one may think it reasonable to conjecture that within each (age, race, hypertension) group, males have lower life expectancy than do females.

3. Patient Care with Partial Personalized Risk Assessment

3.1. Optimal and Reasonable Care

Section 2 studied risk assessment that uses knowledge of P(y|x) and P(w|x), possibly combined with structural assumptions on P(y|x, w). The lesson was that one may draw credible partial conclusions about P(y|x, w), but one can rarely learn it precisely. Thus, partial personalized risk assessment would seem the

[14] The bounds were computed using an online linear program calculator at the website http://comnuan.com/cmnn03/cmnn03004/,accessed September 14, 2017.

norm in clinical practice.

This section considers medical decision making. Normative studies such as Phelps and Mushlin (1988) and Meltzer (2001) have assumed that clinicians maximize expected utility with accurate probabilistic risk assessments conditional on observed patient covariates. My concern is decision making with less information.

A clinician with partial knowledge of P(y|x, w) may nonetheless have sufficient information to determine a care option that maximizes objective expected utility. If so, decision making using the partially identified P(y|x, w) necessarily yields at least as high welfare as does decision making using either of the two polar prediction options—the point-identified P(y|x) or subjective assessment of P(y|x, w). The decision using the partially identified P(y|x, w) is strictly better when a polar option yields a sub-optimal choice.

For example, consider choice between surveillance and aggressive treatment for a woman who is at risk of developing breast cancer.[15] I show in Section 3.2 that the optimal strategy is surveillance if the probability P(y = 1|x, w) of developing the disease is below a certain personalized threshold and aggressive treatment if the probability is above the threshold. Determination of the optimal strategy does not require precise knowledge of the risk of disease. It suffices to know whether P(y = 1|x, w) is below or above the threshold. Prediction using the point-identified P(y = 1|x) or a subjective assessment of P(y = 1|x, w) is sub-optimal if this prediction is on the wrong side of the threshold.

How might a clinician choose patient care when credible risk assessment is not sufficiently informative to maximize objective expected utility? Bayesian decision theorists, citing axioms for decision making 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 is attractive if subjective risk assessment has a credible foundation. However, the psychological literature on clinical judgment shows that it may be harmful otherwise.

---

[15] Aggressive treatment strategies in this context range from intensive screening to drug treatment intended to reduce the risk of cancer onset to preventive mastectomy.

A clinician who acts without making a precise subjective probabilistic risk assessment faces a problem of decision making under *ambiguity* (Ellsberg, 1961). The optimal strategy for patient care may not be known 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, 2017) 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.

Section 3.2 adds to this small recent literature by considering choice under ambiguity between surveillance and aggressive treatment. An apt example is choice between periodic screening for breast cancer and aggressive 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. While motivated by the ecological inference problem, the analysis below applies more generally to any setting in which credible assumptions yield a partial personalized risk assessment.[16]

---

[16] The present analysis differs from my earlier work on medical decision making under ambiguity in that 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. In most of this work, the source of ambiguity was not the ecological inference problem but rather the selection problem; that is, partial knowledge of the process of treatment selection in a study population. However, one of these earlier studies analyzed a situation in which a planner faces the ecological inference problem.

Manski (2000) studied treatment choice by a planner who knows the distribution of treatment outcomes in a classical randomized experiment conditional on covariates x but not conditional on (x, w). Thus, the planner knows P[y(t)|x] for each potential treatment t but he does not know P[y(t)|x, w]. This

3.2. Choice Between Surveillance and Aggressive Treatment

Suppose that a clinician caring for a currently healthy patient with covariates $(x = k, w = j)$ chooses between two care options for a possible disease. Option $c = A$ denotes surveillance and $c = B$ denotes aggressive treatment. Let $y = 1$ if a patient will develop the disease and $y = 0$ if not. Assume that the chosen care option does not affect whether the patient will develop the disease, but it does affect the severity of the consequences. The two care options may also have different side effects.[17]

Let $P_{jk} \equiv P(y = 1|x = k, w = j)$ for short. Using the available evidence and credible structural assumptions, suppose that the clinician 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.

The utility of each care option depends on whether the patient will or will not develop the disease. Let $U(c, y)$ denote the utility of option c when outcome y occurs. Utility function $U(\cdot, \cdot)$ expresses patient preferences and is specific to the patient under consideration. Assume that the clinician knows U.

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

(14) $U(B, 1) > U(A, 1)$ 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,

---

informational situation is common in patient care. The medical journal articles that report the findings of clinical trials typically do not publish outcomes conditional on extensive covariate information. They usually describe outcomes only within broad risk-factor groups.

[17] A variant of the analysis applies to another scenario in which aggressive treatment always prevents development of the disease but has side effects. In this scenario, there are two potential outcomes, y(A) and y(B). The uncertain outcome of interest is $y = y(A)$. It is known with certainty that $y(B) = 0$.

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

I assume that these inequalities hold in the analysis below.

3.2.1. 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 both $U(\cdot, \cdot)$ and P_{jk}, and that he maximizes objective expected utility. Thus, the clinician is assumed to act as follows:

(16a) Choose c = A if $P_{jk}\cdot U(A, 1) + (1 - P_{jk})\cdot U(A, 0) \geq P_{jk}\cdot U(B, 1) + (1 - P_{jk})\cdot U(B, 0)$,

(16b) 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 (16) is easy to characterize when inequalities (14) hold. Let P^* denote the threshold value of P_{jk} that makes options A and B have the same expected utility. This value is

$$(17) \qquad P^* = \frac{U(A, 0) - U(B, 0)}{[U(A, 0) - U(B, 0)] + [U(B, 1) - U(A, 1)]}.$$

Observe that $0 < P^* < 1$. Option A is optimal if $P_{jk} \leq P^*$ and option B if $P_{jk} \geq P^*$.

My concern is decision making when the clinician 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 setting 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 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 3.2.2 and 3.2.3 study maximin and minimax-regret care respectively.

3.2.2. 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 (14) and (15) hold, option B is the maximin choice.

The other approach considers the possible values for the objective expected utility of each option. When inequalities (15) 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:

(18a) Choose c = A if $P_H \cdot U(A, 1) + (1 - P_H) \cdot U(A, 0) \geq P_H \cdot U(B, 1) + (1 - P_H) \cdot U(B, 0)$,

(18b) 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. A clinician using the former version acts as if the patient will become ill for sure, disregarding probabilistic risk assessment entirely. One using the latter version acts as if the patient will become ill with probability P_H, the upper bound on his risk assessment.

3.2.3. 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 to define maximum regret when considering patient care. Hence, there are two ways to implement the criterion. Again, one approach considers the two possible outcomes, y = 0 and y = 1. Inequalities (14) state that aggressive 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) ≤ U(A, 0) − U(B, 0) and B if U(B, 1) − U(A, 1) ≥ 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 (14), maximum regret under option A occurs when P_{jk} equals its upper bound P_H and equals the expected utility difference

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

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

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

Option A is an MMR choice if the difference in expected utilities in (19) is less than or equal to that in (20).

A more transparent representation of this finding emerges if we define P_M to be the midpoint of interval $[P_L, P_H]$. Rearrangement of terms in (19) and (20) shows that a clinician using this version of the MMR criterion acts as follows:

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

(21b) 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 neutral perspective that illness will occur with probability ½ or P_M.

3.3. Rethinking Care with Actuarial Prediction

The psychological literature on clinical judgment exemplified by Dawes, Faust, and Meehl (1989) has not recommended clinical use of any of the decision criteria discussed in Section 3.2—not maximization of subjective expected utility, nor maximin, nor MMR. Instead it recommends that the clinician suppress his knowledge of the patient covariates w 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}. $P(y = 1|x = k)$ necessarily lies within $[P_L, P_H]$ when no structural assumptions are imposed, but it may lie outside $[P_L, P_H]$ given structural assumptions. More generally, when the outcome is real-valued, P(y|x) may be an infeasible value for P(y|x, w) when structural assumptions are imposed. The application in Section 2.3.2 demonstrated this. We found there that life expectancies unconditional on hypertension status lie substantially outside the bounds on life expectancies conditional on hypertension status that hold given the specified bounded-variation assumptions.

One can rationalize acting as if $P_{jk} = P(y = 1|x = k)$ when $P(y = 1|x = k) \in [P_L, P_H]$, as the short probability is a possible value of P_{jk}. However, one can similarly rationalize acting as if P_{jk} is any element of $[P_L, P_H]$. Section 3.2 showed that decision making with the second version of the maximin or MMR criterion is equivalent to acting as if P_{jk} takes specific values in $[P_L, P_H]$, P_H for maximin and P_M for MMR. Using these values has a decision-theoretic foundation because each yields a care choice that is uniformly satisfactory in the maximin or MMR sense. Acting as if $P_{jk} = P(y = 1|x = k)$ has no such foundation.

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 subjective 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 is unrealistic to expect clinicians to understand the mathematics of ecological inference or to accurately perform it in their heads. What my analysis suggests is that it may

be possible to improve on both actuarial prediction and subjective clinical judgment by providing clinicians with decision support tools that formalize credible clinical judgment.

4. Directions for Further Applications and Methodological Research

This paper has reported new analysis of the ecological inference problem and its implications for decision making. Section 2.3 added to my previous analysis of partial identification by studying ecological inference with bounded variation assumptions. Section 3 initiated patient-centric analysis of treatment under ambiguity by considering choice between surveillance and aggressive treatment using partial personalized risk assessment.

These contributions notwithstanding, I do not view the primary significance of the paper to be development of new methodology per se. It is rather that the paper calls attention to a difficult identification problem that arises in personalized risk assessment and shows how analysis of partial identification and decision making under ambiguity may potentially be used to improve decision making. I used two substantive questions to illustrate, prediction of breast cancer conditional on multiple patient covariates and prediction of life span conditional on demographic covariates and hypertension status. I discuss directions for future research here.

I anticipate that econometricians and other readers who may not be specifically concerned with medical decision making will nonetheless find the paper useful. As mentioned in the Introduction, many decision makers in realms other than health care face the ecological inference problem as they attempt to predict person-specific outcomes. I cited judges deciding how to sentence convicted offenders, lenders considering loan applications, firms deciding whether to employ job applicants, and colleges deciding whether to admit student applicants.

As with clinicians, these and other decision makers have in the past chosen between actuarial

prediction and exercise of subjective clinical judgment. The analysis of this paper suggests that it may be possible to improve on both polar options by formalizing clinical judgment. There are rich possibilities for future research, but there also are challenges. One issue is that evidence-based predictors P(y|x) of non-health outcomes often are proprietary rather than in the public domain. Another is that auxiliary data revealing the distribution P(w|x) may be less available that it is in health care.

There is considerable scope for new methodological research that would usefully extend the present analysis. To focus attention on the ecological inference problem, I assumed complete knowledge of P(y|x) and P(w|x) throughout the paper. In practice, a decision maker may have only finite-sample estimates of these distributions and thus have to cope with sampling imprecision as well as the ecological inference problem. It may be possible to draw on the literature on statistical inference for partially identified parameters to develop confidence sets for features of P(y|x, w). However, a decision maker's objective is to make good decisions, not to perform statistical inference. Wald's statistical decision theory can potentially inform this objective. A few studies to date have applied statistical decision theory to settings where partial identification stems from the selection problem (Manski , 2007b; Stoye, 2012; Tetenov, 2013; Dominitz and Manski, 2017), but none when the source of ambiguity is the ecological inference problem.

Another open methodological question is characterization of the identification region for P(y|x, w) and study of decision making under ambiguity when the outcome is multi-dimensional. The starting point for consideration of ecological inference is the Law of Total Probability (1) whatever the outcome space may be. However, the analysis in Sections 2 and 3 restricted attention to settings in which y is binary or real-valued. This analysis suffices for some important applications but not for others. In particular, it does not cover settings in which a decision maker has to predict outcomes y(t), $t \in T$ for a vector of potential outcomes, T being the set of feasible treatments. Manski (2000) reported some findings for social planning when there are two treatments, the potential outcomes are binary (that is, treatment success or failure), the covariate w is binary, and no structural assumptions are imposed. Research studying more general settings with multi-dimensional outcomes would be useful.

Appendix A: Basic Bounds on Means and Quantiles, with Application to Prediction of Life Expectancy

A.1. Form of the Bounds

Consider $E(y|x = k, w = j)$ 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 Horowitz and Manski (1995) find it easier to prove a general result for parameters that respect stochastic dominance. They then apply this result to the mean and quantiles.

To simplify notation, let $p_{jk} \equiv P(w \neq j|x = k)$. Horowitz and Manski show that the identification region for $P(y|x = k, w = j)$ contains a "smallest" member L_{jk} that is stochastically dominated by all other feasible values of $P(y|x = k, w = j)$ and a "largest" member U_{jk} that stochastically dominates all other feasible values of $P(y|x = k, w = j)$. These distributions are truncated versions of $P(y|x = k)$: L_{jk} right-truncates $P(y|x = k)$ at its $(1 - p_{jk})$–quantile and U_{jk} left-truncates $P(y|x = k)$ at its p_{jk}–quantile. Formally, L_{jk} and U_{jk} are defined as follows (Horowitz and Manski, 1995, Proposition 4):

$$\begin{aligned} L_{jk}[-\infty, t] &\equiv P(y \leq t|x = k)/(1 - p_{jk}) && \text{for } t < Q_{(1-p_{jk})}(y|x = k), \\ &\equiv 1 && \text{for } t \geq Q_{(1-p_{jk})}(y|x = k). \end{aligned}$$

$$\begin{aligned} U_{jk}[-\infty, t] &\equiv 0 && \text{for } t < Q_{p_{jk}}(y|x = k), \\ &\equiv [P(y \leq t|x = k) - p_{jk}]/(1 - p_{jk}) && \text{for } t \geq Q_{p_{jk}}(y|x = k). \end{aligned}$$

With this background, it follows that if $D(\cdot)$ is a parameter that respects stochastic dominance, the smallest feasible value of $D[P(y|x = k, w = j)]$ is $D(L_{jk})$ and the largest feasible value is $D(U_{jk})$. Thus, sharp bounds on $E(y|x = k, w = j)$ are $E(L_{jk})$ and $E(U_{jk})$, and sharp bounds on $Q_\alpha(y|x = k, w = j)$ are $Q_\alpha(L_{jk})$ and $Q_\alpha(U_{jk})$. In the case of the mean, all values between the bounds are feasible and, hence, the identification region is $[E(L_{jk}), E(U_{jk})]$.

A.2. Application to Prediction of Life Expectancy Conditional on Hypertension Status in Section 2.1.2

Let remaining life span y be integer-valued. Consider a specified racial subset of males who survive at least to age 50. Let $y^* = y + 50$ denote complete life span. It is simplest to consider y^* for most of the derivation and transform to y near the end.

The distribution of y^* in the specified group is $P(y^* \leq A | y^* \geq 50)$. Its q-quantile is

$$\begin{aligned}\min A: P(y^* \leq A | y^* \geq 50) \geq q &= \min A: P(50 \leq y^* \leq A)/P(y^* \geq 50) \geq q \\ &= \min A: P(50 \leq y^* \leq A) \geq q \cdot P(y^* \geq 50) \\ &= \min A: P(y^* \geq 50) - P(y^* \geq A + 1) \geq q \cdot P(y^* \geq 50) \\ &= \min A: P(y^* \geq A + 1) \leq (1 - q) P(y^* \geq 50).\end{aligned}$$

Writing the q-quantile in the final form is useful because column 2 of the life table considers a population of initial size 100,000 and shows the number who survive at least A years for all values of A. $P(y^* \geq A)$ is the number of survivors to age A divided by 100,000.

The distributions L and U are determined by the $(1 - p)$ and p quantiles of $P(y^* | y^* \geq 50)$. These are

$$\begin{aligned}Q_{(1-p)} &= \min A: P(y^* \geq A + 1) \leq p \cdot P(y^* \geq 50), \\ Q_p &= \min A: P(y^* \geq A + 1) \leq (1 - p) \cdot P(y^* \geq 50).\end{aligned}$$

The bounds on $E(y^* | x)$ are E(L) and E(U).

The fact that outcome y^* is integer-valued makes exact computation of E(L) and E(U) subtle because only part of the probability masses occurring at the truncation points are used to form L and U. Reasonably narrow bounds on E(L) and E(U) are easy to compute. These are

$E(y^*| 50 \le y^* \le Q_{(1-p)} - 1) \le E(L) \le E(y^*| 50 \le y^* \le Q_{(1-p)})$,

$E(y^*|Q_p \le y^*) \le E(U) \le E(y^*|Q_p + 1 \le y^*)$.

My computations use the conservative lower bound on E(L) and upper bound on E(U).

The derivation below explains computation of these bounds when x = (age 50, NH black male, not HBP). The derivation is analogous for other values of x.

For the specified value of x, $P(y^* \ge 50) = 0.892$, $p = 0.426$, $1 - p = 0.574$, $(1 - p)P(y^* \ge 50) = 0.512$, and $p \cdot P(y^* \ge 50) = 0.380$.

To obtain the conservative lower bound on E(L), one uses the life table to determine that $Q_{(1-p)} = 80$. This is so because $P(y^* \ge 81) = 0.357 \le 0.380$ and $P(y^* \ge 80) = 0.383 > 0.380$. Hence, the conservative lower bound is $E(y^*|50 \le y^* \le 79)$.

Now consider computation of $E(y|50 \le y^* \le 79)$. The life table shows that, given 100,000 initial persons, 89,156 have life span at least 50 and 38,383 have life span at least 80. Hence, 89156 – 38383 = 50773 have life span in the interval [50, 79]. The number of person years beyond 50 lived by persons who survive to 50 or more is 2,373,723. The number of person years beyond 50 lived by persons who survive to 80 or more is $30 \cdot 38383 + 305{,}178 = 1{,}456{,}668$. Hence, the number of person years beyond 50 lived by those with life span in the interval [50, 79] is 2,373,723 – 1,456,668 = 917055. Hence, $E(y|50 \le y^* \le 79) = 917055/50773 = 18.1$.

To obtain the conservative upper bound on E(U), one uses the life table to determine that $Q_p = 74$. This is so because $P(y^* \ge 75) = 0.511 \le 0.512$ and $P(y^* \ge 74) = 0.535 > 0.512$. Hence, the conservative upper bound on $E(y^*|x)$ is $E(y^*|75 \le y^*)$.

Now consider computation of $E(y|75 \le y^*)$. The life table shows that, given 100,000 initial persons, 51097 have life span in the interval [75, ∞]. The number of person years beyond 75 lived by these persons is 529,456. Hence, $E(y|75 \le y^*) = 25 + 529456/51097 = 35.4$.

Appendix B. Identification with Parametric Models

Some authors point-identify $P(y|x, w)$ by asserting a parametric model that places this distribution in a finite-dimensional family. 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 Y, and assume there exists a unique $\theta \in \Theta$ such that

(B1) $P(y|x = k, w = j) = F(k, j, \theta)$, all $(k, j) \in X \times W$.

Combining the Law of Total Probability (1) with assumption (B1) yields

(B2) $$P(y\,|\,x = k) = \sum_{j \in W} F(k, j, \theta)P(w = j\,|\,x = k), \quad k \in X.$$

For each $k \in X$, the data reveal $P(y\,|\,x = k)$ and $[P(w = j\,|\,x = k), j \in W]$. Hence, (B2) is a system of $|X|$ distributional equations restricting 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 (4) yields

(B3) $$E(y\,|\,x = k) = \sum_{j \in W} e(k, j, \theta)P(w = j\,|\,x = k), \quad k \in X.$$

System (B3) is similar to Goodman's system (5) except that $e(k, j, \theta)$ generally varies nonlinearly with θ. Nonlinearity in θ implies that solution of (B3) is more complex than is 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 (B1) 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. A model that received considerable attention was proposed 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; Cho and Gaines, 2004). Wakefield (2008) cautions against application of King's model or other parametric models to public health research.

Illustration

Organizations developing clinical practice guidelines for breast cancer screening have acknowledged 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) states (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 teams made varying predictions of life expectancy after age 40 by combining the shared data with auxiliary data and alternative structural assumptions strong enough to yield point identification.

References

Arias, E. (2015), *United States Life Tables, 2011*, National Vital statistics Reports, Vol. 64, No. 11, Centers for Disease Control and Prevention, www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_11.pdf, accessed May 2, 2017.

Berger, J. (1985), *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag.

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.

Dominitz, J. and C. Manski (2017), "More Data or Better Data? A Statistical Decision Problem," *Review of Economic Studies*, DOI: https://doi.org/10.1093/restud/rdx005.

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.

Go, A., D. Mozaffarian, V. Roger, E. Benjamin, J. Berry, W. Borden, D. Bravata, S. Dai, E. Ford, C. Fox, S. Franco, H. Fullerton, C. Gillespie, S. Hailpern, J. Heit, V. Howard, M. Huffman, B. Kissela, S. Kittner, D. Lackland, J. Lichtman, L. Lisabeth, D. Magid, G. Marcus, A. Marelli, D. Matchar, D. McGuire, E. Mohler, C. Moy, M. Mussolino, G. Nichol, N. Paynter, P. Schreiner, P. Sorlie, J. Stein, T. Turan, S. Virani, N. Wong, D. Woo, and M. Turner; on behalf of the American Heart Association Statistics Committee and Stroke Statistics Subcommittee (2013), "Heart disease and stroke statistics—2013 update: a report from the American Heart Association," *Circulation*, 127, e6-e245.

Good, I. (1967), "On the Principle of Total Evidence," *The British Journal for the Philosophy of Science*, 17, 319-321.

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

Groves, W., D. Zald, B. Lebow, B. Snitz, and C. Nelson (2000)," Clinical Versus Mechanical Prediction: A Meta-Analysis," *Psychological Assessment*, 12, 19-30.

Grünwald, P. and J. Halpern (2004), "When Ignorance is Bliss," in *Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence*, 226-234.

Herron, T., T. Seidenfeld, and L. Wasserman (1997), "Divisive Conditioning: Further Results on Dilation,", *Philosophy of Science,* 64, 411–444.

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

Kadane, J., M. Shervish, and T. Seidenfeld (2008), "Is Ignorance Bliss?" *Journal of Philosophy*, 105, 5-36.

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

Manski, C. (2007b), "Minimax-Regret Treatment Choice with Missing Outcome Data," *Journal of Econometrics*, 139, 105-115.

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. (2017), "Mandating Vaccination with Unknown Indirect Effects," *Journal of Public Economic Theory*, DOI:10.1111/jpet.12234.

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 (2017), "How Do Right-to-Carry Laws Affect Crime Rates? Coping with Ambiguity Using Bounded-Variation Assumptions," *Review of Economics and Statistics*, forthcoming.

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," *Journal of the American Medical Association,* 314, 1615-1634.

National Cancer Institute (2016), *Breast Cancer Risk Assessment Tool*, http://www.cancer.gov/bcrisktool/, accessed May 2, 2017.

National Comprehensive Cancer Network (2016), *Breast Cancer Screening and Diagnosis*, Version 1.2016, www.nccn.org/professionals/physician_gls/pdf/breast-screening.pdf, accessed May 2, 2017.

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

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

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.

Stoye, J. (2012), "Minimax Regret Treatment Choice with Covariates or with Limited Validity of Experiments," *Journal of Econometrics*, 166, 157-165.

Tetenov, A. (2012), "Measuring Precision of Statistical Inference on Partially Identified Parameters,"

Collegio Carlo Alberto, Moncalieri (Torino), Italy.

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 May 2, 2017.

Susan G. Komen (2016), *Estimating Breast Cancer Risk*, ww5.komen.org/BreastCancer/GailAssessmentModel.html, accessed May 2, 2017.

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.