

CONTENTS

Preface	xiii
Introduction	1
<i>Surveillance or Aggressive Treatment</i>	2
<i>Evolution of the Book</i>	3
<i>Summary</i>	6
1 Clinical Guidelines and Clinical Judgment	8
1.1. <i>Adherence to Guidelines or Exercise of Judgment?</i>	8
Variation in Guidelines	10
Case Study: Nodal Observation or Dissection in Treatment of Melanoma	11
1.2. <i>Degrees of Personalized Medicine</i>	16
Prediction of Cardiovascular Disease	17
The Breast Cancer Risk Assessment Tool	17
Predicting Unrealistically Precise Probabilities	18
1.3. <i>Optimal Care Assuming Rational Expectations</i>	20
Optimal Choice between Surveillance and Aggressive Treatment	21
1.4. <i>Psychological Research Comparing Evidence-Based Prediction and Clinical Judgment</i>	21
1.5. <i>Second-Best Welfare Comparison of Adherence to Guidelines and Clinical Judgment</i>	24
Surveillance or Aggressive Treatment of Women at Risk of Breast Cancer	25

viii CONTENTS

2	Wishful Extrapolation from Research to Patient Care	27
2.1.	<i>From Study Populations to Patient Populations</i>	28
	Trials of Drug Treatments for Hypertension	30
	Campbell and the Primacy of Internal Validity	31
2.2.	<i>From Experimental Treatments to Clinical Treatments</i>	32
	Intensity of Treatment	32
	Blinding in Drug Trials	33
2.3.	<i>From Measured Outcomes to Patient Welfare</i>	35
	Interpreting Surrogate Outcomes	35
	Assessing Multiple Outcomes	36
2.4.	<i>From Hypothesis Tests to Treatment Decisions</i>	38
	Using Hypothesis Tests to Compare Treatments	38
	Using Hypothesis Tests to Choose When to Report Findings	40
2.5.	<i>Wishful Meta-Analysis of Disparate Studies</i>	41
	A Meta-Analysis of Outcomes of Bariatric Surgery	43
	The Misleading Rhetoric of Meta-Analysis	43
	The Algebraic Wisdom of Crowds	44
2.6.	<i>Sacrificing Relevance for Certitude</i>	45
3	Credible Use of Evidence to Inform Patient Care	47
3.1.	<i>Identification of Treatment Response</i>	48
	Unobservability of Counterfactual Treatment Outcomes	48
	Trial Data	48
	Observational Data	49
	Trials with Imperfect Compliance	51
	Extrapolation Problems	51
	Missing Data and Measurement Errors	52
3.2.	<i>Studying Identification</i>	52
3.3.	<i>Identification with Missing Data on Patient Outcomes or Attributes</i>	53
	Missing Data in a Trial of Treatments for Hypertension	55

	Missing Data on Family Size When Predicting Genetic Mutations	57
3.4.	<i>Partial Personalized Risk Assessment</i>	59
	Predicting Mean Remaining Life Span	59
3.5.	<i>Credible Inference with Observational Data</i>	61
	Bounds with No Knowledge of Counterfactual Outcomes	61
	Sentencing and Recidivism	62
	Assumptions Using Instrumental Variables	63
	Case Study: Bounding the Mortality Effects of Swan-Ganz Catheterization	65
3.6.	<i>Identification of Response to Testing and Treatment</i>	66
	Optimal Testing and Treatment	67
	Identification of Testing and Treatment Response with Observational Data	68
	Measuring the Accuracy of Diagnostic Tests	69
3.7.	<i>Prediction Combining Multiple Studies</i>	70
	Combining Multiple Breast Cancer Risk Assessments	71
	Combining Partial Predictions	72
4	Reasonable Care under Uncertainty	74
4.1.	<i>Qualitative Recognition of Uncertainty</i>	74
4.2.	<i>Formalizing Uncertainty</i>	76
	States of Nature	77
4.3.	<i>Optimal and Reasonable Decisions</i>	78
4.4.	<i>Reasonable Decision Criteria</i>	80
	Decisions with Rational Expectations	80
	Maximization of Subjective Expected Utility	80
	Decisions under Ambiguity: The Maximin and Minimax-Regret Criteria	81
4.5.	<i>Reasonable Choice between Surveillance and Aggressive Treatment</i>	83
4.6.	<i>Uncertainty about Patient Welfare</i>	84

x CONTENTS

- 5 Reasonable Care with Sample Data 87
 - 5.1. *Principles of Statistical Decision Theory* 88
 - Some History, Post-Wald 90
 - 5.2. *Recent Work on Statistical Decision Theory for Treatment Choice* 91
 - Practical Appeal 92
 - Conceptual Appeal 92
 - 5.3. *Designing Trials to Enable Near-Optimal Treatment Choice* 94
 - Using Power Calculations to Choose Sample Size 94
 - Sample Size Enabling Near-Optimal Treatment Choice 95
 - Choosing the Near-Optimality Threshold 96
 - Findings with Binary Outcomes, Two Treatments, and Balanced Designs 97
 - Implications for Practice 98
 - 5.4. *Reconsidering Sample Size in the MSLT-II Trial* 98
- 6 A Population Health Perspective on Reasonable Care 100
 - 6.1. *Treatment Diversification* 101
 - Treating X-Pox 102
 - 6.2. *Adaptive Diversification* 104
 - Adaptive Treatment of a Life-Threatening Disease 105
 - 6.3. *The Practicality of Adaptive Diversification* 106
 - Implementation in Centralized Health-Care Systems 106
 - Should Guidelines Encourage Treatment Variation under Uncertainty? 107
- 7 Managing Uncertainty in Drug Approval 108
 - 7.1. *The FDA Approval Process* 108
 - 7.2. *Type I and II Errors in Drug Approval* 109
 - 7.3. *Errors Due to Statistical Imprecision and Wishful Extrapolation* 110
 - 7.4. *FDA Rejection of Formal Decision Analysis* 111

<i>7.5. Adaptive Partial Drug Approval</i>	114
Adaptive Limited-Term Sales Licenses	114
Open Questions	116
Conclusion	117
<i>Separating the Information and Recommendation Aspects of Guidelines</i>	118
<i>Educating Clinicians in Care under Uncertainty</i>	119
Complement 1A. Overview of Research on Lymph Node Dissection	120
1A.1. <i>Sentinel Lymph Node Biopsy</i>	120
1A.2. <i>Observation or SLN Biopsy</i>	123
1A.3. <i>Observation or Dissection after Positive SLN Biopsy</i>	126
Complement 1B. Formalization of Optimal Choice between Surveillance and Aggressive Treatment	129
1B.1. <i>Aggressive Treatment Prevents Disease</i>	130
1B.2. <i>Aggressive Treatment Reduces the Severity of Disease</i>	131
Complement 2A. Odds Ratios and Health Risks	132
Complement 3A. The Ecological Inference Problem in Personalized Risk Assessment	134
Complement 3B. Bounds on Success Probabilities with No Knowledge of Counterfactual Outcomes	136
Complement 4A. Formalization of Reasonable Choice between Surveillance and Aggressive Treatment	138
Complement 5A. Treatment Choice as a Statistical Decision Problem	140
5A.1. <i>Choice between a Status Quo Treatment and an Innovation When Outcomes Are Binary</i>	141
Complement 6A. Minimax-Regret Allocation of Patients to Two Treatments	144

xii CONTENTS

Complement 6B. Derivations for Criteria to Treat X-Pox 146

6B.1. Maximization of Subjective Expected Welfare 146

6B.2. Maximin 146

6B.3. Minimax Regret 146

References 149

Index 159

Introduction

There are three broad branches of decision analysis: normative, descriptive, and prescriptive. Normative analysis seeks to establish ideal properties of decision making, often aiming to give meaning to the terms “optimal” and “rational.” Descriptive analysis seeks to understand and predict how actual decision makers behave. Prescriptive analysis seeks to improve the performance of actual decision making.

One might view normative and descriptive analysis as entirely distinct subjects. It is not possible, however, to cleanly separate prescriptive analysis from the other branches of study. Prescriptive analysis aims to improve actual decisions, so it must draw on normative thinking to define “improve” and on descriptive research to characterize actual decisions.

This book offers prescriptive analysis that seeks to improve patient care. My focus is decision making under uncertainty regarding patient health status and response to treatment. By “uncertainty,” I do not just mean that clinicians and health planners may make probabilistic rather than definite predictions of patient outcomes. My main concern is decision making when the available evidence and medical knowledge do not suffice to yield precise probabilistic predictions.

For example, an educated patient who is comfortable with probabilistic thinking may ask her clinician a seemingly straightforward question such as “What is the chance that I will develop disease X in the next five years?” or “What is the chance that treatment Y will cure me?” Yet the clinician may not be able to provide precise answers to these questions. A credible response may be a range, say “20 to 40 percent” or “at least 50 percent.”

2 INTRODUCTION

Decision theorists use the terms “deep uncertainty” and “ambiguity” to describe the decision settings I address, but I shall encompass them within the broader term “uncertainty” for now. Uncertainty in patient care is common and has sometimes been acknowledged verbally. For example, the Evidence-Based Medicine Working Group asserts that (Institute of Medicine, 2011, p. 33): “clinicians must accept uncertainty and the notion that clinical decisions are often made with scant knowledge of their true impact.” However, uncertainty has generally not been addressed in research on evidence-based medicine, which has been grounded in classical statistical theory. I think this a huge omission, which this book strives to correct.

Surveillance or Aggressive Treatment

I pay considerable attention to the large class of decisions that choose between surveillance and aggressive treatment of patients at risk of potential disease. Consider, for example, women at risk of breast cancer. In this instance, surveillance typically means undergoing periodic mammograms and clinical exams, while aggressive treatment may mean preventive drug treatment or mastectomy.

Other familiar examples are choice between surveillance and drug treatment for patients at risk of heart disease or diabetes. Yet others are choice between surveillance and aggressive treatment of patients who have been treated for localized cancer and are at risk of metastasis. A semantically distinct but logically equivalent decision is choice between diagnosis of patients as healthy or ill. With diagnosis, the concern is not to judge whether a patient will develop a disease in the future but whether the patient is currently ill and requires treatment.

These decisions are common, important to health, and familiar to clinicians and patients alike. Indeed, patients make their own choices related to surveillance and aggressive treatment. They perform self-surveillance by monitoring their own health status. They choose how faithfully to adhere to surveillance schedules and treatment regimens prescribed by clinicians.

Uncertainty often looms large when a clinician contemplates choice between surveillance and aggressive treatment. The effectiveness of surveillance in mitigating the risk of disease may depend on the degree to which a patient will adhere to the schedule of clinic visits prescribed in a surveillance plan. Aggressive treatment may be more beneficial than surveillance to the extent that it reduces the risk of disease development or

the severity of disease that does develop. It may be more harmful to the extent that it generates health side effects and financial costs beyond those associated with surveillance. There often is substantial uncertainty about all these matters.

Evolution of the Book

I am an economist with specialization in econometrics. I have no formal training in medicine. One may naturally ask how I developed an interest in patient care under uncertainty and feel able to contribute to the subject. It would be arrogant and foolhardy for me to dispense medical advice regarding specific aspects of patient care. I will not do so. The contributions that I feel able to make concern the methodology of evidence-based medicine. This matter lies within the expertise of econometricians, statisticians, and decision analysts.

Research on treatment response and risk assessment shares a common objective: probabilistic prediction of patient outcomes given knowledge of observed patient attributes. Development of methodology for prediction of outcomes conditional on observed attributes has long been a core concern of many academic disciplines.

Econometricians and statisticians commonly refer to conditional prediction as *regression*, a term in use since the nineteenth century. Some psychologists have used the terms *actuarial prediction* and *statistical prediction*. Computer scientists may refer to *machine learning* and *artificial intelligence*. Researchers in business schools may speak of *predictive analytics*. All these terms are used to describe methods that have been developed to enable conditional prediction.

As an econometrician, I have studied how statistical imprecision and identification problems affect empirical (or evidence-based) research that uses sample data to predict population outcomes. Statistical theory characterizes the imprecise inferences that can be drawn about the outcome distribution in a study population by observing the outcomes of a finite sample of its members. Identification problems are inferential difficulties that persist even when sample size grows without bound.

A classic example of statistical imprecision occurs when one draws a random sample of a population and uses the sample average of an outcome to estimate the population mean outcome. Statisticians typically measure imprecision of the estimate by its variance, which decreases to zero as sample size increases. Whether imprecision is measured by variance or another

4 INTRODUCTION

way, the famous “Laws of Large Numbers” imply that imprecision vanishes as sample size increases.

Identification problems encompass the spectrum of issues that are sometimes called *non-sampling errors* or *data-quality problems*. These issues cannot be resolved by amassing so-called big data. They may be mitigated by collecting better data, but not by merely collecting more data.

A classic example of an identification problem is generated by missing data. Suppose that one draws a random sample of a population, but one observes only some sample outcomes. Increasing sample size adds new observations, but it also yields further missing data. Unless one learns the values of the missing data or knows the process that generates missing data, one cannot precisely learn the population mean outcome as sample size increases.

My research has focused mainly on identification problems, which often are the dominant difficulty in empirical research. I have studied probabilistic prediction of outcomes when available data are combined with relatively weak assumptions that have some claim to credibility. While much of this work has necessarily been technical, I have persistently stressed the simple truth that research cannot yield decision-relevant findings based on evidence alone.

In Manski (2013a) I observed that the logic of empirical inference is summarized by the relationship:

assumptions + data \Rightarrow conclusions.

Data (or evidence) alone do not suffice to draw useful conclusions. Inference also requires assumptions (or theories, hypotheses, premises, suppositions) that relate the data to the population of interest. Holding fixed the available data, and presuming avoidance of errors in logic, stronger assumptions yield stronger conclusions. At the extreme, one may achieve certitude by posing sufficiently strong assumptions. A fundamental difficulty of empirical research is to decide what assumptions to maintain.

Strong conclusions are desirable, so one may be tempted to maintain strong assumptions. I have emphasized that there is a tension between the strength of assumptions and their credibility, calling this (Manski, 2003, p. 1):

The Law of Decreasing Credibility: The credibility of inference decreases with the strength of the assumptions maintained.

This “Law” implies that analysts face a dilemma as they decide what assumptions to maintain: Stronger assumptions yield conclusions that are more powerful but less credible.

I have argued against making precise probabilistic predictions with *incredible certitude*. It has been common for experts to assert that some event will occur with a precisely stated probability. However, such predictions often are fragile, resting on unsupported assumptions and limited data. Thus, the expressed certitude is not credible.

Motivated by these broad ideas, I have studied many prediction problems and have repeatedly found that empirical research may be able to credibly bound the probability that an event will occur but not make credible precise probabilistic predictions, even with large data samples. In econometrics jargon, probabilities of future events may be *partially identified* rather than *point identified*. This work, which began in the late 1980s, has been published in numerous journal articles and synthesized in multiple books, written at successive stages of my research program and at technical levels suitable for different audiences (Manski, 1995, 2003, 2005, 2007a, 2013a).

Whereas my early research focused on probabilistic prediction per se, I have over time extended its scope to study decision making under uncertainty; that is, decisions when credible precise probabilistic predictions are not available. Thus, my research has expanded from econometrics to prescriptive decision analysis.

Elementary decision theory suggests a two-step process for choice under uncertainty. Considering the feasible alternatives, the first step is to eliminate dominated actions—an action is dominated if one knows for sure that some other action is superior. The second step is to choose an undominated action. This is subtle because there is no consensus regarding the optimal way to choose among undominated alternatives. There are only various reasonable ways. I will later give content to the word “reasonable.”

Decision theory is mathematically rigorous, but it can appear sterile when presented in abstraction. The subject comes alive when applied to important actual decision problems. I have studied various public and private decisions under uncertainty. This work has yielded technical research articles and a book on public policy under uncertainty written for a broad audience (Manski, 2013a).

I have increasingly felt that patient care is ripe for study as a problem of decision making under uncertainty. I therefore have sought to learn enough about research on evidence-based medicine to make original contributions that build on my methodological background in econometrics

6 INTRODUCTION

and decision analysis. The results include studies of diagnostic testing and treatment under uncertainty (Manski, 2009, 2013b), personalized care with partial assessment of health risks (Manski, 2018a), analysis and design of randomized clinical trials (Manski, 2004a; Manski and Tetenov, 2016, 2019), drug approval (Manski, 2009), and vaccination policy with partial knowledge of disease transmission (Manski, 2010, 2017). I have also written a review article (Manski, 2018b).

The idea of writing a book evolved as I have accumulated background in evidence-based medicine and have developed an enlarging set of original research findings. A book provides the space to present major themes and to show how they become manifest in various contexts. A book enables an author to speak to a broader audience than is possible when writing research articles on particular topics.

I hope that this book will prove useful to a spectrum of readers. I would like it to help clinicians and public health planners recognize and cope with uncertainty as they make decisions about patient care. It may help patients to become informed about and participate in their own care. I anticipate that the book will help medical researchers design randomized trials and interpret the evidence they obtain from trials and observational studies. I will be pleased if the book encourages the biostatisticians who assist medical researchers to make constructive use of modern methodological advances in econometrics and statistical decision theory.

Some readers with certain types of expertise will correctly view the book as critical of the methodologies they have advocated. These include biostatisticians who have used the statistical theory of hypothesis testing to advise medical researchers on the design and analysis of randomized trials. They include personnel at the US Food and Drug Administration and other governmental agencies who regulate approval of new drugs, biologics, and medical devices. They include developers of clinical practice guidelines who have argued that evidence-based medicine should rest either solely or predominately on evidence from randomized trials, disregarding or downplaying evidence from observational studies. I hope that these readers will make the effort to understand the bases for my criticisms and that they will view the prescriptive decision analysis presented here as constructive suggestions.

Summary

Aiming to make the book accessible to a wide readership, the exposition in the main text is almost entirely verbal rather than mathematical. For readers

who want to dig deeper, I include a set of complements that formalize or elaborate on key parts of the discussion in the main text. I also provide references to the technical articles that present the full analysis.

The eight chapters of the book move from review and critique in chapters 1 and 2 to prescription in chapters 3 through 7 and conclusion. Chapter 1 reviews the continuing discourse in medicine regarding the circumstances in which clinicians should adhere to evidence-based practice guidelines or exercise their own judgment, sometimes called “expert opinion.” Chapter 2 critiques how evidence from randomized trials has been used to inform medical decision making.

Chapter 3 describes research on identification, whose aim is credible use of evidence to inform patient care. Chapter 4 develops decision-theoretic principles for reasonable care under uncertainty. Chapter 5 considers reasonable decision making with sample data from randomized trials. Moving away from consideration of a clinician treating an individual patient, chapter 6 views patient care from a population health perspective. Chapter 7 considers management of uncertainty in drug approval. The final chapter provides concluding suggestions that encourage putting the themes of the book into practice.

INDEX

- actuarial prediction, 3
- adaptive diversification, 104–6; in centralized health-care systems, 106–7; of drug treatment, 114–16; practicality of, 106–7
- adaptive minimax-regret (AMR) criterion, 104–6; FDA computation of treatment allocation of, 115
- adaptive partial drug approval, 114–16
- aggressive treatment: formalization of optimal choice of *vs.* surveillance, 129–31; formalization of reasonable choice between surveillance and, 138–39; with positive testing (ATPT), 68–69; preventing disease, 130–31; reducing severity of disease, 131; *vs.* surveillance, 2–3, 21, 25–26, 83–84
- Altman, D., 70, 95
- ambiguity, 2; decisions under, 81–83
- American Society of Clinical Oncology (ASCO) recommendations, 26
- Amir, E., 58, 71
- anticoagulation drug trial, primary and secondary endpoints for, 37
- Arias, E., 59
- ASCVD Risk Estimator, 17; unrealistically precise predictions with, 18–19
- assumptions, strength *vs.* credibility of, 4–5
- atherosclerotic cardiovascular disease (ASCVD) risk, 17
- Atlas of Variation in Healthcare* (UK), 9
- attributable risk, 132–33
- average treatment effect (ATE): bounds on, 61–62; calculation of, 140
- Balke, A., 65
- bariatric surgery, meta-analysis of outcomes of, 43
- Barsky, R., 46
- Basu, A., 21, 84–85
- Bayes, Thomas, 80
- Bayes risk, 89
- Bayes rule, 142–43
- Bayesian decision making, 80, 89
- Bayesian process, 89
- Beckett, N., 30
- Beckett trial, 30–31, 33; duration of, 36
- Bentham, Jeremy, 20
- Berger, J., 88
- Berkson, J., 133
- Berry, D., 91
- Bhattacharya, J., 66
- bias: risk of, 50; of sample population estimates, 50
- binary outcomes, 132; bounds for success in, 61–62; findings with, 97–98; status quo treatment *vs.* innovation with, 141–43
- binary risk factor, 132
- Bland, J., 70
- blinding, in drug trials, 33–34
- Blümle, A., 30
- BOADICEA Model, 71
- Boland, G., 14
- Boston Area Anticoagulation Trial for Atrial Fibrillation Investigators, 37
- bounded-variation assumptions, 60
- BRCA mutations, meta-analysis of breast and ovarian cancer risk with, 44
- BRCAPRO Model, 71
- breast cancer risk: bounding, 83–84; eight attributes in, 17–18; meta-analysis of, 44; surveillance *vs.* aggressive treatment for, 25–26
- Breast Cancer Risk Assessment (BCRA) Tool, 17–20, 26; unrealistically precise predictions with, 18–19
- breast cancer risk assessments, combining multiple, 71–72
- Buchwald, H., 43
- Camerer, C., 22
- Campbell, Donald, 31–32
- cancer drugs: European Union approval of, 35–36; U.S. and European approval of, 36
- Canner, P., 91

160 INDEX

- cardiovascular disease, prediction of, 17
case-control sampling, 132
Caulley, L., 14
centralized health-care systems, adaptive
 diversification in, 106–7
certitude, sacrificing relevance for, 45–46
Chen, S., 44
Cheng, Y., 91
Cheville, A., 99
choice-based sampling, 132
Chomsky, Noam, 46
Claus Model, 71, 72
Clemen, R., 45
clinical breast examination (CBE), variations
 in guidelines for, 10
clinical judgment: *vs.* adherence to guidelines,
 8–15; *vs.* evidence-based prediction,
 21–24; integrating with evidence-based
 research, 23–24; second-best welfare
 comparison of with adherence to guide-
 lines, 24–26
clinical practice guidelines (CPGs), 8–9;
 for breast cancer screening, 18; for rat-
 ing treatment recommendations, 75–76;
 rigorous, 76
clinical significance, 96
clinical treatments, from experimental treat-
 ments to, 32–34
clinicians, education in care under uncer-
 tainty, 119
CLND (completion lymph node dissection),
 128
*Cochrane Handbook for Systematic Reviews
 of Interventions*: on risk of bias, 50; on
 trials, 27–28
combined findings, 41–42
comorbidities, extrapolating trial finding to
 patients with, 28–29
compliance: imperfect, 51; perfect, 49
conditional prediction, 3
confidence intervals, 87–88
Connors, A., 65–66
consensus estimates, 44
conventional asymmetric error
 probabilities, 39
Cook, R., 39, 65
Cornfield, J., 133
cost-effectiveness analysis, 85
counterfactual outcomes: bounds with no
 knowledge of, 61–62; with observational
 data, 68–69; unobservability of, 48,
 49, 50
credibility, *vs.* strength of assumptions, 4–5
credible bounds, identification of, 53
credible inference, with observational data,
 61–66
Crits-Christoph, P., 40–41
crowds, algebraic wisdom of, 44–45
data-invariant rules, 142
data-quality problems, 4
Davis, C., 35–36, 111, 135
Dawes, R., 22–23
decentralized decision making: *vs.* adher-
 ence to guidelines, 9–10; using clinical
 judgment, 24
decision analysis, branches of, 1
decision making: under ambiguity, 81–83;
 criteria of for treating new disease,
 146–47; optimal and reasonable, 78–79;
 state of nature in, 77–78; with subjective
 distribution, 80, 83; three-step process
 of, 88–89; under uncertainty, 5–6
decision theory, 5; in public health
 treatment, 102–3
deep uncertainty, 2
DeGroot, M., 91
Demets, D., 35, 111
DerSimonian, R., 42–44, 70
descriptive analysis, 1
diagnostic tests, 67; measuring accuracy of,
 69–70
disease development, personalized prob-
 abilities of, 16
disease-free survival rate, definition of, 124
disease risk, bounding, 83–84
disease-specific survival rate, 124
disparate studies, wishful meta-analyses of,
 41–45
distant disease-free survival, 124
diversification: adaptive, 104–7; of
 treatment, 101–4
Domchek, S., 71, 72
dominated actions, eliminating, 79
drug approval: adaptive partial, 114–16;
 errors in due to statistical imprecision
 and wishful extrapolation, 110–11; lon-
 ger Phase 3 trials in, 114–15; managing
 uncertainty in, 108–16; process of, 108–9;
 type I and II errors in, 109–10, 114
drug licensing process, 114–15
drug trials: assessing multiple outcomes of,
 36–37; blinding in, 33–34; for hyperten-
 sion, 30–31; intensity of treatment in,
 32–33; interpreting surrogate outcomes
 of, 35–36; Phase 3, 108–9, 114–15; phases
 of, 108–9; primary and secondary end-
 points in, 37

- drugs: adaptive limited-term sales licenses for, 114–16; post-market surveillance of, 115
- Dummer, R., 15
- Duncan, O., 135
- Dunnnett, C., 39
- ecological inference problem, in personalized risk assessment, 134–35
- Eddy, David, 77
- effectiveness, *vs.* efficacy, 34
- efficacy of treatment, 34
- Eggermont, A., 15
- Eichler, H., 114
- empirical success rule, 92, 97–98, 142
- errors: magnitudes of losses with, 39; Type I, 38–40, 93–95, 109–10, 114; Type II, 38–40, 95, 109–10, 114
- European Medicines Agency, 108
- evidence: classifications of quality of, 75; credible use of to inform patient care, 47–73
- Evidence-Based Medicine Working Group, on uncertainty, 2
- evidence-based prediction, *vs.* clinical judgment, 21–24
- evidence-based research, integrating with clinical judgment, 23–24
- expected value of individualized care (EVIC), 85
- expected value of information, 67
- expected welfare criterion, 91, 103
- experimental treatments, to clinical treatments, 32–34
- external validity, 31–32; with missing data, 57
- extrapolation, from study to patient populations, 28–32
- extrapolation problems: in drug approval errors, 110–11; in identifying treatment response, 51–52
- Ezekowitz, M., 37
- false negative result, 67–68
- false positive result, 67–68
- family: difficulty obtaining histories of, 58; missing data on in genetic mutation prediction, 57–58
- Farewell, V., 39
- Faries, M., 12, 120, 126–28
- Faust, R., 22–23
- Federal Food Drug and Cosmetic Act, Section 505(d) amendment of, 111–12
- Ferguson, T., 79, 88
- Fischhoff, B., 89
- Fisher, E., 54–65
- Fisher, L., 27
- Fleiss, J., 97, 132–33
- Fleming, T., 35, 111
- Food and Drug Administration (FDA), 108; adaptive partial approval of, 114–16; Adverse Event Reporting System, 110; approval process of, 108–9; benefit-risk considerations of, 113; quantitative assessment of, 113; rejection of formal decision analysis, 111–13; structured risk-benefit assessment framework of, 111–12; type I and II errors in drug approval of, 109–10
- Food and Drug Administration Safety and Innovation Act, 111–12
- Food, Drug, and Cosmetics Act, 109, 114; 1962 Amendments to, 109
- Fowler, R., 65
- Freedman, L., 89
- Gafni, A., 9
- Gail, M., 18, 19, 71
- Gail Model, 18, 19, 71, 72
- Galen, on medical “sects,” 12
- Garthwaite, P., 89
- Gartlehner, G., 34
- Gassenmaier, M., 14
- genetic mutation prediction, missing family size data in, 57–58
- Ginsburg, G., 16
- Go, A., 32–33
- gold standard, 27
- Goldberg, L., 22
- Goldberger, A., 64
- Good, I., 21
- GRADE system (Grading of Recommendations Assessment, Development and Evaluation), 11, 28; in assessing multiple outcomes, 85–86; classifications of quality of evidence, 75
- Green, S., 27–28, 50
- group drug counseling (GDC), comparing with other therapy, 40–41
- Groves, W., 22, 23
- guideline adherence: *vs.* exercising judgment, 8–15, 117; second-best welfare comparison of with clinical judgment, 24–26
- guidelines: definition of, 8–9; separating information and recommendation aspects of, 118–19; variation in, 10–11
- Guyatt, G., 10–11, 85–86

162 INDEX

- Harrington, J., 109
Haynes, A., 14
health risks, odds ratios and, 132–33
Henrion, M., 89
Higgins, J., 27–28, 44, 50
high blood pressure, evidence-based CPG
 for managing, 24
Hodges, E., 90
Horowitz, J., 55, 56–57, 135
Hoyt, D., 8, 9
Hsieh, D., 133
hypertension treatments: drug trials for, 30–31,
 33–34; missing data in trial on, 55–57
hypothesis tests: to choose when to report
 findings, 40–41; to compare treatments,
 38–40; to treatment decisions, 38–41
- IBIS Model, 71
identification: with missing data on patient
 outcomes or attributes, 53–58; of
 response to testing and treatment, 66–70;
 studying, 52–53
identification problems, 3, 4, 47; of treat-
 ment response, 48–52
identification region, 54
identified set, 54
imprecision measurement, 87–88
inference, logic of, 4
innovation, *vs.* status quo treatment with
 binary outcomes, 141–43
Institute of Medicine (IOM): clinical practice
 guidelines of, 79; report of on guideline
 development, 76–77
instrumental variables: assumptions using,
 63–65, 66; definition of, 64; monotone, 69
intention-to-treat analysis, 51, 55–56
internal validity: favoring of, 47; focus on,
 46; missing data effects on, 56–57; pri-
 macy of, 31–32
international Conference on Harmonisa-
 tion: on clinical significance, 96; on trial
 design, 38–39
intersection bound, 65
Ioannidis, J., 41
- James, P., 24, 28, 30, 42, 50, 75–76
Jarrell, S., 44
Jensen's Inequality, 45
Johnson, E., 22
Joint National Committee, Eighth, guide-
 lines for hypertension management, 30
Jonker Model, 71
juvenile offenders, sentencing and recidi-
 vism analysis of, 62–63
- Kadane, J., 89
Karlin, S., 142
Karmali, K., 17
Kasumova, G., 14
Kindig, D., 101
Kitagawa, T., 65, 91
Koriat, A., 89
- Laird, N., 42–44, 70
Law of Decreasing Credibility, 4–5, 54, 64
Law of Total Probability, 134, 136–37
Laws of Large Numbers, 4
Lazar, N., 41
Lehmann, E., 90
Leiter, U., 13–14, 127–28
Lerman, S., 132
Lichtenstein, S., 89
life span: ambiguity in predicting, 81–83;
 missing data on, 54; predicting mean
 remaining, 59–60
life tables, 59; bounded-variation assump-
 tions and, 60
life-threatening disease, adaptive treatment
 of, 105–6
lung cancer, attributable and relative risks
 of, 133
lymph node dissection: ASCO-SSO guide-
 lines for, 125–28; definition of, 11; as diag-
 nostic test, 120; *vs.* nodal observation,
 11–15; *vs.* observation after positive senti-
 nel lymph node biopsy, 126–28; research
 on, 120–28; survival outcomes with, 15
- Mai, P., 71
Mandelblatt, J., 72
Manski, C., 4–6, 21, 25, 28, 39, 45, 54–57,
 59–66, 68–69, 72, 87, 89, 91, 92, 95–100,
 102–4, 114, 118, 132, 133, 135, 140, 145
Materson, B., 33, 36, 55–56, 57, 96
maximin criterion, 84, 146; in public health
 decisions, 103; Wald's version of, 89–90
maximin decisions, 81–83
maximin rule, 141, 143; in treatment
 choice, 92
maximum regret, conceptual appeal of,
 92–94
McFadden, D., 133
McGregor, J., 15, 125
McNees, S., 45
measured outcomes, to patient welfare,
 35–38
measurement errors, on treatment
 response, 52
Meehl, P., 22–23

- melanoma: case vignette of, 14–15; definition of early stage, 120–21; desirability of lymph node dissection for early stage, 120–28; incubator and marker hypotheses of spread of, 12–14; nodal observation or dissection for, 11–15; patient care recommendations for, 124–25; reasonable care for today, 14–15; treatment of in patients with comorbidities, 29
- Meltzer, D., 21, 67, 84–85
- Mercuri, M., 9
- meta-analysis: of bariatric surgery outcomes, 43; development of, 42; of disparate studies, 41–45; misleading rhetoric of, 43–44
- meta-regression, 44
- metastasis, factors in route of, 14
- minimax criterion, 89, 90
- minimax-regret allocation, of patients in two treatments, 144–45
- minimax-regret criterion, 82–83, 90, 102, 141, 143, 146–47; in public health decisions, 103; in treatment choice, 92
- minimum clinically important difference (MCID), 96
- minimum welfare, maximization of, 146
- missing data, 4; effects of on conclusions, 54; on family size in genetic mutation prediction, 57–58; in hypertension treatment trial, 55–57; on patient attributes, 53–58; on patient outcomes, 54–58; random, 54; on treatment response, 52
- monotone instrumental variable, 69
- monotone treatment response, 69
- monotone treatment rules, 142
- Montgomery, R., 74
- Morgan, G., 89
- Morton, D., 120–22, 123, 125–26
- Moyé, L., 27
- Mullahy, J., 49
- Mullins, D., 74, 110–11
- Multicenter Selective Lymphadenectomy Trial II (MSLT-II), 123; findings of, 123–25, 126, 128; reconsidering sample size in, 98–99
- multiple studies, prediction combining, 70–73
- Mushlin, A., 67–68
- Nagin, D., 62–63, 64
- National Comprehensive Cancer Network recommendations, 26
- National Guideline Clearinghouse, Agency for Healthcare Research and Quality, 8
- National Health and Nutrition Examination Survey (NHANES), on mean life spans, 59–60
- National Institute of Health and Care Excellence (NICE), British, SNL biopsy guidelines of, 125–26, 128
- nature, states of, 77–78
- near-optimal treatment choice, designing trials to enable, 94–98
- near-optimality threshold, choosing, 96, 98
- New Drug Application, 109; review of, 111
- nodal observation: *vs.* lymph node dissection, 11–15; survival outcomes with, 15
- non-sampling errors, 4
- normative analysis, 1
- null hypotheses, trial data to test, 38–39
- observational data: counterfactual outcome unobservability and, 50; credible inference with, 61–66; on treatment response, 49–51
- observational studies: extrapolation problems with, 51–52; partial patient data from, 68; *vs.* trials, 48–49
- odds ratios, 46; health risks and, 132–33
- “off-label” drug uses, 110–11
- O’Hagan, A., 89
- optimal care, assuming rational expectations, 20–21
- optimal decision: factors in, 24; *vs.* reasonable decisions, 78–79; between surveillance and aggressive treatment, 21, 129–31
- ovarian cancer, meta-analysis of, 44
- overpowered trials, 95
- Parmar, M., 89
- Parmigiani, G., 44
- partial identification: modern research on, 63–64; quantifying severity of missing data problems, 57–58
- partial personalized risk assessment, 59–60
- partial predictions, combining, 72–73
- partially identified probabilities, 5
- partially identified responses, 53
- patient attributes, 117; in disease risk predictions, 59; identification with missing data on, 53–58
- patient care, credible use of evidence in, 47–73
- patient outcomes: bounds on with no knowledge of counterfactual outcomes, 136–37; identification with missing data on, 53–58

- patient populations, extrapolation of study populations to, 28–32
- patient preferences, 85
- patient welfare: maximization of, 101; from measured outcomes to, 35–38; uncertainty about, 84–86
- Pauly, M., 20
- Pearl, J., 65
- Peltzman, S., 109
- Pepper, J., 69
- perfect foresight, 20
- personal state of nature, 78; in optimal *vs.* reasonable decisions, 78–79
- personalized medicine: in breast cancer risk assessment, 17–18; definition of, 16; degrees of, 16; in predicting cardiovascular disease, 17; in predicting unrealistically precise probabilities, 18–19
- personalized risk assessment, ecological inference problem in, 134–35
- Phase 1 trials, 108–9
- Phase 2 trials, 109
- Phase 3 trials, 35, 109
- Phelps, C., 67–68
- point-identified probabilities, 5
- point-identified responses, 53
- point-identity, 133
- point predictions, probabilistic predictions and, 24–25
- population health, *vs.* public health, 101
- population health perspective, 91; on reasonable care, 100–107
- power calculations, 110; to choose sample size, 94–95
- Prasad, V., 36, 111
- precise probabilities, unrealistically, 18–19
- prediction errors, of mean prediction, 44–45
- predictions: combining multiple studies, 70–73; problems with, 5
- Prescription Drug User Fee Act, 111
- prescriptive analysis, 1
- primary endpoint, 37
- probabilistic predictions: credible precise, 5; disparate, 71
- probabilistic risk assessments/diagnoses, accuracy of, 24–25
- procedures, 88
- Psaty, B., 111
- psychological research, on evidence-based prediction *vs.* clinical judgment, 21–24
- public health, *vs.* population health, 101
- publication bias, 41
- Pure Food and Drug Act, 109
- quality of evidence classifications, 75
- random-effects models, 42–43, 44
- random treatment assignments, 49
- randomized controlled trials (RCTs), 27; design guidelines for, 38–39; quality of evidence from, 75; supposed superiority of, 27–28
- rare-disease assumption, 133
- rational expectations, 80; optimal care assuming, 20–21
- reasonable, multiple interpretations of, 79
- reasonable care: population health perspective on, 100–107; with sample data, 87–99
- reasonable decision, 78–79; criteria for, 80–83; between surveillance and aggressive treatment, 83–84, 138–39
- recidivism, sentencing and, 62–63
- regression, 3
- regret, 82; of treatment allocation, 144–45
- Reiersol, O., 64
- relative risk, 132–33
- relevance, sacrificing for certitude, 45–46
- reporting findings, choosing when, 40–41
- response-based sampling, 132, 133
- retrospective studies, 132
- risk assessments: combining multiple, 71–72; partial personalized, 59–60; personalized, 134–35; testing sensitivity and specificity in, 70
- risk evaluation models, 58
- Rosenbaum, P., 32
- Rubin, H., 142
- Sackett, D., 23–24
- sample data, reasonable care with, 87–99
- sample size: enabling near-optimal treatment choice, 95–96; implications of for practice, 98; minimum, 97–99; in MSLT-II trial, 98–99; power calculations to choose, 94–95
- sampling: in personalized risk assessment, 134; types of, 132
- Sarbin, T., 22
- Savage, L., 83, 89, 92
- Schlag, K., 91, 92
- Schünemann, H., 11
- screening, 67
- secondary endpoint, 37
- Sedgwick, P., 96
- sensitivity, 69–70
- sentencing and recidivism analysis, 62–63
- sentinel lymph node (SLN) biopsy, 14; British guidelines for, 125; as less invasive

- than dissection, 121; logic under, 121, 122–23; Morton et al. study on, 120–23; observation of, 123–26; observation or dissection after positive, 126–28; success of, 121–22; two steps in, 121
- Shaikh, A., 66
- SHEP Cooperative Research Group, 30
- SHEP trial, 31, 33; duration of, 36
- specificity, 69–70
- Spiegelhalter, D., 89, 91
- Staessen, J., 30
- Staessen trial, 31, 33; duration of, 36
- Stanley, T., 31–32, 44
- state of nature, 77–78
- state space, 78
- statistical decision function, 88
- statistical decision problem, treatment choice as, 140–43
- statistical decision theory: Bayesian, 91–94; conceptual appeal of, 92–94; post-Wald, 90–91; practical appeal of, 92; principles of, 88–91; for treatment choice, 91–94; Wald development of, 88
- statistical imprecision, 3–4, 47; with breast cancer risk assessment tools, 18–19; drug approval errors due to, 110–11
- statistical inference, Bayesian prescription for, 89
- statistical precision, conventional ideas of, 41
- statistical prediction, 3
- statistical theory, in medical research, 87–88
- statistical treatment rule (STR), 91, 140–41; feasible, 142; regret of, 93–94
- status quo treatment: *vs.* innovation with binary outcomes, 141–43; of life-threatening disease, 105–6
- Stoddart, G., 101
- Stoye, J., 87, 91, 92, 97
- study populations: differences of from patient populations, 29; extrapolation of to patient populations, 28–32; incomplete descriptions of, 29–30
- Su, F., 91
- subjective expected utility, maximization of, 80, 89
- subjective expected welfare, maximization of, 146
- subpopulations, treatment responses of, 64–65
- success probabilities: bounds of, 56; bounds on with no knowledge of counterfactual outcomes, 136–37
- surrogate outcomes: FDA reliance on, 111; interpreting, 35–36; with missing data, 57
- surveillance: *vs.* aggressive treatment, 2–3, 21, 25–26, 83–84, 138–39; for breast cancer, 25–26; formalization of optimal choice of, 129–31; formalization of reasonable choice between aggressive treatment and, 138–39
- Swan-Ganz catheterization, bounding mortality effects of, 65–66
- systematic review, 41–42
- Temin, P., 109
- tests: identification of response to, 66–80; optimal, 67–68; sensitivity and specificity of, 70
- Tetenov, A., 6, 39, 91, 95–99
- Thompson, S., 44
- Torjesen, I., 125
- treatment: diversification of, 101–4; hypothesis tests to compare, 38–40; identification of response to, 66–80; intensity of, 32–33; optimal, 67–68; personalized outcome probabilities of, 16; rating system for recommendations of, 75–76
- treatment allocation, minimax-regret, 144–45
- treatment choice: from hypothesis tests to, 38–41; statistical decision theory for, 91–94, 140–43
- treatment response: extrapolation problems in identifying, 51–52; identification problems with, 48–52; intersection bound on, 65; missing data and measurement errors on, 52; observational data on, 49–51; partially identified *vs.* point-identified, 53; predicting based on evidence-based guidelines, 117; in study subpopulations, 64–65; trial data on, 48–49
- treatment variation: argument for, 100; encouraging under uncertainty, 107; rationale for, 118
- trial data: revealing outcome distributions, 87; on treatment response, 48–49
- trials: assessing multiple outcomes of, 36–37; designing to enable near-optimal treatment choice, 94–98; extrapolation problems with, 51–52; for hypertension treatment, 30–31; with imperfect compliance, 51; intensity of treatment in, 32–33; interpreting surrogate outcomes of, 35–36; *vs.* observational studies, 48–49; patient welfare and outcomes of, 35–37; with perfect compliance, 49; primacy of internal validity in, 31–32; primary and secondary endpoints in, 37; on subjects with no comorbidities, 28–29;

166 INDEX

- trials: assessing multiple outcomes of
(*continued*) under- and overpowered, 95.
See also Drug trials; Randomized controlled trials (RCTs)
- Tukey, John, 46
- Tunis, S., 74
- Type I errors, 38; chance of, 93–94; in drug approval, 109–10, 114; probability of, 39, 40, 95
- Type II errors, 38–39; in drug approval, 109–10, 114; probability of, 39, 40, 95
- ultrapessimistic criterion, 92
- uncertainty: care under, 119; in choosing between surveillance and aggressive treatment, 2–3; deep, 81–83; encouraging treatment variation under, 107; formalizing, 76–78; managing in drug approval, 108–16; about patient welfare, 84–86; planning population health under, 101; qualitative recognition of, 74–76
- uncontrolled case series, 43
- underpowered trials, 95
- utilitarian social welfare, maximizing of, 101
- utilitarian welfare function, 20
- Vernon, J., 109
- Viscusi, K., 109
- Visvanathan, K., 26
- Vytlačil, E., 66
- Wald, A., 88–89
- Wald decision making theory, 88–91; statistical decision theory after, 90–91
- warfarin: possible outcomes for patients taking, 37–38; in trial *vs.* clinical treatments, 32–33
- Wasserstein, R., 41
- welfare transformation, decision making for, 103–4
- Wennberg, J., 9
- Willard, H., 16
- wisdom of crowds, 44–45
- wishful extrapolation, 28; drug approval errors due to, 110–11
- wishful meta-analyses, of disparate studies, 41–45
- Wong, S., 29, 125, 127–28
- worst-case prediction, 81–82
- x-pox, 102–3, 146–47