Bayesian Methods and Ethics in a Clinical Trial Design

Edited by
Joseph B. Kadane

Wiley Series in Probability and Statistics

This page intentionally left blank

Bayesian Methods and Ethics in a Clinical Trial Design

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: *Vic Barnett, Ralph A. Bradley, Nicholas I. Fisher, J. Stuart Hunter, J. B. Kadane, David G. Kendall, David W. Scott, Adrian F. M. Smith, Jozef L. Teugels, Geoffrey S. Watson*

A complete list of the titles in this series appears at the end of this volume

Bayesian Methods and Ethics in a Clinical Trial Design

Edited by

JOSEPH B. KADANE
Department of Statistics
Carnegie Mellon University

A Wiley-Interscience Publication
JOHN WILEY & SONS, INC.
New York • Chichester • Brisbane • Toronto • Singapore

A NOTE TO THE READER
This book has been electronically reproduced from digital information stored at John Wiley & Sons, Inc. We are pleased that the use of this new technology will enable us to keep works of enduring scholarly value in print as long as there is reasonable demand for them. The content of this book is identical to previous printings.

Library of Congress Cataloging in Publication Data:

Bayesian methods and ethics in a clinical trial design / edited by Joseph B. Kadane.

p. cm. — (Wiley series in probability and mathematical statistics. Applied probability section)

Includes bibliographical references and index.

ISBN 0-471-84680-5 (cloth : alk. paper)

1. Clinical trials—Moral and ethical aspects. I. Kadane, Joseph B. II. Series: Wiley series in probability and mathematical statistics. Applied probability and statistics.

[DNLM: 1. Research Design. 2. Clinical Trials—methods. 3. Bayes Theorem. 4. Ethics, Medical. 5. Jurisprudence—United States. 20.5 B357 1996]

R853.C55B39 1996

615'.072--dc20

DNLM/DLC

95-14352

10 9 8 7 6 5 4 3

Contents

This page intentionally left blank

Preface

The ideas in this book, the use of Bayesian methods to give patients a better break in clinical trials, have taken over a decade to bring to fruition. They have, in general, their technological roots in the development of Bayesian ideas, and in particular in progress in practical elicitation of prior opinions.

These ideas have been developed by a talented interdisciplinary group involving philosophy (Schaffner, Seidenfeld), law (Kairys), medicine (Heitmiller, Blanck), statistics (Kadane, Sedransk, Emrich), and statistical computing (Galway). This group was helped, critiqued, and commented upon by other participants, including Coulehan, Katz, Popp, and Moore. All have their say in the chapters that follow.

That each of these people has been trained in a particular way does lead to certain intellectual leanings on their part, but it does not determine their perspectives by any means. There was a lot of debate in our meetings, and we came to appreciate each other's viewpoints more as a result. We also came to see that we are all potential consumers of clinical research, and we are all potential patients in clinical trials as well. We are particularly reminded of the human stakes in our research by the untimely death of Larry Emrich, coauthor of Chapter 18 of this book.

My hope in editing this book is to provide some of the flavor of the debate. To do so, I have encouraged each author to tell a personal story in a personal way. The consequence is that the book is somewhat uneven from chapter to chapter. I hope that the burden on the reader imposed by this policy is compensated for by the genuineness of the resulting expression. The discussion did not always lead to agreement; sometimes we found the opinions of others wrong and/or offensive. In order to expose the variety of opinions offered, there are three chapters of commentary (7, 15, and 16), and one of rebuttal (17).

The book is organized as follows: Chapter 1 gives an overview of the project and touches on the main ideas. Most readers will be well served by reading it first. The next three chapters, constituting Part I, deal with important issues for the class of designs of clinical trials proposed here: Chapter 2, by Schaffner, reviews current ethical theory and how it relates to our design; Chapter 3, by

Sedransk, examines the key concept of the admissibility of a treatment assignment to a particular patient and offers advantages and disadvantages for each of the several choices. Finally, Chapter 4, written by Kadane and Seidenfeld, shows how the data from a trial designed as we suggest, can be analyzed to yield uncontaminated information about the effect of treatment on outcome. I think a first reader would want to at least skim these chapters.

The heart of the book is Part II, the test case of the verapamil/nitroprusside trial as agents for treatment of hypertension immediately after open-heart surgery. This material, in Chapters 5 through 13, discusses the process and results of the trial, as experienced by the investigators. Probably this will be the most heavily studied aspect of the book, since it is more specific than the generalities that precede and follow it.

Part III takes up other issues that we explored in this context. In Chapter 14 Kairys explores American law and how it relates to our design for a clinical trial. This chapter attracted comments from Popp and Moore, and from Katz, to which Kairys replies in Chapter 17. Each of the chapters in the legal section is dated to reflect when it was first written. Each of the authors had a recent opportunity to revise and declined to do so. Finally, Chapter 18 reports work by Sedransk and Emrich about when a rational patient would agree to participate in a clinical trial. The book concludes with an Epilogue in Chapter 19.

The work reported here is the subject of research funded by the Ethics and Values in Science and Technology Program of the National Science Foundation and by the National Endowment for the Humanities, through Carnegie Mellon University. Those supported by the grant included Lionel Galway, Joseph B. Kadane, David Kairys, Ken Schaffner, Nell Sedransk, and Teddy Seidenfeld, advised by Thomas J. J. Blanck, Jack Coulehan, Preston Covey, Jerome J. DeCosse, Arvin S. Glicksman, Eugenie Heitmiller, Rachelle Hollander, Kathryn D. Katz, Alan Meisel, A. John Popp, and John C. Ruckdeschel. Others whose comments were helpful include John Bailar III, Robyn Dawes, Clark Glymour, and Juana Sanchez. Chapters 1 and 12 appeared in earlier forms in the *Journal of Medicine and Philosophy*, 11 (1986), 325–404, and in the *Journal of Statistical Planning and Inference*, 40 (1994), 221–232, respectively.

J. B. KADANE

Pittsburgh, PA
September 1995

About the Authors

Thomas J. J. Blanck (M.D., Ph.D.) is Professor of Anesthesiology, Pharmacology, Physiology, and Biophysics at Cornell University Medical College. He is a practicing anesthesiologist and has written extensively about drug action on the cardiovascular system. He has been a member of the Human Investigations Committees at both Johns Hopkins University and Cornell University Medical College for the past ten years.

Thomas J. Conahan (M.D.) is Associate Professor of Anesthesia at the Hospital of the University of Pennsylvania and is a member of the cardiac anesthesia group at that hospital. He has conducted research in both cardiac and ambulatory anesthesia and is the author of a textbook of cardiac anesthesia.

John L. Coulehan (M.D.) is a Professor of Medicine and Preventive Medicine at the State University of New York at Sony Brook, where he is also Codirector of the Medical Humanities Program and a Senior Fellow at the University's Institute for Medicine in Contemporary Society. He is the author of a widely used textbook *The Medical Interview* (F. A. Davis Co., 1991), as well as over a hundred papers in the medical literature ranging in topic from clinical trials and epidemiologic studies to personal essays on the physician-patient relationship.

Lawerence J. Emrich (M.S.), until his untimely death from Hodgkins' Disease, was Head of Statistical and Computing Laboratories at the Roswell Park Memorial Cancer Institute and Associate Research Professor of Biometry at State University of New York at Buffalo where he also was the principal consulting statistician for the School of Dentistry. He was highly regarded by clinical researchers and by statisticians for his work in designing and analyzing clinical research.

Lionel A. Galway (Ph.D.) is Associate Statistician at the RAND Corporation in Santa Monica, CA. He works on public policy research projects in the areas of military logistics, environmental affairs (particularly air pollution), and civil justice.

Eugenie S. Heitmiller (M.D.) is an Associate Professor of Anesthesiology and Pediatrics at The Johns Hopkins University Medical School. A member of the American Society of Anesthesiology and the American Academy of

Pediatrics, she is the author of several articles and book chapters on the anesthetic care of children and adults during cardiac surgery.

Joseph B. Kadane (Ph.D.) is Leonard J. Savage Professor of Statistics and Social Sciences at Carnegie Mellon University. A Fellow of the American Statistical Association and the Institute of Mathematical Statistics, he is the author of over a hundred papers, ranging from statistical theory to various applications.

David Kairys (Esq.) is Professor of Law at Temple University. He is the author of many articles and books on constitutional law, most recently *With Liberty and Justice for Some*, and on law, science, and technology. He has also litigated some leading cases on a range of civil rights and civil liberties issues, including unwitting and nonconsensual application of drugs.

Katheryn D. Katz (Esq.) is Professor of Law at Albany Law School of Union University. She teaches primarily in the areas of family and constitutional law and reproductive technology. She is the author of a number of articles addressing issues such as the scope of personal autonomy, the allocation of authority over decision making among the family, the individual, the medical profession, and the state as well as articles dealing with the protection of children's rights. She served as a member of the Human Studies Committee of Albany Medical Center for 12 years.

Robert G. Merin (M.D.) is Professor of Anesthesiology at the Medical College of Georgia. He has been an academic anesthesiologist for more than 30 years at a variety of institutions including the University of Rochester and the University of Texas-Houston. His major interests have been cardiovascular pharmacology. He has been an editor of the journal *Anesthesiology* and a consultant for the U.S. Food and Drug Administration. More recently his clinical activities have concentrated on cardiac anesthesia.

Dale Moore (Esq.) is Professor of Law at Albany Law School and Adjunct Professor at Albany Medical College, where she also serves as vice chair of the Committee on Research Involving Human Subjects. She teaches and writes on health law and bioethics issues.

A. John Popp (M.D.) is presently the Henry and Sally Schaffer Chairman of Surgery and Head of the Division of Neurosurgery at Albany Medical College. Dr. Popp also serves as Surgeon-in-Chief of the Albany Medical Center Hospital. His publications reflect a diverse interest in neurosurgery including, most recently publications on head injury, research design, and neurosurgical history.

Richard L. Prager (M.D.) is the Head of the Section of Cardiac and Thoracic Surgery at the Michigan Heart and Vascular Institute at St. Joseph Mercy Hospital in Ann Arbor, Michigan. He is a Clinical Associate Professor in the section of Thoracic Surgery at the University of Michigan. He has authored many articles in the field of cardiac and general thoracic surgery. He is currently the Medical Director of the HCFA Coronary Artery Bypass Demonstration Project at St. Joseph Mercy Hospital.

James J. Richter (M.D., Ph.D.) is Director of the Department of Anesthesiology at Hartford Hospital in Hartford, Connecticut. He has conducted research programs in neurochemistry and has continued interest in the growth of basic science applications to clinical medicine. Hartford Hospital has a very large graduate medical education program and is increasingly active in clinical research.

Kenneth F. Schaffner (M.D., Ph.D.) is University Professor of Medical Humanities at the George Washington University. A fellow of the American Association for the Advancement of Science and the Hastings Center, he is the author of numerous publications in the philosophy of medicine. His most recent book, *Discovery and Explanation in Biology and Medicine*, was published in 1993 by the University of Chicago Press.

Nell Sedransk (Ph.D.) is Professor of Statistics at Case Western Reserve University in Cleveland. She is a Fellow of the International Statistical Institute and has published research in the literatures both of theoretical statistics and of clinical research in medicine. Her experience includes the design and analysis of a wide variety of clinical trials in medicine and dentistry.

Teddy Seidenfeld (Ph.D.) is Professor of Philosophy and Statistics at Carnegie Mellon University. His research interests include "foundational" problems in statistics and decision theory.

This page intentionally left blank

Bayesian Methods and Ethics in a Clinical Trial Design

This page intentionally left blank

PART I

Major Issues

This page intentionally left blank

CHAPTER 1

Introduction

Joseph B. Kadane

The circumstances surrounding the administration of experimental drugs and treatments to human beings trouble the conscience of the medical and scientific community. Not that I think what is done is bad. On the whole the system works surprisingly well. There are a few atrocity cases, however. I remember especially the Tuskegee syphilis experiment, in which black men with syphilis remained untreated for years so that the experimenters could observe the natural course of the disease (Brandt 1978). But on the whole it is my belief that standard experimental practice takes reasonable care of patients in clinical trials. I think that experimentation bothers the conscience because it is not clear that the patient is receiving the best possible care in the experimental situation (see Lellouch and Schwartz 1971; Clayton 1982). This is a quest without a definite end: To be sure it is a challenge to our collective applied cleverness to see if we can somehow devise alternatives that are both arguably better for patients and scientifically responsible.

1.1 DO PATIENTS GET A FAIR SHAKE IN CURRENT CLINICAL TRIALS?

Most clinical trials use some form of randomization to assign patients to treatments. Patients may be told that the treatment they will be given is decided by the flip of a coin. Often this is not literally true, for neither the patient nor the attending physician knows which treatment will be assigned. The patient is asked to sign an informed-consent statement agreeing to be in the experiment, and agreeing to treatment assigned in a random manner. Informed consent is like a legal contract between the patient and the physician. Usually the contract states the patient's diagnosis, the treatments under study, and the possible adverse side effects of the treatments. On this basis the patient signs the form, accepting the randomly assigned treatment.

The informed-consent procedure affirms the rights of the patients to determine the uses of their bodies. But if informed consent is to be regarded as a valid contract, the patient and experimenter must be reasonably equal in

Bayesian Methods and Ethics in a Clinical Trial Design, Edited by Joseph B. Kadane
ISBN 0-471-84680-5

bargaining power and have their wits about them. On this point the evidence is negative. In several studies the experimenter was in the waiting room to interview the patient leaving the physician's office after the informed-consent interview. The patients had poor recall of what they had signed and what the treatments and side effects were. Many patients had interpreted the process of informed consent as a form they had to sign in order to get treatment. (See Meisel and Roth 1983 for a review).

To say this is not to criticize the physicians or the patients. The patients are, after all, sick, and depending on the nature and severity of their illness, their cognitive functioning may be impaired. They may, earlier in the informed-consent interview or in the recent past, have been given bad news about their health or prognosis. Who among us might not be upset and functioning poorly when given such news? The physician and the patient are not anywhere near equal in bargaining power in this situation.

Despite the evidence showing that informed consent is rarely "informed" and may or may not be consent, I would not recommend abolishing the ceremony. Rather, my interpretation of these results is that they impose a greater burden on the medical-scientific community to ensure that the contract offered the patient in informed consent is as advantageous as possible for the patient. We must look out for the patient's interests, since the patient may be unable to do so. We cannot use the excuse that the patient has agreed, via informed consent, to a disadvantageous procedure. To do so has the ring of foisting disadvantageous treaties on Indian chiefs in a language they did not know.

So I am led to consider the fairness of the substance of what the patient is asked to sign in informed consent. Here the picture is somewhat gloomier. Let us accept that clinical trials only occur when the medical community is not agreed upon the best treatment for the condition under study. In such a circumstance the patient can be told truthfully that the best treatment among those compared in the trial is not currently known. But this does not justify random assignment. A knowledgeable patient might say, "Doctor, you know about me and about my disease. You must have a hunch about which treatment would be better for me. Please give me the treatment, and forget about flipping coins." If the attending physician is not so knowledgeable, we can suppose that a consulting expert could be found to make such a judgment. Even if such an expert were truly neutral at the start of a study, after the first few patients are studied and their outcomes are at least partially known, there would be a reasonable basis for a hunch that could be useful to the patient, though without the established validity we usually associate with scientific knowledge (see Chalmers 1967).

One road out of this conundrum is to keep the attending physicians ignorant of the results of the trial to date. This is useful in that it may help reduce their tendency, even unconsciously, to change the pool of patients in the trial or to misinterpret "recovery." But as a way to deal with the legitimate patient objective of getting competent and appropriate care, it seems to me to be very shaky. If anything, there might be a duty to inform the attending

physicians of the data to date so that their advice to, and treatment of, patients might be better informed. To prevent the person who is supposed to be using his or her expert judgment to help the patient from having the very information that might help the patient seems to me to be unethical. After all if that person had that information and conveyed it to the patient, the patient might make a decision the designers of the trial do not want the patient to make.

Sometimes a physician believes that a drug or treatment available only through a specific clinical trial would be advantageous to the patient. This argument presupposes the current U.S. law that requires FDA approval before a new drug can be made available commercially. It also assumes that the patient is unable or unwilling to go to another country where the treatment is available without being in a trial. Such availability of the treatment would mean that the patient could be certain of getting the treatment, without undergoing randomization. This is then a weak sense of advantageousness to patients, and it applies to only a few trials.

There is another line of argument that supports the current system from a very different premise. This line is utilitarian, and admits that the deal currently offered patients is suboptimal in a narrow accounting of the patient's interests. Quite frankly, the patient is being asked to sacrifice some prospect of recovery for the sake of scientific progress. Of course this is hubris, the kind of assertion that has led in our century to much good but also much mischief. I would feel more comfortable with it if I thought that informed consent worked better than it apparently does. While an appeal to this argument may in the end be necessary to support doing clinical trials on human beings at all, surely the circumstance that the patients, as a practical matter, are not in a good position to defend themselves from overly great and unfair claims that they should "help science" must admonish us to design trials to reduce the burden on patients to a minimum. Whether patients get a fair shake in current clinical trials, then, depends critically on our ability to propose a system that would be better for them and still permit the scientific analysis of the resulting data.

1.2 WHY ALLOWING PATIENTS TO CHOOSE THEIR OWN TREATMENTS IS NOT A SOLUTION

Clinical trials would be a fruitless exercise if the data could not be used scientifically. There is, in medicine, a long history of false conclusions reached through observational studies or from clinical trials lacking proper control. This has led the medical/scientific community to be methodologically cautious, and properly so.

To take a position opposite to current practice, suppose that the patient and the physician jointly decide on a treatment. The trial may then follow the course of the patient. This would remove nearly entirely the burden imposed by current trials on patients of being assigned a possibly disadvantageous treatment. However, how could interpreters of the data separate effects due to the treatments themselves from effects due to the kinds of people who choose

the treatments? To take a recent example, a study was done on women with breast cancer to determine whether segmental mastectomy, which removes only sufficient tissue to ensure that what remains is free of tumor, is as efficacious as the more traditional total mastectomy, which removes the entire breast and some chest muscles (Fisher et al. 1985). Women may value differently the benefits of saving the breast against the possible increased risk of recurrence of the cancer (and early death). How they do so may have something to do with their personality, which, for all we know now about cancer, may have something to do with their outcome. Consequently, had they been allowed to choose their own treatments, it would have been very difficult to interpret the results. A summary of the data might be "Among those who chose segmental mastectomy, the five-year survival rate was x, while among those who chose total mastectomy, the five-year survival rate was y." If $x > y$, the advice to patients is to be like those who chose segmental mastectomy. This is unhelpful both scientifically and therapeutically. I do not mean that such data must be valueless—as a statistician I occasionally work with data sets with as much ambiguity. But I do mean that, had it been required that patients be allowed to choose their treatments, it might well have been decided that such a study would not be a cost-effective way to make progress on cancer, and consequently the study might well not have been done. And this would have been a real loss to the thousands of women who develop breast cancer every year. We would not be doing a service if, in the name of protecting patients, we protected them from the possibility of medical progress using clinical trials. (For a contrary view outside of the context of experimentation, see Schultz 1985).

What I seek, then, is a middle position, one that offers patients a better deal in the design of a trial but that still offers data interpretable as bearing on the effect of the treatment on the disease. This will be a compromise of some sort. It is in the nature of compromises that they are uneasy positions, liable to attack from both sides.

1.3 DISAGREEMENT AMONG EXPERTS

Disagreement among medical experts about the advisable treatment for the condition seems inherent to the decision to conduct a clinical trial. Clinical trials are expensive, and, according to this model, will be conducted only when serious disagreement exists. Many medical procedures are supported by custom, and do not have a rigorous scientific basis, so orthodox treatment is not necessarily good medicine. The radical mastectomy operation for breast cancer, mentioned above, was the standard treatment since the turn of the century. A trial to compare it to the alternative lumpectomy was not begun until the mid-1970s, however, when sufficient expert opinion supported the alternative to create serious disagreement.

Sometimes there are trials where nearly the entire medical community is quite convinced of the outcome before the trial starts. This was the case in the

test of a derivative of apricot pits (Laetrile) as a treatment for terminal cancer. Although outlawed as a treatment in the United States, many very sick and desperate patients were going to Mexico for treatment using it, in the hope that it might be effective. Finally, sufficient political pressure was brought to bear through Congress on NIH that a clinical trial was authorized. (The treatment proved ineffective.)

What happened here, in my view, is that the definition of who is a medical expert was expanded to include physicians who believed in this unorthodox treatment. The lesson I learn from this episode is that the decisions as to who is an expert and what clinical trials will be conducted have a political component. Often expenditure of public money is involved. Always the credibility and trust put in the medical/scientific community by the public is at stake. While the politics involved rarely includes Congress, it usually does have to do with the pecking order among physicians and scientists (and also between physicians and scientists), which is political in the larger sense. I do not regard this as pejorative; I think of peer review as a political mechanism to reach political decisions about the allocation of resources. "Political" need not mean partisan in the sense of political parties. Modern scientific politics often involves fascinating mixtures of expertise and general judgment in matters such as environmental and space policy, energy, and the construction of large laboratories to study fundamental particles, as well as medicine. Whom to trust to do what is a matter of continuing discussion; I will return to this question later, after explaining the proposal discussed in this book, to show that the particular way information is treated here ameliorates this problem to some extent.

Disagreement in the medical community often colors what a physician might say to a patient about the relative merits of the treatments in a clinical trial. To say that the best treatment is not known is obvious. This usually does not imply that the physician has no opinion about which of the treatments might be better for the patient. I would argue that what a patient seeks from his or her physician is informed opinion. Medical certainty and agreement, although pleasant when available, are the exception. (This is not to deny the possible therapeutic benefit of a patient's naive belief in the infallibility of the physician.)

1.4 A PARTIAL REDEFINITION OF THE RESPONSIBILITIES OF THE DESIGNERS OF A CLINICAL TRIAL TO THE PATIENT

In the context of clinical trials medical experts will likely disagree about the desirability of the treatments. A responsible clinical trial cannot offer patients a choice of treatments. What might then be said about the responsibilities of the designers of a clinical trial to the patient?

For the long-term survival of clinical trials as well as for the shorter-term effect of feeling better about what one is doing, it seems to me that a clinical trial should offer the patient whatever benefits the patient might reasonably be

able to obtain outside the trial, as long as it can do so and still fulfill its scientific mission. What might this consist of in the context of lack of knowledge in the medical community and conflict over the best courses of treatment? Suppose that there is a group of experts on the disease, identified in the scientific-political way described above. A vigorous patient with adequate financial resources and good medical connections might get to see one of these experts and take the advice of that expert about what treatment to take. Thus I think that a clinical trial should try to replicate this, as best it can, for the patients in the trial.

A second thing I think a clinical trial ought to do is to offer information or use information developed during the course of the trial for the benefit of the patient to the extent that it can do so without jeopardizing the scientific merit of the study. The proviso is there because totally unblinded patients might implicitly be choosing their treatment in the guise of choosing whether to be in the trial. Thus, for example, in a randomized trial, telling the patient what treatment would be assigned if the patient were in the trial, might have this effect. There can be legitimate differences of opinion about what information might "jeopardize the scientific merit of the study," a matter to which this chapter returns later.

1.5 THE BASICS OF SUBJECTIVE STATISTICAL INFERENCE

There are several forms of classical statistics and subjective Bayesian statistics. A good general review of these schools can be found in Barnett (1982).

What is particularly important for our purposes about the subjective Bayesian view point is that it offers legitimacy and methods of calculation for dealing with opinions, in our case, opinions of medical experts. Modern Bayesian research concerns, among other topics, the elicitation of opinion in the form of probability distributions (Kadane et al. 1980), and this is the new technical tool being brought to bear on clinical trials. In an elicitation an expert is asked questions about his or her median for a dependent variable (the one being predicted) given specified values of the predictor variables (being used in the prediction). The answers to these questions, and other similar questions discussed later, are put into a computer, and they lead to a computer model of how the expert would answer any such questions. Thus Bayesian analysis allows us to study how experts are similar and different in their views, and also allows us to use these opinions for the benefit of patients without having physically to consult the expert about each patient.

1.6 PROPOSAL FOR A MORE ETHICAL METHOD FOR CLINICAL TRIALS

The possibilities opened by the subjective Bayesian technique led Sedransk and me (1980) to propose the following modification of the standard clinical trial.

Suppose that a number of experts on the disease in question, say, five, are identified. Suppose also that the goal of treatment is agreed upon (i.e., length of life in the case of cancer), and that a short list of important predictor (or concomitant) variables are identified that arguably are the most important variables for predicting prognosis (for primary breast cancer, at least extent of disease, whether there is clinical nodal involvement, and pre- or post-menopausal status). Then the opinion (as a probability distribution for the goal of the treatment) of each expert could be elicited as a function of the predictor variables and treatment. This would be done, in the example, by asking questions like: For a pre-menopausal woman with no nodal involvement treated with segmental mastectomy, what is your median for how long she will live (i.e., what is the length of time you think it is as likely she would live longer than as shorter than)? Another part of the elicitation involves finding out how surely these opinions are held, by asking questions like: If we had already tried segmental mastectomy on two patients, the first pre-menopausal with nodal involvement who then lived 2.7 years, and the second post-menopausal with no nodal involvement who then lived for 6.5 years, what would your median now be on how long a pre-menopausal patient with no nodal involvement would live? An expert who has treated many patients, and is very sure, would not let small amounts of data influence his or her judgment: An expert who is unsure (e.g., who may be relying solely on animal and other laboratory studies) might be much more influenced by hypothetical data. A single indicator (or utility), such as expected survival time, would also be chosen.

Using standard Bayesian probabilistic methods, these opinions can be updated, once the trial is underway, for the evidence as it is gathered. Thus it is possible to have available a computer model for what each expert would think, if asked before that patient is assigned a treatment by the trial, about the prognosis of each individual patient in the trial as a function of treatment. These opinions would take into account only the values of the predictor variables, and would not therefore be necessarily the same as an opinion based on a full medical examination.

The proposal then is that *unless at least one expert (as modeled on the computer) finds a treatment to be the best for someone with the predictor variable values of the patient, it will not be assigned.* Thus, if only one expert finds a treatment the best for a patient of a given type, it could, under this rule, be assigned to that patient. Within this constraint to admissible treatments, patients may be assigned to treatments in any one of many ways—randomly, or to maximize balance on the important predictor variables, or in any other way. Other concepts of admissibility are considered in Chapter 3.

How might data collected in this way be analyzed? If random selection of treatment is made among the admissible treatments, an analysis could be conducted based on classical randomization theory. From a Bayesian perspective, the analysis is straightforward. At each stage the assignment of patients to treatments is a known (albeit complicated) function of the predictor variables of the current and past patients, and the results to date. They involve no dependence on unrecorded aspects of the patients and in particular do not

involve unexplained patient choice. Hence the data on treatment outcomes, given treatment and the predictor variables, are independent of treatment assignment, as discussed more fully in Chapter 4.

An appropriate Bayesian analysis, then is conditional on the predictor variables used in the designs. This means that the conclusions would be stated separately for each value of predictor variables (for post-menopausal women with nodal involvement, etc). Alternatively, conclusions might be stated by a probability distribution conditioned on the predictor variables.

It might happen that there is evidence E such that, given E, no expert prefers treatment T_1, and hence no more subjects are given it. However, it might be the case that if they had, T_1 would have been preferred in the long run. Stable estimation, in the sense of Edwards, Lindman, and Savage (1963), is lost. Note, however, this would happen only if the alternative to treatment T_1 continues to do better than each expert thinks T_1 would. It is the thesis of this book that in such an instance it would be unethical to assign T_1 to patients.

It is one thing to propose in principle how a trial might be conducted; it is quite another thing to do it. Several things might be learned from such an experience: whether institutional review boards will permit the trial, whether the calculations proposed can be performed, what the response of patients, physicians, experts and others is, whether the trial is scientifically successful in answering the question posed.

Due to a fortunate circumstance, I was able to attract a team of anesthesiologists at Johns Hopkins to these ideas. Jointly we set up a trial to compare nitroprusside and verapamil infusions as treatment for hypertension after separation from cardiopulmonary bypass during cardiac surgery. Nationally, nitroprusside infusion is the more common treatment, but both treatments had been used at Johns Hopkins Hospital. In fact the Joint Committee on Clinical Investigation, the human subjects institutional review board at Johns Hopkins, had previously approved an unrestricted randomized trial comparing these two treatments. However, that trial had not begun. This history is recounted in Chapter 6.

Our first step was to find experts to serve as guardians of the interests of the patients. They should be selected so that they adequately represent the range of responsible medical opinion on the issues at hand. In this case they were chosen so that only one of the five is at Johns Hopkins, to avoid conflict of interest. They include one anesthesiologist, one cardiac anesthesiologist, one cardiac surgeon, one physiologist and anethesiologist, and one biochemist and cardiac anesthesiologist.

A decision had to be made about what predictor variables to use. Those selected were whether the patient was already receiving beta blockers or calcium antagonists, whether the patient's heart had demonstrated wall-motion abnormalities as measured by X-ray contrast studies or echocardiography, and whether the patient had a previous history of hypertension. These choices were made by Dr. Blanck, the anesthesiologist in charge of the study, and were not objected to by any of the experts. Potential study patients were excluded from the study if they had any other serious illness beside the heart problem that led

them to have cardiac surgery. Only patients whose mean arterial pressure after cardiopulmonary bypass exceeded 100 mmHg or whose systolic blood pressure exceeded 120 mmHg were treated, and hence included in the study.

The choice of dependent variable was more difficult. The choice initially made supposed that the principal danger patients faced was that their blood pressure might drop too quickly by the treatment. The dependent variable was chosen to be the lowest value that mean arterial pressure reached in the half hour after commencement of treatment. Higher numbers are judged better for patients than lower numbers on this measure. There can be a variety of views about what measure appropriately reflects danger to the patient. Whatever measure is chosen, though, it must be one that the experts can use comfortably.

Having chosen the experts, and the predictor and dependent variables, the next step was to elicit the opinions of each expert about the dependent variable as a function of the predictor variables and treatment. For each treatment and expert, this took approximately one hour in a telephone interview using the methods described in Kadane et al. (1980).

The interviews have two rather distinct phases. In the first phase, the expert is asked for a median, a 75th percentile, and a 90th percentile for the dependent variable at various values of the predictor variables. For each expert this permits estimation of how much each predictor variable matters in determining the dependent variable. Those estimated values were read back to the expert, who in each case confirmed that they reasonably represented his view. The difference between the expert's medians and the model's fit to those medians were computed and made available to the expert. At several points in this process, experts could change their minds about values already given. Several did.

In the second phase, a hypothetical data set is built and shown to the expert. The expert is asked for a median at previously elicited values of the predictor variables, to see how much this opinion is influenced by the accumulating data. This in turn is used to estimate how sure the expert is about the estimates previously described. An expert who is very uncertain will be much influenced by the hypothetical data, and the converse is also true. Experts are asked not to forget any given hypothetical data, since this is an important design consideration in creating the method. To be useful, the hypothetical data must be somewhat different from what the expert thought would be most likely but must not be outlandish. One expert, incidentally, was sufficiently certain of his judgments that the hypothetical data set did not change any of his opinions, although he did say that for the less common treatment, verapamil, having seven patients all with higher than predicted mean arterial pressures was coming close to changing his mind a bit.

All the experts found that they were able to answer the questions posed. One factor that may have been important might have been that all the questions were about their medians for mean arterial pressure of patients with various characteristics (and here they had varying amount of information, so that they were answering in units that were within their experience). Their views on the elicitation process are recounted in Chapter 9.

The probability model underlying the elicitation is technically a normal linear model with a conjugate prior. The built-in model checks, based on redundancy of questions, indicated that it fit reasonably well in each case.

Once institutional review board approval was received, the elicitations were completed, and the requisite computer programs were written, we began the actual trial. When a potentially suitable patient, who has given informed consent, is found, the anesthesiologist at Johns Hopkins Hospital telephones the statistician at Carnegie Mellon and gives the patient's name, number, and description according to the four predictor variables. These are entered into the computer, which then computes, for each "expert," the predicted mean arterial pressure for a patient with the specified predictor variable values. If calculations for all five experts show that this is likely to be higher with one particular treatment, that treatment is assigned. If not, the treatment is assigned from a table in the computer that is created to maximize balance among the treatments, based on designs of Sedransk (1973), as described in Chapter 10. The assigned treatment is read over the telephone to the anesthesiologists at Johns Hopkins. Should the patient experience hypertension after cardiopulmonary bypass, but prior to the end of the operation, the assigned treatment is used. The patient's lowest mean arterial pressure is entered into the computer. If no hypertension develops, no treatment is used. Whichever the case, this information is reported to Carnegie Mellon and entered into the computer. When a treatment is used, the opinions of each expert are updated using Bayes's theorem to reflect the new data. These are then read for application on the next eligible patient. The computations are discussed in Chapter 8.

Chapter 11 gives my view of how successful we were in carrying out the trial. While our execution was certainly not flawless—we made mistakes—I nonetheless came away from the experience confident that the method proposed is feasible. In Chapter 12 and 13 we report the results of the trial.

1.7 CHARGES AND RESPONSES

This section repeats certain charges made or questions asked about the method of clinical trial described above, and to give answers that reflect the analysis developed to deal with them.

1. *Only strictly randomized trials are valid.*

   Response: I have four replies to this charge. First, many studies that appear to be randomized do not have a truly random mechanism determining treatment. As one example, some trials assign patients to treatments to maximize balance. Thus with the given characteristics and assignments to treatment of patients treated previously, and with the given characteristics of the current patient, the same assignment would be made each time. Such a trial is controlled, however, in that neither

the patient nor the attending physician knows what treatment would be assigned when the patient agrees to be in the trial. True randomization may be much less frequent in practice than is believed.

Second, practice is split on whether randomized trials are or should be checked for balance dynamically as the trial progresses. Not to check is more in accord with Fisher's original theory of randomization, and this is simpler procedurally. However, checking avoids obvious design catastrophes, such as assigning all men to one treatment and all women to the other. In some clinical trials the patients are entered sequentially, and balance is only examined (if at all) when the trial is over.

Fisher himself (see Savage 1962, p. 88) agreed that he would exclude certain regular designs when drawing a random Latin Square. Some clinical trials use this idea, dynamically checking balance on various covariates as patients are accrued. This has the advantage of avoiding design catastrophes, but it complicates the randomization analysis, making it more costly because the researcher has to specify exactly which (sequential) designs would have been excluded had they come up and what would have been done instead. I know of no clinical trial to have done so. Neither alternative strikes me as satisfactory.

One might use the admissibility criterion to limit eligibility to patients for whom every treatment in the trial would be admissible and then use random assignment. That approach is consistent with the ethical principle suggested here, but it is not what is done in the verapamil/nitroprusside experiment. Rather, patients for whom only one treatment is admissible are included in the study and assigned the admissible treatment. The alternative seems to me impractical in that it would exclude too many patients, particularly in a trial with several treatments.

Randomization is not the only way of thinking about statistics. If other considerations, such as treating patients well, suggest other methods, statisticians should be flexible enough to adjust. Alternative designs allow one to design more efficiently to achieve the purpose of the design.

Fourth, if one really insists on the randomization and is willing to do it right, that commitment is not barred by the ethical proposal made here. More about randomization and its place in statistical theory can be found in Kadane and Seidenfeld (1990).

2. *If patients cannot be allowed to choose their own treatment, at the very least, before agreeing to be the trial the patient should be told how each expert would vote on a patient with his or her characteristics.*

Response: To be told how each expert would vote is sometimes to be told what treatment would be assigned if the patient agreed to be in the study, and hence would be objectionable on the reasoning in Section 1.2. It would not have the same impact in the other cases. Since the expert

panel is chosen to represent the spectrum of the medical community but not to reflect the appropriate weight (if such could be specified) that should be given to each view, the "votes" could be quite misleading to an unknowledgeable patient.

3. *How should experts be chosen?*

   Response: As previously noted, with the admissibility rule used here, the experts should reflect the span of current respectable medical/scientific opinion on the disease and treatments in question. Adding an expert with views identical to someone already on the panel would have no effect on the trial. The composition of the panel should be included in the proposal to conduct the trial, and the reviewers should understand the spread of views on the panel and in the wider medical/scientific community.

4. *Physicians' opinions are unreliable and often wrong.*

   Response: Opinion is what patients go to doctors for. Medicine may not be so wonderfully scientific, but it will have to do until something better comes along. This design additionally incorporates a self-correcting feature through the Bayesian updating.

5. *Physicians, like other experts, tend to be too sure of their opinions.*

   Response: Although this is widely believed to be true, in my judgment, the evidence for over-confidence should be treated with some caution. It should be possible to correct for over-confidence by increasing the variance associated with various aspects of the opinion. Philosophically the trial's view of an expert opinion could be different from the expert's own view, as pointed out by Lindley et al. (1979). The trial of verapamil and nitroprusside described here took expert opinions at face value.

6. *How can a single, simple function of the patient's outcomes represent all the values and concerns of the patient?*

   Response: The function of patient outcomes used here to protect patients should not substitute for a full analysis of the data at the end of a trial, when a full and careful balancing of advantages and disadvantages of each treatment can be conducted. It has a more modest intent, to protect the patients from the single most pressing concern a physician might have. Because of this limited purpose, it may appropriately be changed as the trial takes place if the most pressing concern for patient safety changes.

7. *There is substantial evidence (see Kahneman et al. 1982) that opinion cannot be described well by Bayesian axioms. How reliable, then, are the calculations based on Bayes's theorem?*

Response: For this subtle matter, answers on several levels are appropriate. First, the Bayesian axioms are normative, not descriptive. Thus, they describe how a reasonable and rational person ought to make judgments under uncertainty, and not as they actually do. For example, there is evidence (Tversky, 1969) that in some circumstances subjects are not transitive in their preferences (i.e., they prefer A to B, B to C, and C to A). However, it would not benefit patients to model experts' behavior this way.

Second, it is reasonable to inquire about the extent to which the calculated posterior distributions approximate the experts' posterior distributions as they might be reelicited. As explained above, the elicitation questions asked using the Kadane et al. (1980) methods are predictive. Indeed the questions the experts are asked in the elicitation are very close to the predictions the computer model made about their opinions. Thus the Bayesian model can be used to smooth answers but would be more questionable if used to extrapolate far beyond the elicitations. For this reason, the results might be expected to be robust.

Additionally, if at any time, the statisticians in the trial suspect that the computed elicitations are no longer good approximations, they can re-elicit. In the trial described in Part II, we did just that, though not for the same reason (see Chapter 11 for a discussion).

8. *Patients do not withhold informed consent for the reasons addressed here, so why change?*

   Response: As discussed above, patients do not cognitively process informed consent very well. If you subscribe to the argument that it is incumbent upon the medical/scientific community to make the contract proposed to the patient as favorable to the patient as possible, then some change is warranted. The experimental results summarized in Meisel and Roth (1983) suggest that patients may fail to protest because they do not understand what they are agreeing to. While patients may not refuse to participate, physicians may refuse if they are not satisfied with the ethics of the design.

9. *The patients are not being individually examined by the experts, and they should be.*

   Response: If we could engage the attention of five world-renowned experts to personally examine each patient, that would be ideal. Clearly this is not practical, so we use instead computer models of the experts' opinions. It is possible to recalibrate them by giving the experts the data and reeliciting. For logistical reasons we have not done so in the trial reported here. A recalibration of these models would give some idea how good the models of experts' opinions are.

10. *Suppose that after a clinical trial is conducted, research shows some new predictor variable (e.g., some biochemical marker that was previously*

unknown) to be important in prognosis given treatment. In such a case, would it be necessary to redo the clinical trial?

Response: When a new predictor variable is discovered, neither Bayesian nor randomization analysis will be directly informative. A treatment judged best in a randomization analysis may be best only for some particular values of the predictor variables. Similarly the inclusion of a new predictor variable may change the results of a likelihood or Bayesian analysis.

11. *The Zelen (1979) alternative of randomly choosing some patients for the standard treatment, and others for informed-consent statements is preferable.*

Response: The Zelen design divides patients into two groups. Those in group 1 are given the standard treatment. Data about them is collected and used in the analysis whether or not the patient has agreed to this. Those in group 2 are asked to be in the trial and are given their choice of treatments. There is a serious difficulty with the Zelen design in that it requires patients to be in the study regardless of their preferences. Technically this is an invasion of their privacy. Nonetheless, the Zelen design is an interesting alternative worthy of discussion (see Zelen 1982).

12. *Definitions of admissible treatments for patients might be used other than the one-expert preference definition used here. Have these alternatives been explored?*

Response: Alternative admissible treatments were discussed at great length by the authors of this book. We considered, for example, circumstances of three treatments and two experts, whereby expert 1 rates treatments A, B, and C as 1.0, 0.9, and 0.0, respectively, while expert 2 rates treatments A, B, and C as 0.0, 0.9, and 1.0. In this case, by definition of admissibility, only treatments A and C would be admissible. Yet treatment B, a compromise treatment reasonably acceptable to both experts, may be a reasonable choice. We could even invent an expert 1.5, with preferences 0.5, 0.9, and 0.5, who would make treatment B admissible as well. This example suggests an extended definition of admissibility in which a treatment would be admissible if an expert or any convex combination of experts thinks that treatment is best for a particular patient. Chapter 3 looks at alternative definitions of admissibility.

The definition of admissibility used in the verapamil/nitroprusside trial ameliorates the problem of choosing experts because two experts in complete agreement will not affect the assignment of a patient any more than if there is only one such expert. Thus only the range of expert