The Decision To Initiate Clinical Trials of Current Medical Practices

James P. Kahan, C. R. Neu, Glenn T. Hammons, Bruce J. Hillman
The work reported here was supported by Contract No. 282-88-0068 with the Department of Health and Human Services, under the sponsorship of the National Center for Health Services Research and Health Care Technology Assessment.

ISBN: 0-8330-0680-0

The Rand Publication Series: The Report is the principal publication documenting and transmitting Rand’s major research findings and final research results. The Rand Note reports other outputs of sponsored research for general distribution. Publications of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

Published by The Rand Corporation
The Decision To Initiate Clinical Trials of Current Medical Practices

James P. Kahan, C. R. Neu,
Glenn T. Hammons, Bruce J. Hillman

September 1985

Prepared for the
National Center for Health Services
Research and Health Care Technology Assessment
FOREWORD

The Public Health Service, in recognition of the need to understand the process that stimulates the conduct of a randomized clinical trial of a medical practice in current use, contracted with The Rand Corporation in the fall of 1984 for a study of that process. Specifically, because clinical trials are conducted for various reasons, we asked the contractor to determine whether there are consistent criteria governing the decision to support a clinical trial of a current medical practice. It was our hope that from this study a model would emerge to guide policymakers. We are pleased that such a set of criteria has been identified and a model proposed that can lead to a more informed decision-making process with respect to when a clinical trial is appropriate and necessary.

John E. Marshall, Ph.D.
Director, National Center for Health Services Research and Health Care Technology Assessment
PREFACE

The randomized clinical trial is methodologically the strongest way to study the merits of medical practices. But such a trial is expensive and complicated, poses ethical issues, and is slow to reveal insights. Although its most common use is to test the efficacy and safety of innovative medical procedures, it can also test other features—such as cost-effectiveness or patient acceptability—not only of new but also of current practices. In the United States, the federal government funds most clinical trials through the National Institutes of Health (NIH).

Under contract to the National Center for Health Services Research and Health Care Technology Assessment (NCHSR), The Rand Corporation has examined clinical trials of current medical practices with the aim of obtaining a better understanding of the processes and criteria underlying NIH's decision to initiate such trials. This report presents findings based on a review of the literature, interviews with Public Health Service personnel and clinical trial investigators, and four case studies of proposed clinical trials for current treatments.

The report should be of interest to government policymakers and to members of the research and professional health care community, particularly those concerned with the use of clinical trials in medicine and with the question of how NIH policy directs the path of medical research in the United States.

The work reported here was supported by contract with the U.S. Department of Health and Human Services, under the sponsorship of NCHSR. Four case studies of the decision whether to conduct a clinical trial of a current practice were conducted as background information to this report. They are being published separately as the following Rand Notes:

SUMMARY

This report examines the decisionmaking process by which the National Institutes of Health (NIH) initiate a randomized clinical trial (RCT) of a medical practice in current use. By that we mean treatment, diagnostic, and preventive practices, including procedures and the use of devices and drugs, that are currently used routinely by some fraction of the practitioner community and are not regarded as experimental. These practices are sometimes called into question and RCTs are proposed to determine whether or to what extent they should remain part of medical practice. Our central concern is how NIH decides to apply clinical trials to the evaluation of these problematic current medical practices and how well NIH's approach works.

BACKGROUND

Our investigation was constructed of three parts: a review of the literature, interviews with Public Health Service (PHS) personnel and clinical trial investigators, and case studies of proposed clinical trials.

We selectively reviewed the literature, focusing on the question of when to initiate a clinical trial. Because of the great number of articles on clinical trials, we largely restricted our examination to large-scale, multicenter trials of current medical practices.

We interviewed PHS personnel involved in sponsoring clinical trials and investigators who conduct them to determine which factors were important in deciding which current medical practices are subjected to clinical trials and what the role of NIH is in supporting such trials. In-person interviews took place largely at offices in the Washington, D.C. area and at the 1985 meeting of the Society for Clinical Trials in New Orleans, Louisiana; in addition, we interviewed several individuals by telephone.

We conducted four case studies of clinical trials of current medical practices that were considered by NIH. For each, we interviewed NIH personnel involved in the project, the principal investigator(s), and other centrally involved people. NIH and the investigators provided access to the materials involved in processing the applications for funds. The case studies were:

- "Extracranial/Intracranial Arterial Anastomosis," a grant application funded by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS).
• “Intensive Care for Acute Myocardial Infarction,” a grant application that was not funded by the National Heart, Lung and Blood Institute (NHLBI).

• “Total Mastectomy vs. Segmental Mastectomy With or Without Radiation for the Treatment of Breast Cancer,” primarily a contract to the National Surgical Adjuvant Breast and Bowel Project (NSABP), supported by the National Cancer Institute (NCI).

• “Intermittent Positive Pressure Breathing,” contracts supported by NHLBI.

There is a consensus that the randomized clinical trial represents the “gold standard” of clinical research from a methodologic point of view, but some debate about when this standard should be applied in defining “medical truth,” especially with regard to current medical practices. Moreover, even among those advocating extensive use of clinical trials, there is some debate about the validity of different variations of basic RCT methodology. Further, although clinical trials consume a considerable fraction of the overall NIH research budget, very few of those funds go to trials of current practices.

The literature provides few clues of how the decision to initiate clinical trials is taken. The British Cancer Centre has an advisory process that, although of great potential influence, does not determine policy. The Veterans Administration Cooperative Studies Program is one of bottom-up initiation. All ideas for clinical trials arise from the individual investigators; no description of the program includes information about how these ideas originate. The National Heart, Lung, and Blood Institute does have an interactive system and provides criteria that are considered, including the state of biomedical science, the feasibility of a proposed RCT, the expected effect of the outcome of the trial, and ethical issues. Finally, in a model developed by Banta and Behney (1981) to describe the interaction between policymaking and technology assessment, empirical research generated by policy goals provides information on the state of medical technology, which in turn suggests the direction of policy. This model appears appropriate to policy decisions of which clinical trials to sponsor.

A DECISIONMAKING MODEL

The information provided by the literature is merged with information gathered by our case studies of four clinical trials of current medical treatments and by our interviews with key PHS personnel and clinical trial investigators. Together, these sources suggest awareness,
relevance, and feasibility as the components of a three-stage model that can both describe the present decisionmaking process and provide a framework within which suggested improvements to that process may be discussed.

**Awareness**

Most current medical practices are accepted for what they are supposed to do without serious question. By and large, this is appropriate; the medical arts comprise many useful and beneficial practices. But some current practices have come to be used without any thorough test of their benefit; for these, one must first become aware of a question of their appropriate usage before one can consider research to answer the question. Awareness can arise either through a common belief that the value of a particular practice is unknown or, more typically, through a difference of opinion among investigators or clinicians about the practice. Several avenues to awareness are possible:

- **Questions of Efficacy and Safety.** Whether a particular treatment does what it is purported to do and whether it is safe are the primary motivators for clinical trials in medicine. However, many medical practices have become established without any formal test of their efficacy or safety. Simple observational experience may lead to the awareness that a test of the appropriateness of the practice is desirable.

- **Advances in Biomedical Theory.** As understanding of disease processes improves, scientific theories of disease and disease processes change. As a consequence, certain medical practices that formerly were regarded as soundly based on scientific theory may be called into question. Clinical trials may then serve the dual purpose of determining the efficacy of a practice and generating or adding strength to a new scientific formulation.

- **Cost-Effective Health Services Delivery.** It may become desirable to know whether a particular medical practice provides equivalent benefits to alternative current practices that may be less costly. In such cases, it is necessary to know both the relative costs and the relative safety and efficacy of the competing practices.

- **Quality of Life.** The awareness that current medical practices are disfiguring, painful, and risky, or that they mandate severely restricted lifestyles, generates a call for research to find replacements that are equivalent in efficacy and safety.

- **Epidemiological Evidence.** Scrutiny of the frequency of medical practices across nations or among geographical regions
may reveal differences that cannot be accounted for by demographic characteristics. In such instances, it is reasonable to conjecture that the practice is either being underused or overused in some locations, and a clinical trial may be performed to clarify the appropriate uses of the procedure.

Relevance

The number of medical practices that are candidates for RCTs far exceeds the research community's capacity to perform them. A winnowing of possibilities will determine which current practices are the most fruitful targets of clinical trials. This takes place largely on the basis of the expected effects of the research on the world of clinical practice.

- **Impact of the Disease.** The impact of the medical problem and its treatment, both on the afflicted individual and on society, is the major determination of which treatments will be investigated. The more dire, the more prevalent, and the costlier the disease, the more it is likely to be the subject of research.

- **Political Environment.** The research community, like every other organization, is subject to pressures from interested parties, especially when they might directly or indirectly control resources. Pressure from the Congress, from interest groups organized about a particular disease, or from political or social movements can cause certain research questions to receive more attention.

- **The Window of Opportunity.** Research topics have their life cycles, and there is a recognition that the time may be too early or too late to perform a clinical trial of a practice. Among the factors that determine the window of opportunity for research are the availability of the practice, the availability of alternatives to it, and the level of disagreement regarding it in the medical community.

- **Anticipated Effect on Clinical Practice.** Even for current medical practices that the medical research community regards as needing scrutiny, clinical trials may not be undertaken if the results will simply be ignored by the practitioner community.

- **The Availability of Alternative Research Strategies.** Because of the large expense and effort of a clinical trial, if alternative means of investigation can answer a question with a good degree of certainty, those alternative means are likely to
be preferred to an RCT. Balanced against this search for alternative means is, of course, the recognition of the "gold standard" quality of clinical trials.

Feasibility

Even with a consensus that a current medical practice should be investigated, that the time is right for an RCT, and that an RCT is, in an ideal sense, the most appropriate way to perform the investigation, several constraints may still prevent the trial from coming into being. These are a mixture of technical, philosophical, and policy barriers that must be overcome.

- **Methodology.** In any major RCT, there are major technical, statistical, and logistical problems to be overcome. Among the specific concerns are the size of a clinical trial, its scope, the choice of variables to measure, the nature of the procedure that assigns patients to practices, and the interaction of the design of the trial with its objectives.

- **Ethics.** The moral appropriateness of the trial may be questioned. For clinical trials of current medical practices, a major ethical issue is that some patients will be deliberately deprived of the standard treatment for their condition.

- **Influence of Proponents.** It is true for clinical trials, as for most human endeavors requiring extensive coordination and organization, that some people have more influence than others. Senior Principal Investigators, influential NIH administrators, and other "people with clout" will have their opinions more attended to and the clinical trials they want to see done performed.

- **Institutional Factors.** The way in which an NIH Institute functions with respect to clinical trials can strongly influence whether potential clinical trials brought to its attention are supported. The institutional interaction with investigators is important. Institutes differ in their preferences for funding modality (grant vs. contract vs. cooperative agreement), in the degree to which they supervise the ongoing research conducted under their sponsorship, and in the degree of guidance they provide to researchers considering submitting grant applications or responding to contract solicitations.

Institutes also differ in the extent to which they have a targeted research program rather than allowing the research community, through the priority scores of Study Sections, to
govern their pattern of research. Some Institutes dedicate substantial funds to clinical trials or to specific research topics, and others minimize their degree of precommitment. Therefore, Institutes differ widely in their dedication to clinical trials as a mode of research and in the number and nature of clinical trials they sponsor.

The model outlined above is in a sense both a description of present practices and a prescription for how an effective system might work. In the prescriptive sense, this model represents a fairly comprehensive view of what the inputs should be into the decision to undertake a clinical trial. The weights that should be given the various factors when formulating policy (e.g., should NIH fund only trials with a strong biomedical science component, or should it also fund trials questioning the cost-effectiveness of health services delivery?) are not determinable by analytic means but rather are an expression of policy preferences. Most reasonable policies should be capable of being expressed within the model.

As a description of current practice, the model can help locate weaknesses in the process. The investigators and NIH administrators who determine which RCTs are conducted are generally aware of this model in only a vague sense; even when an Institute within NIH professed to have a systematic program for deciding what clinical trials to sponsor, many of the people we interviewed believed that the process was still too haphazard, with too much opportunity for error and bias.

CENTRAL ISSUES

In the course of our investigations, we found that our study of the decision to initiate clinical trials of current medical practices was really a study of a few recurrent issues of potential policy importance. The three most prominent of these issues are summarized below.

1. Does the NIH have a uniform process for deciding whether to initiate clinical trials of current practices? We conclude that NIH does not have a single process. Institutes differ in their programmatic interests, in their preferred ways of interacting with investigators, and in the means they use to keep aware of the medical research needs of the public. But because the problems that the Institutes investigate have different characteristics, differences in decisionmaking processes are neither surprising nor inappropriate. In most cases, NIH Institutional policy is an appropriate adaptation to its environment.

2. Is there an adequate system for identifying questionable current medical practices? We conclude that there is not.
Although NIH mechanisms for evaluating the relevance and feasibility of clinical trials for current medical practices appear adequate, questionable medical practices become obvious haphazardly. We advocate the construction of a systematic means of becoming aware of problematic practices, using widely varying sources of data.

3. How should clinical trials on questions of the cost-effectiveness of medical practices be sponsored? At present, no single agency has a mandate to sponsor clinical trials that directly address cost-effectiveness issues. Although such clinical trials should be supported, it is not clear which agency is the appropriate source of that support. Moreover, there are philosophical and practical obstacles to conducting clinical trials of the cost-effectiveness of current practices.
ACKNOWLEDGMENTS

We interviewed many people in the course of this project; they are listed elsewhere. We sincerely thank them for their cooperation. Rand colleagues Geoffrey Anderson and Elizabeth McGlynn researched case studies and critiqued our decisionmaking model as it evolved. Sandra Poindexter aided in data analysis. Robert Brook and Albert Williams listened and read patiently and added their wisdom at important times.

Earlier drafts of this manuscript were constructively critiqued by the late Robert Gordon at NIH, John Marshall and Norman Weissman at NCHSR, Curt Furberg at NHLBI, and Cheryl Austerin at the Office of the Assistant Secretary for Planning and Evaluation, Department of Health and Human Services. We also obtained excellent reviews from Rand colleagues Emmett Keeler, Kathleen Lohr, and Richard Rettig.

Finally, we wish to thank Ann Haendel, our project monitor at NCHSR, for her own constructive criticisms of earlier manuscripts and for effectively and efficiently coordinating the diverse individuals and groups involved in this project so that a final product satisfactory to all might emerge.
CONTENTS

FOREWORD ......................................................... iii
PREFACE ....................................................... v
SUMMARY ........................................................ vii
ACKNOWLEDGMENTS .............................................. xv
FIGURES AND TABLE ............................................... xix
ACRONYMS ............................................................ xxi

Section

I. INTRODUCTION ...................................................... 1
   The Randomized Clinical Trial ............................ 1
   Current Medical Practices ................................. 1
   The Present Study ........................................... 2

II. BACKGROUND ...................................................... 5
   Clinical Trials in Medicine ............................... 5
   Clinical Trials and Current Medical Practices ........... 9
   Federal Support for Clinical Trials ....................... 13

III. DECIDING TO INITIATE A CLINICAL TRIAL:
    FOUR DESCRIPTIVE MODELS .............................. 19
    A British Cancer Centre ................................. 19
    The VA Cooperative Studies Program ................... 20
    Large-Scale Trials at NHLBI ............................. 21
    A Model of Social Policy and Technology Assessment ... 23

IV. A DECISIONMAKING MODEL: OVERVIEW .................... 26
    The Three Stages of the Decisionmaking Model .......... 27
    How Well Does the Model Describe Current Decisionmaking? .... 28

V. THE AWARENESS STAGE ......................................... 29
    Issues of Efficacy and Safety ......................... 29
    Biomedical Theory ........................................ 31
    Issues of Cost-Effective Health Services Delivery .... 33
    Quality of Life ............................................. 38
    Epidemiological Evidence ................................ 39

VI. THE RELEVANCE STAGE ......................................... 42
    Impact of the Disease .................................... 42
    Political Environment .................................... 44
Window of Opportunity ......................... 46
Anticipated Change in Clinical Practice ........ 48
Availability of Alternative Methodologies ...... 51

VII. THE FEASIBILITY STAGE .................. 54
     Methodological Constraints ................. 54
     Ethical Constraints ....................... 60
     Influence of Proponents ................... 64
     Institutional Factors ...................... 65

VIII. ISSUES IN INITIATING CLINICAL TRIALS ...... 69
      Does NIH Have a Uniform Decisionmaking Process? 69
      Identifying Questionable Current Medical Practices 71
      Sponsoring Clinical Trials to Study Cost-Effectiveness 73

Appendix
  A. LIST OF RESPONDENTS ..................... 77
  B. SUMMARIES OF CASE STUDIES .............. 80

BIBLIOGRAPHY .................................. 87
FIGURES

1. Distribution of Total Support for 1979 NIH-sponsored Clinical Trials ........................................ 16
2. Social Policy and Technology Assessment ..................... 24
3. The Decision To Initiate a Clinical Trial of a Current Practice .................................................. 26
4. Cost-effectiveness and Efficacy as Sources of Awareness .. 35

TABLE

1. 1979 NIH Mean Total Support for Clinical Trials .......... 15
# ACRONYMS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCU</td>
<td>Coronary Care Unit</td>
</tr>
<tr>
<td>CDC</td>
<td>Centers for Disease Control</td>
</tr>
<tr>
<td>CSEC</td>
<td>Cooperative Studies Evaluation Committee</td>
</tr>
<tr>
<td>DRG</td>
<td>Diagnosis Related Group</td>
</tr>
<tr>
<td>EC/IC</td>
<td>Extracranial/Intracranial Arterial Anastomosis</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration</td>
</tr>
<tr>
<td>FY</td>
<td>Fiscal Year</td>
</tr>
<tr>
<td>HCFA</td>
<td>Health Care Finance Administration</td>
</tr>
<tr>
<td>HHS</td>
<td>Health and Human Services</td>
</tr>
<tr>
<td>ICU</td>
<td>Intensive Care Unit</td>
</tr>
<tr>
<td>IPPB</td>
<td>Intermittent Positive Pressure Breathing</td>
</tr>
<tr>
<td>IRB</td>
<td>Institutional Review Board</td>
</tr>
<tr>
<td>MRFIT</td>
<td>Multiple Risk Factor Intervention Trial</td>
</tr>
<tr>
<td>NCHS</td>
<td>National Center for Health Statistics</td>
</tr>
<tr>
<td>NCHSR</td>
<td>National Center for Health Services Research and Health Care Technology Assessment</td>
</tr>
<tr>
<td>NCI</td>
<td>National Cancer Institute</td>
</tr>
<tr>
<td>NEI</td>
<td>National Eye Institute</td>
</tr>
<tr>
<td>NHLBI</td>
<td>National Heart, Lung, and Blood Institute</td>
</tr>
<tr>
<td>NIADDK</td>
<td>National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases</td>
</tr>
<tr>
<td>NIH</td>
<td>National Institutes of Health</td>
</tr>
<tr>
<td>NINCDS</td>
<td>National Institute of Neurological and Communicative Disorders and Stroke</td>
</tr>
<tr>
<td>NSABP</td>
<td>National Surgical Adjuvant Breast and Bowel Project</td>
</tr>
<tr>
<td>OTA</td>
<td>Office of Technology Assessment of the United States Congress</td>
</tr>
<tr>
<td>PHS</td>
<td>Public Health Service</td>
</tr>
<tr>
<td>PI</td>
<td>Principal Investigator</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomized Clinical Trial</td>
</tr>
<tr>
<td>RFP</td>
<td>Request for Proposals</td>
</tr>
<tr>
<td>VA</td>
<td>Veterans Administration</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

THE RANDOMIZED CLINICAL TRIAL

The clinical trial is an experiment in medicine in which one group of patients receives one treatment and another group receives another treatment; the two groups of patients are later compared on the basis of some measure (e.g., Byar et al., 1976; Chalmers, 1981; Friedman, Furberg, and DeMets, 1981; Mosteller, Gilbert, and McPeek, 1980, 1983; Schwartz, Flamant, and Lellouch, 1980). Although the patients may be assigned to treatment in many ways, there is an established (but not yet universal) consensus for the value of having this assignment done by a random process whenever possible. In such a study, the trial is called a randomized clinical trial (RCT).\(^1\) The randomization procedure limits to an investigator-specified level the probability that any observed differences in outcome among treatments are due to pre-existing differences among the members of the treatment groups. It is then possible to declare that any observed differences are due to treatment administrations with only the specified probability of error. In this way, unknown factors that could masquerade as treatment effects are statistically “controlled.”\(^2\)

CURRENT MEDICAL PRACTICES

This report examines the process by which the NIH decide to initiate an RCT of a medical practice in current use. By “medical practices in current use” we mean treatment, diagnostic, and preventive practices, including procedures and the use of devices and drugs, that

---

\(^1\)This two-treatment description is the simplest type of RCT. There can be more than two treatments, and patients can be assigned to a sequence of different treatments. “Treatments” can include placebos or just doing nothing. The essential feature of the RCT is the random assignment of the particular patient to a predefined sequence of treatments.

\(^2\)Studies can have many different possible designs, which vary in the nature and degree of the randomization process. At one extreme, all participants in the study are lumped into a single pool and randomly assigned to treatment. This limits the likelihood of “false positive” findings but provides no information about potential biasing characteristics of patients or environments that might influence treatment. An alternative is to “stratify” the participant population along known characteristics that are thought to be related to the effects of the treatment (such as age, sex, severity of illness, treatment, site), and randomize to treatment within strata. The tradeoffs between stratification and pure randomization strategies have been extensively discussed in the statistical literature.
are used routinely by some fraction of the practitioner community. This fraction need not be the majority of the community, but only large enough that the practice in question has some effect on the overall picture of a disease or condition. As most observers would agree, acceptance of a medical practice by the clinical community does not guarantee its appropriateness: The practice may be obsolete, it may never have worked in the first place, it may not be safe, it may be generalized beyond its established and appropriate purposes, or it may not be the most cost-effective technology. These questionable practices may be almost universally accepted by the clinical and academic communities, there may be divisions of belief within either, or there may be a consensus that the validity of the practice is unknown.

Three topics chosen for our case studies provide examples of a range of current medical practices. Extracranial-intracranial arterial anastomosis is a widely used surgical procedure that shunts blood flow to internal arteries that feed the brain. The RCT examined whether this procedure actually helped prevent stroke, as its proponents believed. Intensive care units are routinely used for victims of acute myocardial infarction; the RCT proposed (but not accepted) would have compared the benefits of that treatment for uncomplicated cases with the less intensive and possibly less expensive monitored hospital bed. Total mastectomy (removal of the entire breast) is the consensually recognized standard treatment for early (stage I and stage II) breast cancer; many physicians believed that partial mastectomy (removing only the tumor and some surrounding tissue) was equally effective. Surgeons currently use both procedures. The RCT we examined compared the two treatments.

THE PRESENT STUDY

At the request of the National Center for Health Services Research and Health Care Technology Assessment (NCHSR), The Rand Corporation undertook an investigation of the processes underlying decisions by the different Institutes of the NIH to support RCTs of medical practices in current use. Our investigation comprised three facets, a literature review, interviews, and case studies.

---

3In more traditional terms, current medical practices comprise accepted and established practices, new but nonexperimental practices, and almost abandoned practices. The term is meant to exclude medical technologies that are still in the developmental or experimental stage or "fringe" practices (such as Laetrile) that are consensually disavowed.
The literature review was a very narrowly focused examination of the decision to initiate RCTs, particularly RCTs of current medical practices. This literature, which is not large, is almost exclusively concerned with large-scale, multicenter clinical trials.

Our interviews were with NIH and other governmental personnel involved with RCTs and with investigators who conduct trials. The majority of the interviews were in person, at NIH and other locations in the Baltimore-Washington area or at the 1985 meeting of the Society for Clinical Trials, held in New Orleans, Louisiana. Several people were interviewed when they visited The Rand Corporation in Santa Monica, California. Additionally, a few interviews were conducted by telephone. The list of persons interviewed is presented as Appendix A of this report. The content of the interviews varied with the particular expertise and interest of the respondent, but generally centered around their perceptions of the process by which current medical practices are identified as requiring further study, how RCTs are initiated, and the role of NIH and other agencies in supporting RCTs.

Case studies were done of four proposed RCTs of current medical treatments. The four were selected by the NCHSR and NIH personnel overseeing this project from a list submitted by The Rand Corporation and are intended to represent diverse Institutes of NIH, funding modalities, and medical problems. The senior investigator of each study was an individual associated with Rand who was knowledgeable about the content of the trial. The case studies were of the following projects:

- "Extracranial/Intracranial Arterial Anastomosis," a grant funded by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS). This case study was researched by Geoffrey M. Anderson, M.D.
- "Effectiveness of Intensive Care for Acute Myocardial Infarction," a grant application to the National Heart, Lung and Blood Institute (NHLBI) that was not funded. This case study was researched by Glenn T. Hammons, M.D.
- "Total Mastectomy vs. Segmental Mastectomy With or Without Radiation for the Treatment of Breast Cancer," primarily a contract to the National Surgical Adjuvant Breast and Bowel Project (NSABP) from the National Cancer Institute

---

4All respondents were cooperative, informative, gracious, and generous with their time. We thank all of them once again for their assistance.

5A fifth clinical trial was also examined, but turned out to be a test of a totally new application of a developing technology rather than a current practice. Therefore, it was deemed not suitable for the present project and we dropped it from the analysis.
(NCI). This case study was researched by Bruce J. Hillman, M.D.

- "Intermittent Positive Pressure Breathing," contracts to multiple centers from NHLBI. This case study was researched by Elizabeth A. McGlynn, M.P.P.

Each of these case studies has been separately published as a Rand Note (Anderson and Kahan, 1985; Hammons and Kahan, 1985; Hillman and Kahan, 1985; McGlynn and Kahan, 1985) and is available upon request. Summaries of the case studies are reproduced in Appendix B.

We have combined the information gained from the literature review, interviews, and case studies. The information from the interviews will be presented indirectly, without attribution to individuals or organizations, because we believed it prudent to assure the respondents' anonymity. Except in the last section, where we summarize our own viewpoint, opinions expressed in the text are those reported by our respondents. When a position is attributed to a respondent or when reference is made to a respondent's institutional affiliation, it is with that person's expressed permission.
II. BACKGROUND

CLINICAL TRIALS IN MEDICINE

There is a rapidly growing literature on RCTs in medicine.\(^1\) Interest in the method has spawned a Society for Clinical Trials, which is now in the sixth year of publication of its journal, Controlled Clinical Trials. It is common (e.g., Hopwood, Mabry, and Sibley, 1980) to distinguish five stages of the clinical trial, namely (1) initiation, (2) planning, (3) start up, (4) conduct, and (5) analysis and reporting. Throughout this report, our focus will be almost exclusively on initiation, or how it is decided whether to undertake a clinical trial of a current medical practice. Except as they impinge on the decision to initiate a trial, we shall not discuss issues pertaining to methodological details of clinical trials, their statistical analyses or the dissemination of their results, although all of these are certainly of major importance (e.g., Banta and Behney, 1981; Banta, Behney, and Willems, 1981; Shapiro and Louis, 1983; Simon and Wittes, 1985).

It is popular in the literature on clinical trials (e.g., Feinstein, 1984b; Mosteller, Gilbert, and McPeek, 1983; Office of Technology Assessment, 1983) to accord to Daniel (of lion’s den fame) the honor of having conducted the first clinical trial. He persuaded Nebuchadnezzar’s eunuch to sponsor a clinical trial in which one group (Daniel and his friends) would be fed traditional Judean fare and the other (“the children that eat of the portion of the king’s meat”) would eat the standard Babylonian royal diet. In spite of several weaknesses in the method of the study, including a total absence of randomization, this comparison of two standard nutritional practices was so successful that Daniel and his fellow captives were permitted to eat as their religion dictated.

The Randomized Clinical Trial as a Methodological Yardstick

With the development of the modern scientific method and its application to medicine, the RCT has become the “gold standard” against which other means of investigation are measured. Brown (1984) claims the strength of the RCT is that it uniquely meets all of the following five research standards: (1) the trial directly examines the disease and

\(^1\)Our emphasis here is on RCTs, a specific type of clinical trial. Unless otherwise noted, whenever we refer to a “clinical trial” or a “trial,” we mean an RCT.
type of patients addressed by the research question, (2) it uses the medical treatments in question, (3) it is statistically sound, (4) it requires a well-defined medical protocol, and (5) it can be communicated in unambiguous language. There is increasing reliance on clinical trials as a main source of evidence for the safety and efficacy of medical treatments (e.g., Brandt, 1985; Foster, 1984; Gest, 1984; Office of Technology Assessment, 1978, 1983). In the United States, as well as in many other nations, clinical trials (but not necessarily RCTs) of new drugs are required before a pharmaceutical firm can introduce the new product to the marketplace. Many authors argue that correspondingly stringent requirements should surround the introduction of innovations in diagnostic techniques (e.g., Alperovitch, 1984; Moskowitz et al., 1983; Zweig and Robertson, 1982), surgical techniques (e.g., Bearman, 1983; Haines, 1983; van der Linden, 1980; Spodick, 1982b), prevention techniques (e.g., DeWys and Greenwald, 1983), as well as medical treatment (e.g., Tubiana, 1980, for cancer, and Peto, 1983; Smith, 1983, for rheumatoid arthritis).

The Office of Technology Assessment (1983) views clinical trials as useful in improving the assessment of medical technologies, including the following:

- Comparing the safety and efficacy of new with extant technologies,
- Testing the relative efficacy of alternative technologies,
- Determining the optimal way to use a technology to achieve a desired effect, and
- Determining the generalizability of a technology's effectiveness over a population of potential patients.

In addition, clinical trials can have a role, although not a definitive one, in assessing the cost-effectiveness of technologies (Office of Technology Assessment, 1980). A clinical trial should be undertaken when it is determined that such a study would provide important information for any of these aspects of medical technology assessment (Office of Technology Assessment, 1982).

This is not to say that clinical trials are universally acclaimed as the only way to discover medical truths. New treatments that produce immediate and dramatic effects do not require RCTs to prove their efficacy; many current practices are so well-established by experience, scientific theory, and observable effect that RCTs that deprive patients of them would be unnecessary, not to mention unethical. But even when research about a practice is desirable, the RCT may not be the preferred method. Lorenz (1982) enumerates some major objections to
clinical trials, including inefficiency, immorality, and irrelevancy; he argues that although clinical trials have been unjustly maligned, not every clinical question has to be answered by a controlled trial. Fyfe (1984) believes that RCTs, originally designed for agricultural research, have been overgeneralized in their application to medicine. He believes that the questions an RCT can answer are quite limited in scope and that the method has been overused, sloppily applied, and misinterpreted. Feinstein (1984a, 1984b) and Gehan (1984) list situations in which they believe that observational studies or historical control studies should be preferred to RCTs. We discuss such considerations below. But, as Bailar (1983, p. 1) puts it,

The method of randomized clinical trials (RCTs) is a last resort for the evaluation of medical interventions. It is slow, ponderous, expensive, and often stifling of scientific imagination and creative changes in ongoing protocols.... [But] no other method for studying the merits of clinical treatment regimens can approach the precision of estimating effects and the strength of inference permitted by sound RCTs.

The case is further put by Remington (1979, p. 1607):

Opponents of allocation to clinical trials sometimes argue such studies often produce negative results. This is a curious position, since [the goal of medical research] should be to prove effective treatment and to eliminate treatments that are worthless, whatever opinion exists to support these worthless treatments in the absence of rigorous scientific justification based on controlled trials. Furthermore, our reluctance to embark early on randomized trials often means that such trials are only initiated after some substantial doubt arises concerning conventional therapy.

Thomas C. Chalmers is perhaps the foremost proponent of the absolute need for RCTs in medicine. His position, often expressed, is that when considering potential innovations in medicine, randomization to treatment should begin with the first patient. Chalmers includes such innovations as the artificial heart, insulin pumps, and chemotherapeutic innovations in cancer treatment in his dictum. Moreover, he believes that many current practices are ineffective and that RCTs should be performed to identify and eliminate these practices. Although noting (Chalmers, 1981) that RCTs have become an accepted fixture in medical research, he finds that they have not been accepted as evidence for changing medical practice as much as they could be.

2This extreme position has not gone unchallenged. See Hollenberg, Dzau, and Williams (1980) and Sacks, Kupfer, and Chalmers (1980) for an enlightening exchange of the issues.
He documents this belief with examples of medical practices that RCTs showed to be ineffective yet were not abandoned, if at all, until years after publication of the trial results. He acknowledges that RCTs are expensive to perform, but he argues that the return in not using inappropriate medical technologies will far outstrip the outlay in research.

There appears to be some concern that, with the general overall improvement in health care, tests of practices that compare present outcomes with earlier ones can misleadingly conclude that the more modern practice is superior. Sacks, Chalmers, and Smith (1983) compared RCTs with historical control trials (where outcomes of a treatment are compared before and after the treatment was introduced). They found that in general historical control trials showed the new therapies to be effective, and RCTs showed that the current treatments were not responsible for the improvement in outcomes.

In a study of clinical trials of early contact and maternal-infant behavior, Thomson and Kramer (1984) noted that these studies were generally weak on randomization procedures and subject bias. Of 13 clinical trials they examined, only three showed no effect of early contact on maternal-infant behavior, but of the five methodologically best studies, two showed no effects. Here, too, randomization led to a greater likelihood of treatment being shown to be ineffective and to the strong hypothesis that a study in which mothers were randomly assigned to amount of early contact (impractical from both logistical and ethical points of view) would show no differences.

In summary, RCTs have come to be regarded as the "gold standard" of medical research, not only for testing new ideas in medicine, but also as a debunking tool for dispelling popularly held but erroneous notions and for invalidating the results of less rigorous studies. Although the case for RCTs to investigate innovations is compelling and, indeed, is often taken as the use of RCTs, randomized clinical trials may be usefully conducted for current medical practices as well.

Costs Vs. Benefits of Randomized Clinical Trials

Whatever the benefits that arise from RCTs, they are often very expensive to conduct. It is not uncommon for large, multicenter trials to involve thousands of patients and scores of physicians, nurses, data collectors, and statisticians. The most important obstacle to performing RCTs today may be their enormous costs.

Several authors have argued that despite these high costs, RCTs are good investments, ultimately saving society much more than they cost to perform. Chalmers (1984) notes that in an environment with
limited resources, all patients cannot be given access to all medical services. In this environment, he argues, it is essential to use clinical trials to decide what is and what is not of benefit; the alternative is to risk engaging in unethical medicine, where unnecessary or ineffectacious procedures preempt worthy ones.

Bunker, Barnes, and Mosteller (1977) view clinical trials as a necessary component of cost containment in a society that cannot afford medical care in its fullest for everybody. They state (p. 390):

We have learned from our review of surgical experiments in humans that the costs in dollars and in lives of a poor experiment, or of no experiment, are often much greater than those of a well designed and executed clinical trial.

Although it is not possible or necessary to conduct RCTs to evaluate all new surgical procedures (much less all extant ones), when there is uncertainty about the outcomes of a procedure or its true costs, then an RCT becomes useful.

Hiatt (1975), reiterating the oft-made observation that the pool of resources available for medical care is limited, discusses considerations relevant to the allocation of those resources. Among these is the value of a medical technology: Many current medical practices are of no value or of undetermined value, and we must have a means of identifying these practices. The RCT is a prominent candidate for such a means.

R. Schwartz (1983) makes explicit the argument that cost-benefit analysis should enter into considerations of whether to undertake clinical research; the expected gains to society should outweigh the costs of the research. Mantel (1990, p. 9) agrees:

The cost of continued investigation [by clinical trial of a technology] should not be measured in dollars but in the loss of results of possibly more fruitful investigations that could have been conducted with those same dollars.

**CLINICAL TRIALS AND CURRENT MEDICAL PRACTICES**

An example of RCTs of current medical practices is one by Mathur et al. (1984), who examined the effect of diuretics on cardiopulmonary performance in patients with severe chronic airflow obstruction. Diuretics, standard treatment for this condition, are given to most patients. In this small study, patients were divided into two groups, based on whether they had congestive heart failure. Within each group, they were given either a diuretic or a placebo, in a crossover
design. The group with congestive heart failure was found to need diuretics; indeed, taking them off the medication led to problems so rapidly that the study was abandoned and the entire group was maintained on diuretics. For the second group, however, there were no differences between placebo and diuretic therapy, leading to a conclusion that diuretics benefit only patients with chronic airflow obstruction who have clinical features of congestive heart failure. Mathur et al. conclude that because there are possible negative side effects from diuretics, their administration when unnecessary may even be harmful for the patient.

The Life Cycle of a Medical Practice

To demonstrate a more general utility for trials of current practices requires a model of the life cycle of medical technologies that includes a place for examinations of current practices. Although such a model has not yet been developed and may never be, there are some indications of what it might look like.

McKinlay (1981) identified stereotypic stages in the life of a medical technology. It begins as a promising innovation, based on one study or a small number of pilot studies. Through peer diffusion, including promotion by its advocates in medical and perhaps public media and by word of mouth, it gains professional acceptance. This is followed by public acceptance and acceptability by insurance carriers for reimbursement, which marks its establishment as a standard procedure. Then, doubts begin to arise, perhaps marked by a clinical trial. This leads to professional denunciation, beginning with the scientifically oriented community and eventually filtering down to practitioners. Finally, the technology is discredited, either vanishing quickly or withering away as a new generation of practitioners eschews its use. McKinlay argues that conducting clinical trials before the technology becomes standard procedure could eliminate suspect medical technologies earlier, but for the many technologies that become standard procedures without a formal scientific test of their efficacy, clinical trials should eventually be required.

Barnes (1977) examined case studies of proposed innovations in surgery that had brief periods of popularity but were later shown to be valueless and possibly dangerous. He found, generally, that the

3In a crossover design, each patient serves as his own control, receiving the treatment for one period of time and the placebo for the other. Randomization determines only the order of administration, as all patients receive both treatments. When there are no long-term effects of the medication, this design permits testing with smaller numbers of patients.
The adoption of a procedure came about because it fit the then medical model for the disease and because prestigious surgeons claimed success in its use. A lack of control experience through the adoption phase, uncritical acceptance of “conventional wisdom,” a less-restrictive than present-day set of ethics, and ignorance of statistics all helped maintain the popularity of these techniques. Fisher (1984) found essentially the same process in his examination of the history of the Halsted radical mastectomy for breast cancer.

Finkelstein (1983) looked at the usage patterns of eight drugs introduced and then abandoned in the 1960s and 1970s. He found that models of technology acquisition could not fit the record of technology abandonment; most of the drugs experienced a pattern of growing and then steady usage, only to suffer a sharp drop and then abandonment. This investigation did not look for signal events that could have triggered the drop in usage. There is a need for further research of the life history of abandoned medical technologies to assess the accuracy of the models proposed by McKinlay, Barnes, and Finkelstein.

**Error in Medicine**

RCTs can have an important role in the detection of error in medicine. That there has been error in medicine is denied by nobody. Horror stories abound, even in the “modern era” following the introduction of anesthesia and antisepsics. Robin (1984) coined the term “iatroepidemics” to describe the harm done by systematic errors incorporated into medical practice, and enumerated two dozen of them that have occurred in the last century. Robin strongly recommends clinical trials as the best means of preventing iatroepidemics.

McIntyre and Popper (1983) observe that the medical community is reluctant to accept the findings of studies indicating practices are inefficacious. They view this reluctance as a consequence of a mistaken view of what constitutes progress in medical science (p. 1920):

The old view of the growth of knowledge, especially of scientific knowledge, is still widely held. According to this view, knowledge grows by accumulation: we discover and collect more and more facts. This view is not, of course, totally mistaken. Knowledge does grow, here and there, by accumulation. Yet far more often knowledge grows by the recognition of error—by the overthrow of old knowledge and mistaken theories.

McIntyre and Popper call for a revision in medical ethics to recognize that errors are part and parcel of any system of knowledge. Therefore, there should be research even on consensually accepted practices because there is always the possibility that the consensual opinion may
be in error. For this task, RCTs have a role exactly comparable to the
one they have in testing innovative practices. Lorenz and Rohde
(1979, p. 301) state this role directly:

The aim of the controlled trials is not the introduction of sensational
novelties into medicine, but the abolition and prevention of unneces-
sary and unworthy modes.

Clinical trials have been directly involved in some major discoveries
of error in medical practice. Barsamian (1977) examined how RCTs
led to the abandonment of internal mammary artery ligation for the
treatment of angina pectoris. This technique, which involves surgery
to tie off arteries feeding the chest, gained popularity in the 1930s. It
was simple, fairly safe, seemingly effective, and fit a theoretical model
that cutting off blood to surface areas would increase the supply avail-
able to the heart. But two studies were conducted in which patients
were subjected to a sham operation; that is, the entire operation was
performed, except that the arteries were not tied. Barsamian (1977,
p. 213) views this study as of major importance in the history of medi-
cal research:

Finally, the mammary artery ligation has a classic status as the
operation that forced recognition of the placebo effect of treat-
ment—whether medical or surgical—as no other form of treat-
ment had demonstrated quite so decisively. Because of this very sim-
ple and clear-cut demonstration of the placebo effect, the logical
demand for controlled studies of surgery gained recognition and
acceptance.

Miao (1977) presented a case study of another major instance of an
RCT leading to the abandonment of an established practice. Gastric
freezing for duodenal ulcers was introduced in the late 1950s by Dr.
Owen Wangensteen, a prominent surgeon, as a reasonable, fairly inex-
pensive, and effective treatment. The patient swallowed a balloon,
which, when it was in the patient’s stomach, was filled with a freezing
solution. By contact, the area around the ulcer was frozen, bleeding
was stopped, and the ulcer was given a chance to heal. But RCTs,
using a placebo of water at body temperature instead of the freezing
solution, showed no differences in outcome due to treatment, and gas-
tric freezing was rapidly abandoned.

---

4That such a clinical trial would be impossible given modern ethical standards
presents an interesting paradox. Although we would be aghast at the prospect of a
patient being subjected to the risks of anesthesia and incision as a sham, we are equally
aghast at the prospect of the full treatment when it has no effect. But how is one to
learn about the latter without doing the former?
Chalmers (1974) cited other instances in which clinical trials clearly indicated that an accepted medical practice was in error, including stilbestrol to prevent spontaneous abortion, bed rest for viral hepatitis, a bland diet (the "Sippy diet") for acute peptic ulcers, and oral medicines to control diabetes. But unlike the cases of mammary artery ligation and gastric freezing, these techniques were abandoned, if at all, long after the evidence from clinical trials was available. Chalmers argues from these case studies that RCTs can be used to test standard therapies and that changes in how research is disseminated to the practitioner community is necessary to increase the effect of RCTs.

FEDERAL SUPPORT FOR CLINICAL TRIALS

The federal government, as the major funder of both basic biomedical research and health services research, is the primary source of support for clinical trials. It provides over three-fifths of all health research monies (National Center for Health Statistics, 1980). Because of this preponderance of support, governmental policy strongly determines the direction of health research (Banta, 1982). Although there is no disagreement that clinical trials in general deserve support, the specific details of which agencies should have primary responsibility for funding, the amount of support that should be provided, and the relative roles of private and public funding have been topics for debate (Institute of Medicine, 1983). Clinical trials are seen as a necessary component to the task of providing efficacious, safe, and cost-effective medical services to the American public (Office of Technology Assessment, 1978, 1980, 1982, 1983).

Federal support for clinical trials is largely through the NIH and the Veterans Administration (VA), although other agencies also sponsor clinical trials to a lesser extent (Meinert, 1982). Within the VA, of a 1979 research and development budget of $126.3 million, $8.5 million, or 6.7 percent, went for multicenter trials (Meinert, 1982).

The VA expenditures do not include the costs of patient care, which are assigned to normal VA hospitalization budgets; the true cost of

---

5The Food and Drug Administration (FDA) does not fund clinical trials, but it is instrumental in bringing about many trials. For a new drug to be marketed in the United States, the FDA must approve it as "safe and effective." The FDA decision to approve a drug for marketing is based on information generated by a clinical trial paid for by the manufacturer of the drug in question. The 1976 Medical Device Amendments to the Food, Drug and Cosmetic Act extends FDA authority to medical devices as well as drugs. As with drugs, manufacturers seeking FDA approval for a new medical device must pay for the necessary clinical trials. There is no requirement for FDA approval of new medical procedures.
VA-sponsored clinical trials could well be much higher, depending on how much additional care is provided patients in trials. With respect to the question of patient costs, Yarbro and Mortenson (1985) claim that patients in clinical trials have much greater costs than those not in trials, and they worry about the effects of prospective payment through Diagnosis Related Groups (DRGs) on trials. They argue that prospective payment may cause a disincentive for hospitals to participate in trials and suggest a new diagnosis category for participants in trials that would carry supplemental funding.6

The NIH is the major locus of clinical trial research. Malone (1985) reports that in the fiscal year (FY) 1984, $235.4 million were spent on clinical trials. This represents 6.6 percent of the total NIH FY 1984 outlay of $4.15 billion. About 14 percent of clinical trial money is spent on intramural projects; the remainder supports extramural research, largely conducted through research facilities of academic medical centers. The lion's share of funds for clinical trials goes to the National Cancer Institute (NCI), which spent over $150 million on clinical trials in FY 1984.

Unfortunately, the data available to Malone did not include information about whether a clinical trial tested a current medical practice; to examine this question, we were forced to look at data from 1979, the last year the NIH Inventory of Clinical Trials was conducted.7 This Inventory enumerated all contracts, grants, and intramural projects that were clinical trials; it also provided information about the nature of each trial, including whether the trial was one of a "standard intervention," a "new intervention," or "both." The analyses below are based on 1979 figures. We caution that the proportions might be out of date in terms of current funding policies.8

We examined the 234 trials that were listed as "standard interventions" in the 1979 Inventory. It turned out that many of these were "standard" in the technical sense of involving innovative combinations of standard practices (particularly for tests of drugs for combatting cancer). This use of standard interventions did not correspond to our.

---

6Carolyne Davis (1985), then director of the Health Care Finance Administration (HCFA), replied that it is not the mission of HCFA to support research, and that HCFA already has the mechanisms to pay for the normal hospitalization costs of patients in clinical trials. That HCFA takes this position, even though they could benefit from the findings of cost-effectiveness trials, was repeated in several of our interviews.

7We wish to thank Dr. John James of NIH for helping us obtain a magnetic tape copy of the 1979 Inventory data. Some of the figures below, which we obtained from that data tape, may differ somewhat from results published elsewhere.

8A design for a new NIH inventory of clinical trials was recently proposed. It would have contained almost all the information obtained by the 1975–1979 inventories except the single item of whether the clinical trial addressed new or standard interventions.
notion of a trial of a current medical practice. Therefore, we coded the title and description of purpose of each of these 234 trials as to whether it was a clinical trial of a current practice. We identified 47 trials in this way as “Current Practice” trials; these are the subject of our analysis.

Table 1 shows the mean total support for extramurally funded clinical trials, broken down for all trials and for trials of current practices and by whether the funding modality was a grant or a contract. Table 1 shows that only 4 percent of all extramural dollars spent by NIH on clinical trials went to the study of current practices. However, the means in Table 1 could be regarded as deceptive, as there is a wide range of funding of trials about each mean. Figure 1 illustrates this dispersion by showing the frequency of funding distributed across five intervals.

<table>
<thead>
<tr>
<th>Type of Support</th>
<th>All Extramural Trials</th>
<th></th>
<th></th>
<th>Current Practice Trials</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Non-</td>
<td>NCI</td>
<td>All</td>
<td>Non-</td>
<td>NCI</td>
</tr>
<tr>
<td></td>
<td>NIH</td>
<td>NCI</td>
<td></td>
<td>NIH</td>
<td>NCI</td>
<td></td>
</tr>
<tr>
<td>Contracts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n awards</td>
<td>198</td>
<td>87</td>
<td>111</td>
<td>10 (5%)</td>
<td>10 (11%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>mean $</td>
<td>2910</td>
<td>5986</td>
<td>500</td>
<td>1563 (3%)</td>
<td>1563 (3%)</td>
<td></td>
</tr>
<tr>
<td>Grants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n awards</td>
<td>576</td>
<td>174</td>
<td>402</td>
<td>29 (5%)</td>
<td>13 (7%)</td>
<td>16 (4%)</td>
</tr>
<tr>
<td>mean $</td>
<td>512</td>
<td>698</td>
<td>431</td>
<td>745 (7%)</td>
<td>1110 (12%)</td>
<td>449 (4%)</td>
</tr>
<tr>
<td>All support</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n awards</td>
<td>785</td>
<td>262</td>
<td>523</td>
<td>39 (5%)</td>
<td>23 (9%)</td>
<td>16 (3%)</td>
</tr>
<tr>
<td>mean $</td>
<td>1132</td>
<td>2453</td>
<td>470</td>
<td>955 (4%)</td>
<td>1307 (5%)</td>
<td>449 (3%)</td>
</tr>
</tbody>
</table>

*aRefers to the respective percent (of trials or of total funds) within funding source (all NIH, non-NCI, or NCI) devoted to trials of current practices.

The coding was performed by Terry Hammons and Bruce Hillman, both physicians, and by James Kahan, a psychologist. There was agreement in over 90 percent of the cases; the remainder were resolved by consensus.

Four clinical trials of current practices were intramural projects of the National Institute of Dental Research examining alternative means of fluoridation. We excluded these trials and four others that were mixed intramural and extramural work (they would probably be cooperative agreements if funded now), in order to obtain a purer estimate of the total support provided per trial. This exclusion reduced our sample size by only a small amount.
Fig. 1—Distribution of total support for 1979 NIH-sponsored clinical trials
The most important observation from Fig. 1 is the wide range of 1979 funding amounts, for both NCI and non-NCI contracts and grants. Over one-fourth of all non-NCI awards were under $100,000, and 14 percent were for more than $1 million. NCI awards, although generally at a lower level, still showed a wide dispersion. Almost one-third of all of the NCI's 1979 awards for clinical trials were under $100,000, while over one-fifth were for over $500,000. The pattern is repeated for the 43 current practice clinical trials. For non-NCI awards, the distribution was fairly uniform across all five funding intervals. NCI showed a preference for the $100,000 to $250,000 interval, but was not otherwise systematic.

We return to Table 1 to examine the role of grants vs. contracts. Overall, 198 out of 785, or 25 percent, of extramural clinical trial awards were contracts. These accounted for 65 percent of the total funds spent on clinical trials. Current practice trials were similar to all trials, with 26 percent (10 out of 39 awards) being funded by contracts. These contracts accounted for 42 percent of the total funds, a somewhat smaller proportion than for all trials. The 1979 support pattern for NCI differed considerably from that of the other Institutes; only 21 percent of their clinical trial awards were contracts, and none of the 16 trials of current practices were funded under the contract mechanism. For the other Institutes, one-third of all awards were contracts; for clinical trials of current practices, that figure rises to 43 percent.

Data from Malone (1985) indicate that the 1979 figures may be somewhat inaccurate with regard to present philosophy. As NCI moves to a pattern of using cooperative agreements and cancer control clinical trials, the Institute is taking a more proactive approach than would be suggested by its earlier heavy reliance on investigator-initiated grants. Moreover, the funding modality for the other Institutes is not uniform. NHLBI funds almost all of its large trials by contract, while the National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases (NIADDK) used the grant mechanism for over 90 percent of the trials it funded in 1984 (Malone, 1985). The other Institutes represent the full range between those extremes.

In sum, clinical trials represent a small fraction of the effort of NIH, and clinical trials of current practices represent in turn a small part of

---

11It is enticing but misleading to equate grants with investigator-initiated research and contracts with Institute-initiated research. The interaction between Institutes and investigators is complicated and cannot be easily reduced to the nature of the funding mechanism. As a result, there is no direct way to use extant data to determine the relative expenditures on investigator-initiated vs. Institute-initiated RCTs. See Sec. VII below.
that fraction. But there is no evidence that the funding patterns for clinical trials of current practices are different from the patterns for other clinical trials supported by NIH.
III. DECIDING TO INITIATE A CLINICAL TRIAL: FOUR DESCRIPTIVE MODELS

The literature on clinical trials does not present many discussions of how to decide whether to undertake a trial. Published reports of clinical trials are largely concerned with results and only touch on reasons for initiating a trial to justify the experimental method. Most articles discussing the process of performing clinical trials emphasize the logistics of performing the trial, particularly on the statistical design and the coordination problems among the multiple centers and multiple disciplines within each center. But these articles begin with our endpoint: namely, the decision to initiate a clinical trial. We present here four published descriptions of decisionmaking associated with whether to initiate a clinical trial. The degree to which the description as originally presented dealt with a decision to initiate a trial varies considerably; in no case was that decision the focus of the article.

A BRITISH CANCER CENTRE

The Cancer Research Campaign of Great Britain has funded a Cancer Research Campaign Clinical Trials Centre to coordinate the organization of large-scale prospective clinical trials in cancer (Scientific Advisory Committee, 1981). Clinical trials can be run from this central location in an efficient and cost-effective manner. In addition to providing space, the center also provides consulting expertise on managerial, data processing, medical, and secretarial matters to groups wishing to conduct large-scale clinical trials within Great Britain.

To use the Centre, one first makes a preliminary application to obtain approval of the scientific content of the proposed clinical trial. A scientific committee examines the scientific merit of the proposal, its methodological soundness, and its logistic and ethical feasibility. Promising studies that are weak on methods or logistic grounds may be referred back for revision and resubmission. But aside from this methodological feedback function, the Centre takes no role in the selection of studies to undertake;1 its task is rather one of coordination of clinical trials. Therefore, the process developed by the center for obtaining approval must be considered only a part of the clinical trials initiation process.

---

1However, approval by the Centre might well become a necessary imprimatur for support of major trials.
THE VA COOPERATIVE STUDIES PROGRAM

The VA has long been a locus of clinical trials in medicine, dating from the 1940s, when trials were performed of the effectiveness of chemotherapy in the treatment of tuberculosis. Gradually, support for clinical trials within the VA became centralized, then formalized with the establishment of the Cooperative Studies Program of the Medical Research Service. The functioning of this program is described by Hagans (1974), Henderson (1980), and James (1980). The program coordinates the VA multicenter clinical trials and acts as a gatekeeper deciding which studies are to be done. James (1980, p. 194) provides an overview of the system:

Studies are proposed by medical investigators within the VA system, reviewed by one or more discipline specialists, and by a triage committee consisting of five to seven high ranking officials in Research and Development and Professional Services in the Veterans Administration. Based on assigned priorities and funding availability, the approved proposals are released for planning and are assigned to one of the four Coordinating Centers.

Once at a Coordinating Center, the planning process begins. A study committee, consisting of a chair, potential participating investigators, and specialists (biostatistician, pharmacist, pathologist), is formed to develop the research proposal in full. When planning is completed, the research proposal is evaluated anew before funding, in a four-step process.

1. A review by the Coordinating Center Chief, assisted by a biostatistician and a human rights committee, who determine whether the proposal is ready for further review.
2. If the first step results in approval, submission for external written reviews by three referees.
3. Consideration by the Cooperative Studies Evaluation Committee (CSEC), who review the proposal for merit and how well it fits into the overall VA research program. They make a recommendation for funding priority.
4. Final review by the Director of the VA Medical Research Service, who decides whether the study will be funded.

The review by the CSEC is the most thorough one and addresses, among other elements, the importance of the study, the clarity and feasibility of its objectives, the logistic feasibility of the study, the correctness of the technical details, and the provisions for ethical safeguards. However, there are no formal guidelines for how this list of elements is to be weighted. Within this system, no particular
distinction is made between clinical trials of innovations and trials of current medical practices. Indeed, in our search of the literature on VA clinical trials as well as our interviews, this distinction was very seldom made.

The VA process, in sum, provides a systematic "by the numbers" procedure for determining whether a clinical trial will be funded. This process is bottom-up; that is, the responsibility of proposing a clinical trial lies entirely with the individual investigators. It is apparently assumed that VA investigators, as active participants in the system, will be aware of and responsive to the implicit priorities of the VA; no formal guidelines or explicit priorities for areas of research are provided from above. The investigators we interviewed recognized that priority is given to medical problems that are characteristic of the VA patient population.2

LARGE-SCALE TRIALS AT NHLBI

The NHLBI spends more money on clinical trials than any other Institute within the NIH except NCI. Most of these funds are spent on large-scale trials. Levy and Sondik (1982) outline the procedure involved in planning such trials.

NHLBI requires four separate phases of an RCT, with funding review occurring for each phase. These are initiation, planning, recruitment and intervention, and analysis and dissemination.3 Our principal interest is in the initiation, or first phase. A proposal may be initiated by discussions among the Institute staff, by researchers in the field, or as a consequence of previous clinical trials. In most cases, the idea for a clinical trial is initiated as part of the logical progression of an idea as it passes from basic science through clinical research to the point where large-scale testing is required to determine effectiveness and safety. If the proposal concerns a current medical technology, then the history of that technology is considered. The review at the initiation stage begins with two preliminary requirements. First, the proposed research must be appropriate to NHLBI: The design and

---

2 In our interviews, we noted that the VA had in the past identified areas for research and invited proposals but that "it hadn't worked very well." They are currently considering having the Advisory Board for the Cooperative Studies Program identify areas and move to an internal mechanism similar to NIH's Requests for Proposals (RFP). However, with the imminent retirement of the director of the program, no decisions have been made.

3 One might consider recruitment and intervention separate phases, as well as analysis and dissemination. The distinction made by NHLBI is with respect to funding; once approved for recruitment, intervention follows automatically, but a separate review of the intervention occurs before analysis and dissemination are approved.
management of the trial must require NHLBI research expertise, and the project must be one of actual efficacy and safety, not health service delivery. Second, the research must satisfy basic requirements of scientific merit and potential influence on medical care.

Levy and Sondik (1982) identify four key factors that are considered by NHLBI in deciding whether to pass a proposal beyond the initiation phase:

1. *The state of science.* Is medicine ready to receive the results of the clinical trial and use the conclusions? Here, there is a balance between complete knowledge and total ignorance. If we know a lot already, an expensive clinical trial might provide too little new information; if too little is known, then basic clinical research may be preferable to a clinical trial. For clinical trials of technologies in a state of flux, there is a risk that the design of the trial will have to be overhauled in the middle of the study as new findings invalidate the bases for the design of the trial.

2. *Feasibility.* Here, the trial design parameters, resources, desired results, number of centers, patient sample size, randomization process, stratification rules, etc. are all reviewed. The logistical feasibility of obtaining the objectives of the study are primary issues; in order to fund the project, it must have a good chance of success. At this stage, alternative means of obtaining the same information, including smaller clinical trials, data registries, or other means are considered.

3. *Impact.* The issue here is whether there will be an effect on health services delivery. Basically, if the outcome of a trial will not make a difference, then the incentive to do the trial is reduced. The analysis involves a cost-effectiveness component. Although it is important, it does not by itself determine whether the trial is to be sponsored.

4. *Ethics.* Finally, the ethical aspects of the proposal must pass muster. These involve the ethical aspects of the technologies themselves, the provisions for informed consent, and consideration of oversight of the intermediate results of the trial in case early data mandate changes in treatment protocols.

Levy and Sondik (1982) illustrate their process with two examples, one of an unfunded study of medication for mild hypertension and one for a study of beta blockers, which was performed. In the former case, the medical treatment was widespread and alternative means of testing the efficacy of the medication were available, so it was believed that the trial should not be funded on the basis of expected effect and ethics.

Zukel (1983) also illustrates the initiation process within NHLBI in his discussion of the evolution of the Coronary Drug Project, a long-
term, multicenter project involving several clinical trials exploring the efficacy and safety of cholesterol-lowering drugs. Although most of the article is concerned with how logistical problems such as clinic site selection and quality control were handled and with a chronology of which trials the project sponsored, there is a brief discussion of the initiation process. For the Coronary Drug Project, the original initiation point is well documented. In 1960, a committee met at the request of the National Heart Institute (predecessor to the NHLBI) with the specific mission of exploring the desirability of randomized clinical trials for these drugs. From that point, a mixture of political pressure from Congress and internal policies guided several clinical trials. Most of them were funded by the contract process, as NHLBI maintained more control on contracts than on grants; intramural studies were considered but rejected because they were too large to handle.

In sum, the NHLBI program, like the VA program, provides a systematic way of examining proposals, once the proposals have been presented. No specific search techniques are provided for determining practices needing RCTs, and no special provision is made for trials of current medical practices. Indeed, with the emphasis of the NHLBI on its research mission as opposed to health services delivery, trials of current practices would tend to evolve only when it was believed that the practice was not effective (e.g., Intermittent Positive Pressure Breathing Trial Control Group, 1983), and not for reasons of cost-effectiveness.

A MODEL OF SOCIAL POLICY AND TECHNOLOGY ASSESSMENT

The final process we consider here is not one of decisionmaking with regard to clinical trials, but rather a more general process of the decision to implement a social policy to change the current state of a technology. If the state of technology is the use of a current medical practice and the social policy is the decision to initiate a clinical trial of that practice, then this model is an appropriate one to review. This model will form the basis of the one we propose in the next section.

Banta and Behney (1981) describe a model of the process of social policy as a loop between the current state of a technology and policy manipulations designed to change that current state, with technology assessment as an empirical means of measuring both the current state

---

4 NHLBI, in common with other Institutes, does have various mechanisms for obtaining the advice of the research community. For example, McGlynn and Kahan (1985) document how a special conference on lung disease led directly to the IPPB trial.
of technology and the possible usefulness of a proposed manipulation. Figure 2 displays a much simplified version of their model.

In this model, one begins with a need to determine the current state of a technology. If there is uncertainty as to the current state, then a clinical trial may be an appropriate means of reducing the uncertainty. Then, the current state is compared with a desired state of technology. The differences between the desired and current states result in the

---

5Although it is tempting to try to deal with the problem of determining what the desired state of the technology should be, we leave that to a black box labeled "politics" for the present.
formulation of proposed manipulations to reduce that difference. As the variety of such manipulations is typically wide, and as their effect is unknown, again there is a need for an empirical examination, this time of the effects of the manipulation. Here again, for medical technology, clinical trials are an important investigative tool. Finally, based on the information about the proposed manipulations, a policy is formulated and implemented. This results in a need once again to assess the current state of technology to see if any changes have occurred, and the process recycles.
IV. A DECISIONMAKING MODEL: OVERVIEW

Figure 3 presents a model of the general process that we believe governs the decision whether to initiate a clinical trial of a current medical practice. We view that process as taking place in three sequential stages: awareness, relevance, and feasibility. Within each stage, information is integrated into a decision to proceed further or to cease considering the topic. A current medical practice must successfully pass through each stage to become the object of a clinical trial.

This model is related to that by Banta and Behney (1981) portrayed in Fig. 2; our “awareness” has much in common with their “determination of the current state of technology by empirical examination” and our “feasibility” is their “determine feasibility by empirical

Fig. 3—The decision to initiate a clinical trial of a current practice
examination.” Our “relevance” is an intermediate stage that falls between their two. We have also explicitly looked inside each stage, at the different factors upon which awareness, relevance, and feasibility will be evaluated.

THE THREE STAGES OF THE DECISIONMAKING MODEL

Awareness

With awareness that a current medical practice is perhaps not used appropriately, that practice becomes a candidate for research attention, possibly in the form of a clinical trial. We become aware that current medical practices are questionable if there is no consensus on when they should be used. That might be because of a general recognition that the appropriate indications for the practice are unknown or because there are differences of opinion within the medical community. There are several ways in which we can become aware that a current practice is questionable; these sources of dissensus are portrayed as the components of awareness in Fig. 3.

Of the three stages, awareness is unique to current practices; the medical research community is aware of innovations by virtue of their newness. This stage is therefore the most critical one for an examination of the decision to initiate RCTs of current practices.

Relevance

The inherent imperfection of knowledge in medicine means that there will be awareness of doubt about far more practices than can be investigated in RCTs. Additionally, a clinical trial will not be appropriate for some practices. The process of winnowing the many questionable medical practices down to a set that are candidates for clinical trials is based on a judgment of the relevance of an RCT to test the practice. This second stage involves practical, subjective, political, and methods considerations. If a current practice is deemed relevant, then an investigator considers applying for a grant to conduct the clinical trial or an NIH Institute considers issuing a contract for the trial.

Feasibility

In the third stage, the feasibility of the proposed trial is assessed. Even if an RCT is an appropriate strategy in the abstract, real-world constraints can make it impossible to initiate. In the other two stages, tradeoffs are possible among the dimensions, but a proposed RCT must pass muster on all the dimensions of feasibility to be initiated.
HOW WELL DOES THE MODEL DESCRIBE CURRENT DECISIONMAKING?

This model is qualitative, not quantitative. We believe it is not now possible to specify weights among the various dimensions affecting each stage. Some factors might be important because of the inherent nature of organizations and some might be important because of deliberate policy. The value of the model is in identifying the nature and components of stages and the factors that determine whether a stage will be completed.

As we stated, this model is a general picture of what we believe are the stages of the process underlying a decision to initiate a clinical trial. The model is prescriptive in the sense that we believe investigators and NIH should explicitly and systematically consider the three stages. As we stated in our discussion of the life cycle of medical technologies and error in medicine, there is dissatisfaction with the means of searching for current practices. The model is also to a lesser extent a descriptive model of how NIH makes initiation decisions. Although actual practice generally follows the model, it is not nearly as systematic as the clean lines of Fig. 3 might imply. As we discovered in our interviews and case studies, the processing within each stage can be informal and sometimes even haphazard; this means that the decision-making process is subject to possible inefficiency or bias.

The description is inaccurate to the extent that certain of the factors that could lead to awareness, relevance, or feasibility are not used by an individual Institute. Often this nonuse is deliberate policy on the part of the Institute, but sometimes, as our interviews revealed, the factor is inadvertently ignored. The importance of each factor varies greatly from case to case, and it is often correct to focus on one to the exclusion of others. But when a factor is ignored, current practices that should be questioned according to that criterion will not be considered for a clinical trial, and medical practice in the United States may be that much worse off.
V. THE AWARENESS STAGE

The first step in the process leading to a clinical trial of a current medical practice is the awareness that the practice is suspect. In other words, the need for an RCT must come to the attention of the medical community before such a trial can be initiated. This seemingly tautological point may be overlooked; after all, for the new medical practices that are the topics of the bulk of RCTs, this step is virtually automatic. Current practices, however, may be used without consideration of their appropriateness; if the practice is never questioned, it will never be subjected to clinical trials. We consider several ways in which a current medical practice comes to be questioned by the medical research community. The sources of such questions include the efficacy and safety of the practice, the biomedical theory that explains why the practice works, its cost-effectiveness, the quality of life of patients subjected to the practice, and epidemiological evidence of its use.

ISSUES OF EFFICACY AND SAFETY

Efficacy and safety are the traditional reasons for undertaking a clinical trial. Basically, the question of efficacy asks whether a practice does what it is purported to do, and the question of safety asks whether anything untoward results from the use of the practice. Neither efficacy nor safety is absolute; both are relative to the alternatives available to the physician. As one of our respondents put it, the basic reason for doing a clinical trial is to satisfy the doctor's desire, "I wish I knew how to treat my patients." To the extent of one's faith that physicians will recognize valid scientific evidence when they see it or that medical consumers will attend to scientific evidence, the clinical trial is one of the best ways of fulfilling those desires.

The questions of efficacy and safety are the basic ones of health care technology assessment. Technology assessment is a necessary aspect of medical care, and clinical trials provide a major means of performing such assessments (Banta and Behney, 1981; Office of Technology Assessment, 1978, 1982, 1983). A survey of the Health Technology Assessment Reports (e.g. National Center for Health Services

1Another respondent raised the corresponding and equally valid desire on the part of patients, "I wish I knew why I was doing what the doctor told me to do."
Research, 1983), NIH Consensus Development Conference statements (Office for Medical Applications of Research, 1980), assessments by the Office of Technology Assessment (1978), and reports of the Clinical Efficacy Assessment Project of the American College of Physicians (1982, 1984) demonstrate that clinical trials, when available, are a major component of technology assessment; when clinical trials have not been done, these reports often call for their initiation. Phase III of the testing for new drugs, performed after the drug is shown not dangerous in an absolute sense and has shown promise on a limited scale, is an RCT for the efficacy and safety of the drug in the population of potential users.

The history of clinical trials that ferret out inefficacious treatments, as related in part earlier (e.g., Barnes, 1977; Barsamian, 1977; Chalmers, 1974; Miao, 1977; Robin, 1984) is not small. The literature contains numerous examples. Gotttrup et al. (1985) examined different current practices of preoperative bowel preparation before colorectal surgery in a clinical trial. Their rationale for the trial was clearly based on questions of the efficacy of treatment; among the various alternatives, nobody knew which ones were most effective in preventing postoperative infections. Other clinical trials of treatments in large part instigated by questions of efficacy and safety were brought out in our interviews. These include a test of plasmapheresis for Guillain-Barré syndrome, radial keratotomy for myopia, photocoagulation treatments of various sorts for detached retinas, dietary interventions for pre-end-stage renal disease, early fetal monitoring vs. auscultation, the Multiple Risk Factor Intervention Trial (MRFIT), the aspirin and beta-carotene trials, and many others.

There is also a large list of treatments whose efficacy and safety have been questioned, but for which there has not yet been a clinical trial. Barnett, Plum, and Walton (1984) have expressed a common concern among experts in the field of stroke that the procedure of carotid endarterectomy, one of the most frequent surgical procedures in the United States, may not be an effective means of preventing stroke. They call for a registry study to ascertain what the outcomes of the procedure are and for an RCT of surgery vs. nonsurgery for certain classes of patients to test the efficacy of the procedure. The questioning of carotid endarterectomy also arose in our case study of EC/IC and in some of our interviews.2 Although the procedure was questioned

2It should be noted that Barnett was the PI for the EC/IC trial.
because its efficacy and safety were not established and although its investigation is clearly relevant, it has not yet been subjected to a clinical trial because of feasibility constraints (see below).

The range of treatments that could be added to carotid endarterectomy is at least as great as the range of treatments already subjected to RCTs; among the suggested treatments mentioned in our interviews were a test of DPT vaccine, treatment for multiple sclerosis, enzyme treatment for slipped disks, digitalis for congestive heart failure, and repeat cesarean sections.

Efficacy and safety were important in questioning each one of the treatments considered for trial. The questioned efficacy of the treatment was the primary driving force bringing the EC/IC and IPPB trials into being. Even though the proposed CCU trial was not funded, the reviews of the proposal recognized the validity of the efficacy issue. Although the proponents of the breast cancer trial claimed that theoretical issues were the primary motivation for its initiation, the issues of safety and efficacy were prominently featured in the discussions surrounding the decision to conduct it.

The major reservation we detected about questions of efficacy and safety leading to awareness is that there is no systematic plan to assess them. Here, more than anywhere else, discovery of the need for an RCT to test a practice is a haphazard process. Most of our respondents shared our ignorance when we asked how the medical profession comes to know when it doesn’t know about the efficacy and safety of a practice. In the concluding section, we shall consider some systematic approaches to awareness that could ameliorate this perceived problem.

**BIOMEDICAL THEORY**

Biomedical research is firmly in place as the primary mission of NIH (Brandt, 1985). NIH is the main communication channel for biomedical research and its primary funding source. Many NIH officials perceive that Congress and the administration want a greater emphasis on basic research at the expense of applied clinical research.

It is not surprising, then, that biomedical theory is a major source of the awareness of the possible inadequacy of current medical practices.³

³By "biomedical theory" we mean the part of science that is concerned with how disease occurs and progresses, and why treatments work to retard or eliminate disease. The point here is that the scientific knowledge per se is addressed, not applications of that knowledge.
For example, Capocaccia (1984) called for clinical trials of the standard treatments for hepatic encephalopathy. Most of his reasons deal with the science of treatment. He believes that there is ambiguity about what constitutes the “standard” treatment, and trials are needed to "standardize" its administration. He also believes that the mechanism by which standard treatment (in any of its manifestations) works is unknown; clinical trials may help reveal that mechanism.

Our interviews uniformly supported the proposition that biomedical theory is a major source of awareness. Several respondents noted that for their fields, effective treatments simply do not exist, and the trials that are conducted are tests of new biomedical theory suggesting new treatments that should work. Some respondents held the opinion that too many clinical trials were now done, and that they should be reserved for conceptual breakthroughs. There was consensus that a research appeal other than pure patient management was needed within the research community to question a practice. The argument that a treatment required testing also required theoretical justification.

There are critics of the theoretical emphasis of NIH. The harshest position we heard is that NIH is little more than a conduit of funds to subsidize medical schools and research institutes, and that funding is guided by the professional aspirations of investigators rather than by societal needs. Although this position is certainly overstated, it is true that many NIH officials take the position that research is the investigator's business and that the scientific community, through peer reviews of grant applications and publications, will be a "modestly visible hand" guiding biomedical research in societally optimal directions.

It is instructive to close this discussion of the role of biomedical theory with a brief review of our case studies. As detailed in our case study of breast cancer surgery as well as the published reports of those trials (Fisher et al., 1985a, 1985b), the questioning of scientific models of how cancer spreads led to awareness that the Halsted radical mastectomy might not be the most appropriate treatment for the disease. Also, the case study of IPPB demonstrated that basic biomedical theoretical considerations led to questioning of the treatment. And in the case study of CCUs, the lack of a cogent challenge to the biomedical theoretical basis for intensive care of noncomplicated acute myocardial infarction may well have been a major factor in denying funding to the proposal. However, the remaining case study does not fit this picture. Biomedical theory appears to have taken second place to questions of efficacy of treatment in the EC/IC trial. Based on this small survey, then, the exclusive dedication of NIH to biomedical theory that is claimed by some of its critics is an overstatement of actual practice.
ISSUES OF COST-EFFECTIVE HEALTH SERVICES DELIVERY

To state that the continuing escalation in the cost of medical care is a major problem for American society is to state the obvious. Cures for this particular ill are sought from any possible source. It is not surprising, therefore, that some observers of the U.S. health care system (Bunker et al., 1977; Chalmers, 1984; Gordon, 1985) have proposed that clinical trials be utilized to help identify the most efficient means of providing health care to the American population. In the current jargon, they propose that clinical trials should explore issues of cost-effectiveness.

It is fairly simple to define in a broad way what is meant by cost-effectiveness: A medical practice is cost-effective if its benefits are sufficiently great to justify its costs in some sense and if these benefits cannot be gained through other, less costly means. Few would argue with the general proposition that medical care, particularly that financed by the public purse, should be as cost-effective as possible. No one would support the use of very expensive medical practices to achieve only marginal gains. But when we attempt to apply the concept of cost-effectiveness, it becomes almost impossibly slippery. How, after all, are life, health, and freedom from pain to be quantified in terms that can be compared to costs measured in dollars? Whose benefits (the patient’s, his family’s, society’s) are to be weighed against whose costs? No grand balance of clearly articulated and accurately calculated costs and benefits will ever be possible for any medical procedure. Proponents of using clinical trials, however, generally have much more modest goals—goals that can in many cases be achieved.

Because full balances of benefits and costs cannot be computed does not mean that even the quite narrowly defined monetary costs of a medical procedure are irrelevant to social and private choices about the use of that procedure. Before we choose a practice for ourselves or allow it to be paid for publicly, we want to know what it will cost.

But it is generally not straightforward to compute the monetary consequences of a particular medical practice. Costs may vary from patient to patient depending on whether complications arise, how long the patient remains in the hospital, what sorts of follow-up care are necessary, whether the patient suffers a relapse, and the skill and circumstances of the health care provider. In this regard, monetary outcomes are quite similar to medical outcomes. As with medical consequences, in some circumstances the only valid way to observe

---

4Indeed, the current concern over public outlays has given such narrowly defined costs, if anything, too much prominence.
differences in the monetary consequences of two practices is to conduct a clinical trial, observing costs as closely as medical outcomes are traditionally observed. The result of such a trial may not be a clear determination that one practice is to be used in preference to another; indeed, purely “medical” trials often result in no such conclusions. This sort of trial may, however, contribute to a better understanding of the true monetary costs of a procedure, just as a more traditional trial will contribute to a better understanding of its medical efficacy. For many proponents of the expanded use of clinical trials, then, an examination of cost-effectiveness means only that the design of a clinical trial should include careful observation of the costs of the treatments provided.

There is another sense in which clinical trials can be useful in investigating the cost-effectiveness of particular medical practices. Sometimes there is little question about the relative costs of two practices. Outpatient treatment of some complaints, for example, is clearly less costly than inpatient treatment. But is outpatient care as good as inpatient care, or are the advantages of inpatient care so great as to justify the extra cost? A trial designed to address this question would primarily examine the medical consequences of the two forms of treatment. If these consequences were found to be roughly similar for the two modes of treatment, there would be a strong reason to prefer the less expensive.

Clearly, RCTs are not appropriate for exploring all interesting questions about the cost-effectiveness of particular medical practices. However, clinical trials can be useful in shedding light on some highly relevant aspects of the costs of particular practices or on the additional benefit that derives from a more expensive course of treatment. It is only natural, then, that cost-effectiveness questions may sometimes be seen as an appropriate motivation for a clinical trial. Indeed, some authors have argued that suspecting a particular current practice is not cost-effective may be as valid a reason for a clinical trial as would be suspecting the same practice is ineffective or unsafe.

Fredrickson (1980), for example, noted that the modern doctor’s bag has become a compendium of expensive gadgetry, whose usefulness and cost-effectiveness is unknown. Clinical trials, he felt, might reduce the size of this bag to a leaner, more efficient kit.

Tygstrup et al. (1984) discuss some of the difficulties in designing clinical trials for standard practices within the narrow focus of acute hepatic encephalopathy. Their decision flow model, shown in Fig. 4, explicitly considers cost-effectiveness as a criterion for initiating a
clinical trial of a standard practice and shows the subordinate relationship of cost-effectiveness to efficacy of treatment.

Evans (1983) sees a need to consider, in addition to the efficacy of a medical technology, whether that technology is necessary, the relevant constraints on its availability, its cost-effectiveness, issues of legal risk, and the ethics of its administration. Each of these characteristics should be considered before a medical technology may justly be worth paying for in an environment of limited health care resources. Clinical trials are a means of learning about all of these characteristics of a technology.

McKinlay (1979, 1981) outlined three criteria for a medical practice to be eligible for public funding:

1. It should be of demonstrated effectiveness. That is, it should be better than a placebo or no treatment and should be safe.
2. It should be cost-effective. When two or more technologies of approximately equivalent effectiveness are available, the cheaper one should be preferred.
3. It should be acceptable to the public and equitably distributed.
McKinlay views clinical trials, especially ones of practices already in use, as particularly helpful in establishing the first two criteria. Indeed, he argues that when the efficacy and cost-effectiveness of a proposed technology have not been demonstrated, a clinical trial is mandated.

Drummond and Stoddart (1984) point out that consideration should be given to the economic consequences of a new technology and that therefore a funding agency might decide to fund or not to fund on the basis of the quality of the economic analysis proposed as part of a clinical trial.

Williams, Reading, and Ward (1983, pp. 499–500) take a similar point of view:

The purpose of the trial must be carefully defined and consideration given to the need and worthiness of the treatment, cost effectiveness, risks, and ultimate acceptance by physicians and patients. . . . It is necessary to consider whether the benefit anticipated with such a therapy will be appropriate for the expense incurred and whether the trial results might be misinterpreted or the treatment misused in the usual and regular management of patients.

Gordon (1985) sees a need for clinical trials that begin with competing technologies believed of equivalent effectiveness to test which of them is more cost-effective. Such trials are viewed as a potentially major aid in containing the costs of medical care. The problem, according to Gordon, is that although the need for such trials is readily acknowledged in the abstract, agencies are reluctant to sponsor them, preferring instead to allocate their own research funds to problems of more "scientific" effect, which attract more interest and more publicity, and hold more potential rewards for the investigators. The result is a dilemma; the need for such trials is acknowledged, but everybody prefers that somebody else do the work.

We might add to Gordon's comments that more than the attitudes of sponsoring agencies may lie behind the small number of clinical trials that examine cost-effectiveness issues. Clinical trials are generally proposed and designed by medical researchers, who have not (fortunately so, in some views) seen economic issues as their central concerns and do not immediately include the costs of care among the important attributes of a practice to be tested in a clinical trial. Those for whom a consideration of costs is second nature (economists, say) are seldom in a position to propose a clinical trial.

The few trials that have successfully addressed cost issues demonstrate that such trials are possible. Pineault et al. (1985) conducted an RCT expressly designed to look at costs. They compared doctor costs,
hospital costs, and patient out-of-pocket costs for three surgical procedures—tubal ligation, hernia repair, and meniscectomy. For each of these procedures, the patient can be hospitalized or the surgery can be performed on an outpatient basis. They found that for the first two procedures, outpatient surgery reduced costs, but that for the third, a brief hospital stay was less expensive. Patient preferences were generally in favor of hospitalization, but this may be in part because the study was conducted in Québec, where patients bear no part of the hospitalization costs. This study demonstrates both that RCTs can be used to evaluate cost-effectiveness and that the results of such trials will not necessarily be in favor of the lesser treatment.

Most of the NIH officials we interviewed noted that cost-effectiveness was only beginning to be considered in the deliberations that lead to decisions about supporting particular clinical trials. Most said that the opportunity to look into questions of cost-effectiveness would add only marginally to the attractiveness of a proposed clinical trial but that even the strongest cost-effectiveness question was unlikely ever to lead to support for a trial that was not fully justified on the basis of its direct contribution to medical knowledge. The impression was that a test of two or more medically well understood procedures to determine whether one of them generally involved lower costs would stand little chance of being supported.

Although it is possible to identify NIH-supported trials that have shed light on cost-effectiveness questions, we have identified no case, nor did any of our respondents identify one, in which cost-effectiveness had been the major explicit motivation for supporting a clinical trial. When cost-effectiveness has become emphasized in a clinical trial, it has generally been because influences outside the normal decision-making process have been at work. A consideration of the cost implications of different diets in the treatment of end-stage renal disease, for example, was undertaken only because Congress mandated it. Not surprisingly, no Institute within NIH systematically seeks to identify current medical practices of questionable cost-effectiveness.

All of our respondents insisted that proposals suggesting explorations of cost-effectiveness questions were not at a disadvantage in the NIH decision process, even if that exploration conferred no advantage. Our case study of CCU does not support that, however. The CCU proposal may have been hurt in part because it did include an explicit investigation of the relative costs of two treatment modalities; perhaps those charged with reviewing the proposal did not fully understand the proposed investigation.
QUALITY OF LIFE

As medicine progresses and chips away at mortality, quality of life measures must become more and more important (Mosteller, Gilbert, and McPeek, 1983). This comment is particularly cogent for studying current practices. The mammary artery ligation clinical trial (Barsamian, 1977) and gastric freezing trials (Miao, 1977) both used quality-of-life measures as their major dependent variables, and the major finding of the clinical trials was a demonstration of a placebo effect. Wenger et al. (1983) note that when differences in mortality are not large, knowing the quality of life is important in telling a patient what to do.

The NIH officials we interviewed recognized that quality of life issues were important in motivating clinical trials. Some cited quality of life issues as having been paramount in bringing about particular trials. Perhaps the clearest example of the importance of these issues has been the strong push to study alternatives to total mastectomy as the standard treatment for breast cancer (Hillman and Kahan, 1985). No one has argued that less drastic surgical procedures are likely to reduce mortality; the attractiveness of these alternative procedures was because they resulted in less disfigurement and less psychological stress for breast cancer victims while not increasing mortality. These alternatives would clearly improve the quality of life for breast cancer patients; the question for the clinical trial was whether this improved quality of life had to be paid for with higher mortality rates.

Even if the quality of life is an important consideration in questioning a medical practice, the practice may not be considered for investigation because of a belief that quality of life is not a measurable outcome for a clinical trial. But Meenan et al. (1984) examined quality of life measures in a clinical trial of treatments for rheumatoid arthritis. In addition to the clinical and health status measures of physical functioning, they examined subjective measures of patient well-being and found that these performed comparably to the more traditional outcomes. They conclude that quality of life measures, which are typically less intrusive than other outcome measures, can serve a useful function in evaluating the outcomes of clinical trials.

Although the example of breast cancer surgery demonstrates that quality of life is a potential contributor to the motivation to initiate clinical trials, we found that quality of life is actually only rarely the issue of an RCT. This distinction, which arose in several of our interviews, is a subtle one. In general, there is no question about which treatment provides a superior quality of life. Instead, the superior quality of life of a particular alternative treatment motivates a clinical
trial to establish the equivalence in medical effectiveness of that alternative and the standard treatment. Quality of life can be a source of awareness for clinical trials of current medical practices but is rarely the question that a clinical trial addresses.

EPIDEMIOLOGICAL EVIDENCE

It is generally considered unethical to withhold standard treatment from patients with a given condition when this treatment is thought to be efficacious. Although this view has definite moral appeal, it has the practical disadvantage of restricting knowledge about what might have happened if the standard treatment had been withheld. This in turn restricts our ability to judge the effectiveness of the standard treatment.

But experiments that cannot be ethically performed by design are sometimes performed by accident. Sometimes standard treatments are withheld or used with less frequency in one area or at one time than they are in other circumstances. Perhaps this is because of a temporary incapacity to offer the standard treatment or because of a different prevailing view of what constitutes appropriate medical care. For example, a temporary shortage of intensive care unit (ICU) beds at a hospital can lead to a rationing of beds such that some patients who would normally be in the ICU are in standard beds. The outcomes of these patients can be compared with outcomes of those who, at other times, were in the ICU. For whatever reason, we are sometimes confronted with natural experiments that suggest that the prevailing view of appropriate treatment may not be correct or at the very least may not be generally held.

Wennberg and his colleagues (e.g., McPherson et al., 1982; Wennberg, 1984; Wennberg and Gittlesohn, 1982) have studied such accidental experiments in depth. These investigators have noted markedly different rates of incidence of particular surgical procedures among New England towns and in other locations with at least superficially similar populations. In this research, no attempt is made to assess whether the populations of these towns were better or worse off because of the different incidences of surgery they experienced. Rather, the data are used to conclude that what was considered accepted and standard medical practice in one town was quite different from what was similarly considered in another town. Whatever effect these differing views had on the general health status of the population, the differences were apparently not great enough to force a convergence of medical practices, suggesting that the surgical procedures
in question might be fruitfully subjected to more formal investigation. Moreover, the apparent disagreement over what constitutes appropriate care suggests that withholding a treatment in some cases may be ethically supportable. In essence, these observations provide a source of awareness that a clinical trial might be appropriate to determine the efficacy or effectiveness of these surgical procedures (Brook et al., 1984).

Schacht and Pemberton (1985) examined the incidence rates of several surgical procedures in Australia, the United Kingdom, and the United States. They found utilization differences among the three countries, even after correcting for the sex and age of patients. They offer the hypothesis that cultural differences and consumer preferences, rather than systematic over- or underutilization of the procedures, explain these differences. Even so, they see a need to control practitioners who, on the evidence, may be abusing the system by over-operating. They offer no assessment of how patients in these three countries have fared after experiencing different incidences of surgery, but certainly one must wonder why it is that surgeons in one setting operate more often than those in another. This too suggests a need for more formal study.

These accidental experiments are of course far from ideal. Because they arise by accident, there is no guarantee that the patients assigned to the different arms of the trial are at all comparable. Neither, in general, is there any systematic reporting of patient outcomes. For all this, they can provide useful clues, suggesting which current medical practices might fruitfully be subjected to more formal scrutiny.

All of our respondents recognized such accidental experiments as important means by which procedures that might require a second look were brought to the attention of the medical community. We heard differing views, however, on the value of searching systematically for such experiments. Some expressed faith that persistent differences in practice patterns or patient outcomes would eventually become known to the larger medical community and that appropriate further studies would eventually be undertaken. Others cited results similar to Wennberg’s research group as indications that such differences could be quite large and still not recognized.

This is not to say that such information is unavailable. The National Center for Health Statistics (NCHS) has several databases that could be useful in this regard. At NCI, the SEER database provides information on the incidence of and treatment for tumors of various types in sample parts of the country. In addition, the newly initiated PDQ system, whereby NCI provides information about treatment protocols to inquiring physicians, could provide information
about which treatments are most queried. Some minor modifications in PDQ might provide an epidemiological database for assessing doubt about the efficacy of cancer treatments. We are not aware of any comparable databases for generally widespread treatments among the other Institutes of NIH; that does not mean that they do not exist or that they could not be implemented.

In general, epidemiological evidence is an underutilized source of awareness of which current practices might be fruitfully tested by clinical trial. Nowhere, however, did we find any systematic effort to collect or to analyze raw "epidemiological" data on the incidence of various medical procedures or of medical outcomes to identify candidates for RCTs. Several respondents recognized the potential value of such efforts and decried the scarcity of registries and other low-cost means of accumulating such data. Although these respondents believed that more could be done in this regard, they saw no agency obviously suited to performing this function.
VI. THE RELEVANCE STAGE

When is an RCT an appropriate test of a technology? Louis and Shapiro (1983) echo a consensus response that clinical trials should be initiated only after due consideration of cost, complexity, and alternative approaches. The frequency and severity of the disease treated by the proposed technology and the expectation that the study will be interesting and feasible are also important. But to date, neither Louis and Shapiro nor anyone else has attempted to systematically list these decision criteria. The following dimensions provide a first approximation to such a list: the impact of the disease on society, the political environment surrounding both the disease and the practice in question, whether a window of opportunity to examine the practice is open, the anticipated impact of an RCT on clinical practice, and the alternative methodologies available to test the practice.

IMPACT OF THE DISEASE

Given the financial, organizational, and sometimes ethical costs of a major clinical trial, it is not surprising that considerable attention is directed to whether the disease to be addressed is of sufficient effect, either on the afflicted individual or on society, to justify the effort required. All of our respondents stressed the need for some assessment of the impact of a disease and the treatments for that disease before a trial is undertaken to explore them.

There was general agreement among our respondents about the dimensions in which the impact of a disease should be measured. Clearly, diseases that strike many people or procedures that are widely used are better candidates for clinical trials than are rare diseases or seldom used procedures. This consideration goes a long way toward explaining the decisions to carry out trials on treatments for breast cancer or heart disease. The severity of the illness in question also contributes to its impact and hence the appropriateness of research on it. Disease processes that bring death or severe disability are universally seen as having a stronger claim on clinical trial resources than do disease processes that result in less severe disability or discomfort. Clinical trials sponsored by the National Eye Institute (NEI), for example, are almost all aimed at disease processes that can result in blindness. These diseases are certainly less common than other ail-
ments that impair vision to some extent, but the disability and anguish associated with blindness are such as to give efforts to prevent blindness priority over other possible areas for research.

Our respondents also identified the amount spent on a particular procedure or on treating a particular disease as a legitimate measure of the impact. Several noted that if a procedure apparently causes no harm, does not prevent the patient from obtaining effective treatment, and does not result in large financial outlays either by individual patients or by society as a whole, then there is little reason to test its effectiveness in a clinical trial. Even if it is ineffective, little harm is caused by its use.

These factors—frequency of incidence, severity of illness, and cost of treatment—are often cited (e.g. McKinlay, 1979, 1981) as prescriptive determinants of governmental priorities in reimbursement for medical care. Because RCTs are important in defining the efficacy of treatments, they should be allocated with the same priorities. It was clear that the impact of the disease was considered in all of our four case study trials.

Although all of our respondents could easily identify the factors that made a particular condition or procedure important in their minds, none had made any effort to devise a method for combining these factors to arrive at a systematic assessment of a disease's effect. Moreover, the present procedures were believed to be appropriate. Every case has special features, and attempts at establishing a general scheme for assessing the impact of all possible diseases and practices to be studied were viewed as impractical.

There was no evidence that candidate practices for RCTs were ranked on their importance. Rather, there appeared to be a consensually defined boundary of importance for medical practices. Investigators and NIH officials to a large degree have similar views of which issues are sufficiently important to merit consideration in a clinical trial, and once that criterion of "important enough" was met, candidate practices were not further evaluated on this factor. We were told that it is rare for NIH to turn down a proposed clinical trial because the disease does not have a sufficient impact. Indeed, several NIH officials and clinical trial investigators noted that limitations on the number of clinical trials proposed arose not from any shortage of potential diseases and practices to be studied, but from the small number of investigators and institutions capable of mounting a methodologically sound trial and by the scarce financial resources available for such trials.
POLITICAL ENVIRONMENT

Political influences on medical policy are inevitable. Congressional and other pressures, media attention, and controversiality can all be factors in decisions to initiate a research project (Laumann, Knoke, and Kim, 1985). If a special interest group wants to have a particular trial initiated, it can bring to bear all of these forces. Our interviews established that NIH administrators are acutely aware of political pressures and attempt to balance them against pressures from their medical research constituencies.

NIH, as a federal agency, must respond to the influence of the U.S. Congress. Nine of the 11 NIH Institutes operate under permanent authority derived from section 301 of the Public Health Service Act. NCI and NHLBI are separately funded, with sunset provisions. Some members of Congress want all of NIH to require periodic reauthorization, and others want NHLBI and NCI returned to the status of the other Institutes (Culliton, 1985). For some Institutes, Congress has appropriated special funds to be dedicated to clinical trials. The strict dependence of NIH on Congress for funding is well-appreciated by the Institutes; they are aware of the need to justify the social relevance of their research programs, in particular such expensive projects as multi-center clinical trials.

Political influence is rarely explicitly mentioned in the literature as a motivation for an RCT, but it was readily acknowledged in our interviews. On occasion, Congress has mandated a particular clinical trial. Although the inevitability of bowing to such pressure is recognized (e.g., Brandt, 1985), we heard some desire that Congress allocate supplemental monies for such trials. Our interviews revealed that pressure from disease-oriented lobby groups has a considerable influence in determining the relevance of particular practices. The role of the women's movement in creating an environment favorable for the trial of breast cancer surgery is well-known. Public pressure was responsible for the trials testing the efficacy of vitamin C and Laetrile as cancer treatments.

Political issues surrounding clinical trials can extend to the courts. Norman (1985) reported on a lawsuit by practitioners who claimed that NEI and the investigators were attempting to monopolize the surgical procedure of radial keratotomy by conducting a clinical trial of its safety and efficacy. Even though no in-depth scientific peer-reviewed study of the procedure had been published, the plaintiffs claimed that the treatment was not experimental, and that they were being prevented from receiving reimbursement from third parties because of
the clinical trial. Although the courts ruled that NEI was not liable, the remaining defendants were faced with a trial and have attempted to settle out of court. Norman reports that if clinical trials of treatments become the target of further litigation, there could be an unwillingness by physicians to participate in such investigations, to the detriment of medicine and society in general.

Political influence can include interagency conflicts over who should fund an RCT. For many potential clinical trials, particularly ones with a strong health services delivery or cost-effectiveness component, agencies may agree in the abstract that the trial should be performed but disagree about which agency's mission is best addressed by the trial, hence who should pay for it. The effect of such political wrangling (which is a vital matter for the wranglers) may cause some potentially important trials to fall through the cracks because no single agency finds it relevant to its particular mission. We return to this issue in the concluding section.

Many of our respondents expressed some mild annoyance at the occasional untimely intrusion of political factors into the process of deciding which trials to conduct, but some noted that political influences were occasionally beneficial. One respondent observed that political pressures can sometimes force the initiation of a clinical trial of a current medical practice that otherwise would not be tested because there were no incentives for anybody to initiate it. He illustrated his point with a discussion of FDA drug approval policies. Because FDA does not conduct trials itself but only approves drug company trials, and because the drug companies tend to conduct only those trials necessary to get their products to market, there is no incentive to test either new uses for already approved drugs or promising pharmacological developments that do not appear profitable to the drug companies. Another respondent opined that in the United States, research resources tend to be used unwisely, often supporting the interests of individual investigators and institutions rather than broader societal interests. Political influence can sometimes act as a counterweight to this tendency.

Among our respondents, then, was some consensus that a political element in the determination of relevance was not entirely a bad thing. Clinical trials, because of their expense and magnitude, are necessarily large policy decisions, and there was some expression that when large

---

1Reimbursement here refers to about $1000 per eye for a procedure that lasts only a couple of minutes.

2The Orphan Drug Act is a political intrusion into medicine that attempts to ameliorate this very problem.
questions arise, major policy decisionmaking bodies such as the Congress should have an influence.

WINDOW OF OPPORTUNITY

There was a consensus among the people we interviewed that a clinical trial can be initiated only within a certain “window of opportunity,” which was typically only vaguely defined in our interviews; yet the respondents were firm in their belief of its existence and importance. As Brown (1984, p. 306) notes, “it is always either too early or too late, never the right time, to do a randomized clinical trial.” Chalmers, at the 1985 meeting of the Society for Clinical Trials, noted in the same vein that RCTs are considered impossible because the treatment is either too effective, too ineffective, or too equivocal to be tested. To this we can add that the treatment might also be too controversial to test.

An example of a current treatment whose window of opportunity may never open is provided by Furberg, Yusuf, and Thom (1985). They observe that digitalis has been a mainstay of treatment for patients with congestive heart failure for a very long time. Although that treatment relieves symptoms, its safety and efficacy have never been firmly established. But because it is in such widespread use, few clinicians would be willing to deny it to their patients on a random basis, and there will probably never be a clinical trial to test it.

Levy and Sondik (1982) identified the state of science of a medical treatment as a criterion for initiating clinical trials; their description of this criterion is similar to what we term the window of opportunity. They, along with others (e.g., Ditchley Foundations, 1980; Dudley, 1983; Mosteller, Gilbert, and McPeek, 1983), believe that clinical trials are most appropriate at some middle ground in the state of knowledge about a medical technology. The question here is the potential value to the medical community of the results of the clinical trial,3 manifested in a balance between complete knowledge and total ignorance. If a lot is already known, an expensive clinical trial might provide too little new information; if too little is known, then basic clinical research may be preferable to a clinical trial. For clinical trials of technologies in a state of flux, there is a risk that the design of the trial will have to be overhauled in the middle of the study as new findings invalidate the trial’s foundations. A classical case in point is the very successful clinical trial of the Salk polio vaccine, which had only

3The window of opportunity, then, refers to the potential value of the trial; the anticipated effect of the trial is addressed below.
limited effects because the Sabin vaccine almost immediately replaced it. Had the Sabin vaccine been a couple of years earlier in its development, the test of the Salk vaccine would have been almost impossible to complete.

The ideal state for maximizing the information gained from an RCT is one in which the a priori best guess among the efficacy of the alternative technologies is one of equality (Ditchley Foundations, 1980; Lorenz and Rohde, 1979). This is not to say that current medical practices are already past the stage of being tested by a clinical trial. Indeed, it is when sufficient doubt arises about the effectiveness of a current practice that the time is best for a clinical trial.

Hennekens (1984) defines the window of opportunity in terms of ethical and practical reasons for doing the trial. There must be doubt about the medical practice to be tested, but still enough belief in its potential to be willing to test it. Additionally, there must be enough scientific knowledge about the practice to make the trial scientifically coherent, yet not so much knowledge as to make the results of the trial completely predictable. When the window of opportunity is open, then careful planning can obtain results that are otherwise unavailable. Hennekens cites the Physicians’ Health Study, in which practicing physicians consented to be randomized to heart disease and cancer prevention regimes, as an example of how the window works. This large study examined the efficacy of beta-carotene in the prevention of cancer and the efficacy of aspirin in the reduction of cardiovascular mortality among male physicians 40 to 48 years of age.

The practical manifestation of the window of opportunity is in obtaining physicians who are willing to randomize their patients to the arms of the RCT.4 When a new treatment is introduced, physicians are reluctant to risk sacrificing a patient to an untried treatment. When a treatment is established as a current standard, physicians are reluctant to risk not providing it. If a treatment is controversial, physicians on one side of the controversy are reluctant to treat patients according to the other side. Only during a period of doubt on the part of individual physicians (as opposed to collective doubt expressed as differences of opinion) can the RCT be done.

It is instructive to examine our four case study clinical trials in terms of the window of opportunity. The window was most prominently mentioned in the EC/IC trial. There, the investigators decided to study EC/IC instead of carotid endarterectomy or medical treatment, because the window was closed for those candidates. The

---

4Patients must of course be willing to enter the study as well. But, in practical terms, the bulk of the problem lies in persuading the physicians; they, in turn, are effective in persuading their patients.
Sugarloaf Conference, as well as the vigor of the newly established Lung Division of NHLBI, opened the window for IPPB and other treatments of congestive pulmonary disease. The proposed CCU trial was not funded in large part because the reviewers believed the window of opportunity to be closed, in spite of reviewer and investigator agreement that the topic was important enough to merit study. Finally, the window of opportunity for breast cancer surgery appears to have been opened by the active intervention of the principal investigator and the NCI administrator, who campaigned within NSABP and NCI for the trial.

ANTICIPATED CHANGE IN CLINICAL PRACTICE

Closely related to the window of opportunity is the anticipated change in clinical practice of a trial. Unless one believes that the trial will make a difference in clinical practice, the trial is not worth undertaking. Our respondents varied with respect to the explicit need to consider how a particular trial would affect physician behavior. Some seemed to accept as a matter of faith that physicians are educable, while others seem to worry a good deal about whether the results of a trial would be heeded.

Anticipated change in physician behavior is one of the factors specifically considered by NHLBI in their deliberations about initiating a clinical trial (Levy and Sondik, 1982). For NHLBI, if the outcome of a trial will not make a difference, then the incentive to do the trial is reduced. The analysis involves a cost-effectiveness component that, although important, does not by itself determine whether the trial is to be sponsored.

Why Clinicians Ignore Research Evidence

In an ideal world, the evidence from clinical trials would disseminate to the medical community and would refine medical practice in a process of continuous improvement. Unfortunately, this appears not to be the case. The result is the antinomy that the expected nonresponse by the clinical community becomes the reason for not trying to influence it.

Spodick (1982a) identifies four reasons why medical practices might continue to be used in the face of clinical trial evidence of their inefficacy:

1. Physicians tend to accept established medical practices and current therapeutic hypotheses without testing them. This
has been true since the beginning of medical practice (witch-doctory) through the recent past (bloodletting, purging) to the present.

2. Physicians are understandably always looking for a “magic bullet” to cure the ills of their patients. This leads them to misplaced zeal when a candidate shows promising preliminary results, leading to premature adoption of innovations.

3. Physicians are generally ignorant of the niceties of scientific and statistical methodological correctness and will uncritically accept poor data, leading to a sort of Gresham’s Law whereby large amounts of bad data displace smaller amounts of good data.

4. When faced with intractable medical problems, physicians will grab at proposed solutions, figuring that anything is at least as good as nothing. From both a cost-benefit view and a view based on absolute well-being for the patient, that may not be true.

Spodick attributes these reasons to behavioral attitudes on the part of physicians, including a reverence for authority, a respect for tradition, an Olympian view of medical practice as superior to “mere” research, and a belief that the physician-patient relationship mandates that the physician must do something rather than admit that there is nothing known to be efficacious that can be done.

To this list by Spodick (1982a), we might add that physicians may have both financial and educational investments in extant technologies that they are reluctant to abandon. Both new equipment and new training tend to be expensive in money and time; this leads to an understandable if lamentable proclivity to maintain present habits unless they are demonstrated to be disastrously in error.

The fault does not lie entirely with the practicing clinicians, however. Given their general training, it is unrealistic to expect that they will wade through, much less be attracted to, the methodological detail that accompanies published reports of clinical trials. Their interest is in the results and practical implications. The result is that physicians do not do a good job of evaluating RCTs, even when they read about the trials. Yet, current journal publication customs mandate that the report of a clinical trial provide enough detail to satisfy potential critics of the study, and these customs have survived the test of time. What is lacking is an intermediate publication outlet that would take clinical trials that have passed muster in the formal publications.

\footnote{The point made in this paragraph emerged in a roundtable discussion of journal editors at the 1985 meeting of the Society for Clinical Trials.}
process and distill their contents to a form amenable to practitioners' tastes.

The Office of Technology Assessment (1983) concluded that clinical trials do not have the effect on the medical practitioner community that they should. Their conclusion was based in part on several directed studies. One such study (McGrady, 1982) surveyed family practitioners, providing four prototypical cases, each with two questions about the appropriate treatment, and tabulated whether the responses were concordant with the results of recently published clinical trials. Overall, only 32 percent of the responses were concordant with the recommendations of the trials. This figure, low as it is, is consistent with other evidence on the medical community's slow acceptance of the findings of medical research (Kanouse et al., forthcoming). Indeed, given the low response rate of this survey (39 percent), and the likelihood that responders to surveys might be more aware of scientific findings than nonresponders, the 32 percent figure might be considered too high.

Making Clinical Trials Relevant to Practitioners

Hawkins (1984) looked at studies evaluating the benefit of clinical trials to future patients. She found 20 studies, six examining the effects of clinical trials on the cost-effectiveness of treatment and 14 on physician knowledge and acceptance of trial findings. Her findings on knowledge and acceptance paralleled those of the Office of Technology Assessment (1983): Generally the trials had little effect on physician practice, as measured by physician knowledge, acceptance of the findings of the trial as valid, perception of findings as applicable to their patients, and appropriate application of those findings to their patients. The potential benefit of clinical trial findings for patients, as measured by the cost-efficiency analyses, was however present. If the findings of the trials were incorporated into medical practice, then patients would have improved status and better prognoses, and would incur lower health delivery costs.

Studervant (1977) cautions that controlled trials should not be the sole arbiter of clinical practice. In addition to the scientific findings, one needs to know the varied goals that different technologies can be expected to accomplish, the values of the patient for those goals, and, therefore, what is best for the patient. For example, a patient with ulcers will seek, with differing importance, pain relief, healing of the ulcers, prevention of recurrences, prevention of complications, and minimization of time and monetary costs involved in treatment. As Studervant argued, although ulcers do better with hospitalization, the
side effects in time and money outweigh the marginal benefits for most patients. Although operations often have more risk of immediate mortality and cost than medical treatment, some people elect surgery in these situations. Clinical trials can tell us about the relative safety and efficacy of technologies, but they cannot substitute for the individual cost-benefit calculus that the patient, with the advice of the physician, must perform in deciding which treatment to undergo. Sturdevant (1977, p. 1181) concludes, "My message is that results of controlled trials alone cannot be expected to define standard therapy. Rational therapeutic decision making requires judgments about issues that cannot be certainly settled by controlled trials."

Our case studies showed the influence of anticipated community reaction on the relevance of proposed clinical trials. The IPPB and breast cancer surgery trials were supported because of a belief that the results would influence physicians to alter their practices. The proposed CCU trial was turned down in part because of some belief that no matter how it turned out, few physicians would choose to take their acute myocardial infarction patients out of intensive care. As one respondent put it, "No matter how the trial turned out, I would still put my father in an ICU if he had a heart attack." Only for EC/IC was the anticipated effect on physicians not an apparent major factor in determining the relevance of the trial, and that is because the investigators were apparently genuinely in doubt as to the outcome of the trial.

AVAILABILITY OF ALTERNATIVE METHODOLOGIES

Because of the effort and cost of RCTs, they are often deemed not relevant for a questionable practice if there are alternative research techniques to answer the question. Our respondents took RCTs as the research design of first choice and did not generally seek alternatives. This could be in part a bias in our sample; after all, we were led to people who have been actively involved with clinical trials. Even so, we found ample evidence, both in our interviews and the literature, that alternatives to RCTs are actively considered.

Levy and Sondik (1982) comment that when NHLBI considers sponsoring a clinical trial, alternative means of obtaining the same information, including lesser-scaled clinical trials, data registries, or other means are considered. Zelen (1984) notes that there are several research strategies available to medicine, each with its uses and each with its limitations, and that the RCT is but one, if a major one, of those strategies.
Gehan (1979, 1984) argues that when we know enough from past experience with a technology to determine the population values of measures of interest, then there is no need for a control group, and the experimental group may be compared with history. This has a decided savings in the number of patients required to conduct a trial, the time needed to obtain findings, and dollar expenditure, all at little or no loss to scientific validity.

Clemens and Shapiro (1984) compared RCTs with several alternative designs, including case control studies and various cohort studies. All designs were compared on scientific, ethical, and logistic grounds—susceptibility to bias, capacity to estimate efficacy, ease of obtaining data, etc. For the particular instance of testing the efficacy of pneumococcal vaccine, a test of a standard practice, they recommended a non-randomized study because of the ethical problem of withholding the vaccine if it was useful and the large sample size needed to conduct a prospective study when the incidence of pneumococcal infection in the population is so low.

Feinstein (1984b) sees circumstances when RCTs are inapplicable, including when there are ethical objections to randomization, when a protocol cannot be well-defined, or when they are simply impracticable. But when the issue is prevention, RCTs can pose large logistical problems. First, because not all patients will catch the disease (indeed, the proportion can be very small for even a "common" condition), large sample sizes are needed, and extra care must be taken to stratify the population on possibly confounding dimensions. Also, a longer time must be spent waiting for something to happen, during which time the technology in question might become outdated. There is a greater need to modify the design in mid-experiment, and a consequent need for more control over the study to make those kinds of important decisions. Feinstein concludes that for certain prophylactic issues, although RCTs are the best method in the abstract, they become impractical and must be supplanted by alternative research strategies. His point is that observational studies, when carefully designed, can answer such questions at least as well as RCTs.

Some investigators believe that the "gold standard" set by RCTs is so important that there can be no alternative technique. Chalmers (1981) is the champion of this position; on his side are various methodological experts (e.g., Schwartz, Flamant, and Lellouch, 1980) who argue that only an RCT can guarantee (in a statistical, probabilistic sense) that a study is free from bias. Sacks, Chalmers, and Smith (1983) examined the conclusions of studies using RCTs vs. historical comparisons and found that the studies using historical controls tended to show new therapies to be effective, while RCTs tended to show no
differences. In a medical environment characterized by rapid progress, the very meaning of historical control may be brought into question; it is rare that things are sufficiently the same now as they were even five years ago to make such comparisons legitimate. Our sample of 47 index clinical trials from the 1979 NIH Inventory of Clinical Trials buttresses this conclusion; 70 percent of those trials randomly assigned subjects to practices.

Blackburn (1984, p. 402) represents a sizable contingent of researchers who believe that the RCT has been overemphasized as the uniquely qualified clinical research tool when he states,

We appear to have agreed here that the RCT is the gold standard when the question to be studied can be stated clearly, therapy is simple, the disease is not complex, one can afford to pay for the greater knowledge gained and there is a need to determine relatively small effects in a phenomenon that affects masses of people. It may particularly obtain when the phenomenon is a long-term chronic disease process, the outcome is delayed, treatment effect lags, and there are inadequate ideas or few opinions available for treatment or prevention.

The current state of affairs is fairly close to Blackburn’s recommendation; the tremendous personal, institutional, and financial costs to conducting a decent clinical trial mean that it will be employed sparingly, with alternative means of investigation used if they can in good faith provide reasonable approximations to what would be discovered by a clinical trial.

Although opinions may differ about the relative values of clinical trials and other less demanding methodologies, both clinical and NIH officials are well aware that clinical trials cannot be used to investigate all questions of potential interest. Clinical trials may be inappropriate to the question at hand or there may just not be enough money or talent available to conduct trials as often as one might wish. Inevitably, some questions have to be explored through alternative methodologies.

In some cases, NIH review panels have recommended that questions be explored by means other than a clinical trial. The CCU case study, where a clinical trial was not funded, is an example. Alternative methodologies were considered in the deliberations preceding initiation of the EC/IC and the IPPB trials.

Finally, if an RCT is deemed the best way to test a relevant current practice, that does not mean alternative methodologies are forgotten. If the proposed RCT comes to be regarded as infeasible (see Sec. VII), then the alternatives are reconsidered and the best of them is adopted.
VII. THE FEASIBILITY STAGE

Even if a clinical trial is deemed relevant for a current medical practice, several diverse feasibility constraints must be satisfied before the trial can be initiated. First, there are technical constraints as to whether the important questions can be answered. Then there are moral constraints as to whether we should take the steps necessary to obtain an answer or even whether we should seek an answer. Moreover, there are personal influence constraints, which reflect the inevitable truth that some people's voices carry more influence than other people's. Finally, there is the resource constraint that a trial must be paid for. To interest a sponsoring agency, not only must a proposed clinical trial be relevant in the abstract, it must fit within some agency's mission and be in a format with which the agency feels comfortable.

METHODOLOGICAL CONSTRAINTS

Any research project must be methodologically sound before it can be funded. Clinical trial proposals, in common with other clinical research proposals, suffer more from methodological deficiencies than laboratory research proposals (Cuca, 1983); this puts them at a disadvantage in competition for grants and makes the evaluation of responses to RFPs more difficult. Proposals to test current practices, because they in essence involve tests of a null hypothesis of no benefit for the main practice of interest, present additional complications and are consequently at a greater disadvantage. One respondent commented that any proposed clinical trial can be rejected on the grounds of inadequate methodology if the Institute really doesn't want to support it.1

That methodological constraints are the reason why many potential clinical trials do not take place is universally acknowledged; to properly design an RCT is nontrivial and requires expertise both in the substantive area of the trial and in methodology. Levy and Sondik (1982) identify adequate methodology as one of the most important considerations used by NHLBI in deciding whether to sponsor a trial. Among

1The necessary methodological imperfection of any study does not stop at the initiation phase. Virtually any published report of a clinical trial can be condemned on methodological grounds if the reviewer does not like the direction the results are taking.
the factors that are specifically considered are the trial design parameters, resources, desired results, number of centers, patient sample size, randomization process, and stratification rules. The logistical feasibility of obtaining the objectives of the study are of primary importance; to fund the project, it must have a good chance of success.

The question of how to appropriately conduct a clinical trial is one that has been extensively addressed in the clinical trial literature; several "how-to-do-it" books have provided practical summaries (e.g., Friedman, Furbeg, and DeMets, 1981; Schwartz, Flamant, and Lelouch, 1980; Spilker, 1984). Pocock (1985), noting several deficiencies common to many published clinical trials, nonetheless believes that the essential theoretical work in statistics has already been done and that what remains is to communicate the necessary expertise to medical researchers so that good RCTs can be done.²

There is general agreement that NIH, especially NHLBI and NEI, have the methodological expertise to conduct clinical trials. Those two Institutes have methodological specialists dedicated to insuring that clinical trials are sound. Other Institutes were considered to be more or less methodologically competent, depending on the affiliation of the person doing the evaluations; it is, however, certainly fair to state that all of the NIH officials involved with clinical trials are acutely aware of methodological problems.

The idea of a centralized agency for clinical trials methodology has been introduced by the Institute of Medicine (1983), among others. Opinion about the worth of such a center is mixed. Those in support believe that the major issues of clinical trials cross cut all Institutes and deserve a unique peer review process, with special review boards knowledgeable about clinical trial methodology. There should be coordination in the sequence of clinical trials; some respondents reported a wasteful duplication of effort. Research is required on the methods and processes of conducting multicenter RCTs, and no extant agency has a mission to support such research. A centralized agency could satisfy all of these needs.

Those opposed to the centralized agency object on three grounds. First, each area of medicine has its own unique problems to be dealt

²Both American and European respondents noted that American research is far in advance of European research in this regard. In part they attributed this to Americans having more funds to spend on medical research and so insisting on a higher quality product, and in part to cultural differences leading to different ethical standards. They reported that many European countries have legal and traditional barriers to randomizing patients and the use of placebo. As a consequence, American investigators tend to reject many European findings on methodological grounds, possibly more than is warranted. Some clinical trials in the United States, then, are duplications, although with more methodological sophistication, of European efforts.
with in clinical trials research. To adequately design a study, one needs to be familiar with content as well as with methodology, and the centralized agency would necessarily be weak in the former area. Second, a centralized agency might inhibit the development of decentralized competence throughout the rest of NIH. Third, there is a fear that yet another level of bureaucracy would not solve the problem, but might only make it more difficult to initiate a trial, as there would be another committee to convince of the worth of the project. It is preferable, according to these people, for the Institutes to hire methodological specialists as they see the need. Moreover, as the sophistication of the medical community increases, the problem of methodological expertise as arcane lore will diminish.

The discussion of specific methodological issues that follows will be concerned only with issues of particular importance in determining the feasibility of the design of a proposed clinical trial of a current practice; the general issues of clinical trial methodology are equally applicable but are omitted in the interest of maintaining a narrow focus. The issues that we shall discuss here are: (1) the size of the trial, (2) the scope of the trial, (3) selecting the right variables to measure, (4) choosing a randomization procedure, and (5) explanatory vs. management trials.

The Size of a Trial

The number of patients required by the trial is a methodological issue of major importance in determining feasibility. Yusuf (e.g., Furbeg, Yusuf, and Thom, 1985; Yusuf, Collins, and Peto, 1984; Yusuf et al., 1985) has stated that, especially in RCT's of standard practices, differences among the arms of the trial are likely to be small, thereby requiring large sample sizes to guard against Type II (false negative) errors. Unless these sample sizes can be obtained, it may not be worthwhile to conduct the trial. For prospective studies of prevention techniques, where the proportion of enrolled participants who will contract the disease is fairly small, clinical trials require large numbers of participants. There are alternatives to massive trials, including meta-analytic statistical procedures, but because of the inevitable small differences among investigations, these are never quite as convincing as a direct comparison.

---

3 These are procedures for systematically integrating the results of many individual studies that compare the same practices into a single statistically best estimate of the magnitude of difference between them (Hedges and Olkin, 1985).
The Scope of Clinical Trials

Quite apart from the number of patients in the trial, a major issue is how extensive to make the trial. On the one hand, a multicenter, complicated clinical trial has the capacity to address several issues and, by virtue of its many controls, to make more definitive conclusions. Moreover, in such trials, it is quite easy and inexpensive to add "side studies" of interest to participating investigators, thereby maintaining their active interest and cooperation. Particularly when sample sizes are large, the attraction of partially recouping costs by extending the scope of the study is great. The problems and benefits of these extensive trials have been amply discussed (e.g., Glicksman et al., 1980; Marks et al., 1984; Meinert, 1980).

On the other hand, there is a need to frame questions in answerable terms and not to try to do too much in one study. Peto (1983) advocates clinical trials that address only one major question and are very simple in nature, minimizing the number of forms, the amount of data collected, and the work to be done. No attempt is made to answer any of the potentially interesting questions that are secondary to the main problem of interest. Such simple trials can be quickly mounted, make it easy to enroll physicians' cooperation, and answer the primary question quickly. Gehan (1979) argued that things should be kept as simple as possible, including minimizing the number of patients involved and the length of time the patients are followed.4

The issue addressed by these two positions is the importance of the main objective of a clinical trial. Friedman, Furberg, and DeMets (1981, p. 8) state, "Each clinical trial must have a primary question. The primary question, as well as any subsidiary questions, should be carefully selected, clearly defined, and stated in advance." This seemingly innocuous advice is intended to guard against the design of studies that attempt to do too many things and wind up doing none well. A good study must have specific and well-defined objectives and have a design oriented toward answering the primary objectives. Weiss et al. (1983) further observe that the outcome of a study is in large part determined by its design, so the design of the study should be dictated by the major objectives of the study. Lewis (1982) criticized many trials on, among other bases, their not having clearly defined objectives and on having too complicated experimental protocols.

For clinical trials of current practices, the main question addressed by the trial is likely to be a fairly simple question of whether a

4We hasten to add that these pleas for simplicity are not naive, but are made in full awareness of Einstein's dictum that "Everything should be made as simple as possible, but not simpler."
particular technology works as well or as cost-effectively as the practicing community believes it does. Moreover, because that technology is already being used in the medical community, there should be some history on its claimed efficacy and for which patients it is believed to be effective. These factors argue for the need to select practitioners and patients that are representative samples of the population of current users of the medical practice, hence for large trials with narrow foci; Peto's (1983) examples provide worthwhile models for designing such studies.

Choosing the Right Variables to Measure

The question of which variable to select as the major dependent variable is perhaps not as simple as it seems. The dependent measure should be directly related to the main interest of the study and able to be readily interpreted by practitioners wishing to apply the results of the study. Whether to undertake a clinical trial may well depend on whether a dependent measure can be obtained that is appropriate to the central issue. Lorenz (1982) notes that when inappropriate endpoints are selected, clinical trials will often fail, even when they are otherwise methodologically correct.

For some trials, simple measures of mortality suffice. But for many comparisons of practices, the coarseness of measures of mortality preclude finding differences. Moreover, the modern philosophy of patient outcomes now regards it as important to include measures of functional status, physical capacity, mental health, and the patient's subjective self-rating of health in a health status assessment (Brook and Lohr, 1985). When the quality of life is a component of the question driving the trial or when cost-effectiveness is an underlying motive for the trial, then simple measures will not suffice. To construct these required measures of health status, the participation of behavioral scientists experienced in psychological measurement and of economists experienced in cost-efficiency analysis can be a worthwhile investment; their contributions can help design a feasible clinical trial. In studies examining effective treatment for breast cancer, for example, if survival and disease-free survival are the only dependent measures, then there is no real basis for deciding between the Halsted radical mastectomy and total mastectomy plus radiation (Rockette et al., 1982). It is only on the basis of other measures, including patient acceptance and well-being, that a preference for less-radical procedures is manifested.

Generally, there are no clear-cut differences in outcome expected between the treatment arms of a clinical trial of a current practice; otherwise the trial would be unnecessary. For this reason, then, it is
especially important that trials of current practices should be undertaken only if it is possible to obtain well-defined, reliable, and valid dependent variables that measure the outcomes of central interest.

Choosing a Randomization Procedure

Even when there is agreement that an RCT is the appropriate research design, there are differences of opinion about how to perform the randomization. Although random assignment to treatment when a patient enters the trial is the best way to gather information from a purely abstract point of view, several authors have introduced logistical and ethical objections and advocate the consideration of alternative methods of randomization. There is often a tradeoff between the optimal form of randomization from a statistical point of view and the form of randomization that will be ethically satisfactory and will supply the most patients in the least amount of time.

Fisher et al. (1985b) used a technique they called “prerandomization” in the case study trial of breast cancer surgery. A patient was randomly assigned to a treatment before she was recruited into the experiment. The physician in charge knew what treatment she would obtain and could choose whether to enter her in the experiment. She, knowing which treatment she would obtain, could choose whether to participate. This technique became necessary when the enlistment rate for patients became so dangerously low that continuation of the trial was in jeopardy.\(^5\)

Zelen (1981) discusses alternatives to randomization that avoid the ethical problem of randomly assigning patients to treatments. This ethical problem is exacerbated for current medical practices, when assignment to a control group means withholding a standard technology from that patient. Zelen recommended what he labeled “consent randomization.” Patients who are randomly assigned to receive the standard technology (be that the control group for a test of an innovation or the experimental group for a test of a current practice) are not asked for their consent to participate. But patients receiving the non-standard technology are asked for their consent to participate in the experimental therapy; if they refuse, they are reassigned to the standard group. A full discussion of Fisher’s, Zelen’s, and other randomization procedures would draw us beyond the intended scope of this report. It should suffice to state that alternative randomization procedures have raised considerable debate (e.g., Fletcher, 1984), have

\(^5\)Although prerandomization appears to us to offer the serious possibility of biased patient enrollment, the controversy over this trial was not methodological but rather ethical (Hillman and Kahan, 1986).
been the topics of plenary sessions of the Society for Clinical Trials, and are a major consideration when the decision is made whether to initiate an RCT.

**Explanatory vs. Management Trials**

A distinction is made in the literature (e.g., Ciampi and Till, 1980; Sackett, 1980, 1983) between explanatory clinical trials and management clinical trials. An explanatory trial has the objective of finding out why a particular technology works, and a management trial has the objective of finding out how well it works. Similarly, Schwartz and Lellouch (1967) distinguish between the explanatory question of why something works and the management (to them, “pragmatic”) question of does something do better in practice. Ritter (1980), implicitly adopting management as the sole justification for a clinical trial, fears that statistical controls aimed at explanatory objectives may cause the design to answer only trivial questions instead of the important ones.

Detre (1984) sees the distinction between management and exploratory trials as central to feasibility. If the trial is exploratory, then it is appropriate to wait until the state of science can specify clearly and concisely how the protocol should be constructed. Before that time, the trial is not really feasible. For management trials, the outcomes are more in terms of a patient’s point of view and there is no need to delay constructing a satisfactory, if not perfect, protocol that patients and physicians will find acceptable.

Clinical trials of current practices would tend historically to fall more into the management trial category rather than the explanatory trial one. Sixty-eight percent of the 47 current practice trials we selected from the 1979 NIH Inventory of Clinical Trials were classifiable as management trials. All of the case studies we examined were management trials in that they tested whether the treatment worked in a general sample of afflicted patients. In such instances, the medical practice has been ongoing, and the question is more one of whether it does any good in its natural setting (compared with placebos, alternative simpler treatments, or just doing nothing); it is of less importance to explore the reasons why it works.

**ETHICAL CONSTRAINTS**

In a research world of heightened sensitivity to individuals' rights, extensive regulations governing experimentation with human subjects, detailed informed consent forms, and Institutional Review Boards
(IRBs) that must approve any study using human subjects, ethical constraints can make infeasible even well-intentioned, well-designed, and potentially beneficial clinical trials. In the Levy and Sondik (1982) model of NHLBI decisionmaking the ethical aspects of the proposal must pass muster, including those of the technologies themselves, the provisions for informed consent, and consideration of oversight of the intermediate results of the trial in case early data mandate changes in treatment protocols. For many of these ethical aspects, regulations promulgated by PHS dictate requirements for any proposed RCT.

The consensual opinion (e.g., Rosinga, 1982) is that randomized clinical trials are ethical if done according to certain standards, which include (Lebacqz, 1980) respect for persons, beneficence, and justice. Klein (1979) summarizes the steps that are necessary to ensure ethicality, including:

- Providing for the informed consent of the treated;
- Choosing alternative treatments, placebos, or no treatment controls that are honestly believed to have a likelihood of being as good as the experimental treatment;
- Maintaining the confidentiality of the data;
- Insuring that participation of subjects is voluntary;
- Insuring that the design of the study is not influenced by any conflict of interest that might be held by the source of funding;
- Complying with governmental regulations;
- Providing for compensation for subjects who may be injured as a result of their participation in the study; and
- Making special arrangements to insure the above steps for special groups, such as the elderly, incompetent, or underage, who may not have the ability to make decisions that are in their own best interests.

These ethical constraints apply with full force to clinical trials of current medical practices. If the ethical standards cannot be guaranteed, the trial should not be conducted. Our interviews and case studies indicate that NIH personnel and investigators are acutely aware of these constraints and consider them appropriately.

Our respondents argued that RCTs are inherently ethical if guidelines such as those outlined above are followed. Patients in clinical trials receive more attention and hence better treatment than patients not on a research protocol, no matter to which arm of a clinical trial they are assigned (Lellouch, 1979). It is certainly more ethical to have patients participate in planned, systematic research than it is to subject them to spontaneous experimentation where the results of the
manipulations have little chance of advancing medicine. The advances
gained by clinical trials help prospective future patients and may help
the subject patient with regard to the future course of the disease.
Several interview respondents stated that, particularly in a clinical trial
of what proves to be an ineffective current practice, randomization
means that a certain proportion of the patients in the study will, by
virtue of being in the study, not receive the inferior standard treat-
ment.

Our respondents did raise certain ethical problems inherent in clini-
cal trials. One, particularly important for clinical trials of current
practices, involves the belief about the effectiveness of treatment. If
you believe that Treatment A is better than Treatment B, then you
cannot ethically test A against B. But if you believe that A is no worse
than B, then you can ethically conduct the trial. Moreover, reproduc-
ibility is the basic standard of science. But if a clinical trial has been
done, ethical considerations prohibit its replication unless the original
is dismissed as hopelessly inadequate. For this reason, the method
must be good the first time a trial is conducted. By this last criterion,
which was reflected in a number of our interviews, an unethical trial is
one that is badly done. It wastes resources, serves the patients ill, and
makes good trials harder to initiate.

An ethical consideration that is particular to trials of current med-
cal practices, especially when one of the arms of the treatment involves
a standard practice followed by a substantial proportion of the medical
community, illustrates the overall ethical issue. That is the problem of
denying a patient a standard treatment when it is not known whether
the alternative treatment is better for that individual patient.6

Ethical problems arise when patients who would receive a given
treatment in the normal course of events are instead shunted to a
placebo or a no treatment control group. Schafer (1982) notes that
physicians are rarely totally equivocal about the relative efficacy of the
alternative treatments of a clinical trial. This creates a conflict
between the ethical obligation to do everything in the best interests of
a patient and doing what is best for science. For example, if the
disease model underlying the rationale for the Halsted radical mastec-
tomy had been valid, then patients given a lesser procedure would have
been "sacrificed" in the sense that they were denied effective treat-

6This is a manifestation of the general problem of individualist ethics. If people could
agree in advance of being sick that they would be willing to be tested, denied expensive,
possibly nonbeneficial treatment, etc., then we would collectively be better off in the
long run. But, inevitably, some (randomly chosen) few individuals would, after the fact,
suffer.
syphilis sufferers in order to study the natural course of the disease is a real case in point. Angell (1984) comments that even when there is no preference for one technology over another on a medical basis, the patient may not be indifferent; this creates a conflict between patient preferences, which should be respected, and scientific aims. In West Germany, doctors may be criminally liable for damages resulting from a patient who was randomly allocated to what turned out to be an inferior treatment (Silverman, 1979).

The answer to this ethical dilemma has two components. The first is the higher obligation argument. According to this view (Silverman, 1979), when the value of a technology is doubtful, there is an ethical obligation to find out whether it works that transcends any ethical obligation to provide it to an individual patient. McKinlay (1979), in agreement with this argument, sees that the government, which underwrites a substantial proportion of medical care costs, has an obligation to know that its money is well spent. If there is some doubt that a technology works, it must be tested; if the technology does not work, resources are being spent that could be better allocated elsewhere. Considerations of the philosophical position of McIntyre and Popper (1983) on the role of error in medical judgment (discussed in Sec. II) buttress this argument. The physician does act in what he believes the best interest of the patient to be but must acknowledge the possibility that his beliefs are incorrect. This uncertainty in his own opinion means that the physician may not be serving the interests of the patient, a possibility whose likelihood can only be explored by scientific trial. Therefore, even when a patient is randomly assigned away from a standard practice, it may be in his own interests.7

The second component of the answer to the ethical dilemma is that patients might be more willing to be subjects in clinical trials than is generally believed. If physicians acknowledge their basic uncertainty about the efficacy of different treatment alternatives from the start of their interaction with the patient, then the patient will not feel that random assignment creates the risk of not obtaining the best treatment available. Further evidence for this hypothesis comes from Cassileth et al. (1982) who surveyed hospital patients and the public in Philadelphia to ascertain their feelings about participating in experiments. Most respondents did not mind being subjects; in fact they believed that it was useful and worthwhile, of direct benefit to themselves, and gave them a feeling of esteem by helping future patients.

7The implications of this argument go beyond the scope of this report, and include the need to ascertain in systematic fashion the level of uncertainty of physician's beliefs and to apply Bayesian statistical techniques to decisions concerning both the initiation and planning of clinical trials. See Ederer (1982) for an interesting discussion of the statistical issues thereby engendered.
A practical implication of this ethical problem is that data are often scrutinized in mid-trial to see whether the trial should continue. If the standard treatment is clearly superior to the alternative, then the trial should be stopped. This step introduces statistical problems that, although complicated, are largely resolvable (e.g., DeMets, 1984; Elashoff and Reedy, 1984; Freedman, Lowe, and Macaskill, 1984).

INFLUENCE OF PROPONENTS

Although there is no formal recognition of the fact, personal influence plays a large part in determining which clinical trials will be sponsored by NIH or, for that matter, by any organization. This, of course, is not really news; the “old boy” network has existed for some time and continues to exist, even though some of the old boys are girls. In our interviews, the opinion was often expressed that there are some trusted researchers who can be counted on to do quality work, and whatever they propose has a substantial chance of being funded. Several NIH officials deplored the paucity of qualified clinical trial investigators; it was acknowledged that several fields are dominated by a few senior people. Our case study of breast cancer surgery demonstrated the influence of the NCI administrator and of the principal investigator in convincing the research and practitioner communities and NIH itself that the trial was feasible; without their personal influence, that trial might not have been undertaken. Our case study of EC/IC showed similar influences. The principal investigator of that study was given a pilot grant to formulate the trial after two previous attempts to obtain funds for the study by another researcher had been denied. Clearly, NINCDS had an interest in the study, and it is not unreasonable to conjecture that the principal investigator helped sustain that interest.

Several times our respondents observed that the principal investigators of major clinical studies are older, established investigators. Younger individuals do not appear to have the influence, the moral profile, or the authority to convince the medical community of the need for a trial. It is also harder for less-established investigators to obtain the cooperation of the clinical sites needed to stage a multicenter RCT. Finally, young investigators, looking at tenure and promotion decisions, cannot take the risk of waiting years for the fruits of their research to appear in print. These reasons may be partly why the investigator who submitted our case study of CCUs chose not to pursue the topic after his grant application was denied.
The influence of established investigators may be a manifestation of a general tendency to avoid risk on the part of NIH staffers. By funding proven investigators, they can minimize the risk of “wasting” funds on poor research. The negative consequence of this risk aversion is, of course, that new investigators with potentially major innovative ideas find it more difficult to obtain support.

INSTITUTIONAL FACTORS

The final constraints we shall consider are those imposed by the typically large resources needed to conduct an RCT. To be sponsored, a proposed clinical trial must be seen to fit within the mission of a sponsoring agency and to function within the format that the agency believes best. These institutional constraints are largely of two sorts. First, the research must fit within the programmatic interest of the agency. Second, the proposed interaction between the agency and investigator must meet the agency’s standards of control. We shall discuss those institutional constraints within the context of NIH; other agencies both within and outside the federal government operate in fairly similar manners.

Programmatic Interest

For an Institute of NIH to sponsor a proposed RCT, no matter how worthwhile in the abstract, the trial must fit the Institute’s programmatic interest. For different Institutes, this requirement can mean very different things. NIH regards itself as a biomedical research agency and considers its first responsibility to be to biomedical science. Therefore, among most Institutes, there is a great reluctance to do clinical trials for health care service delivery or cost-effectiveness reasons. For most of our NIH respondents, a biomedical scientific component was a necessary aspect of a clinical trial if it was to be funded by their Institutes. On occasion, political pressures could in effect force an Institute to support a trial that did not have a basic biomedical emphasis, but these were not frequent.

The Interaction of Institutions and Investigators

Institutes within NIH vary considerably in the way that they prefer to interact with investigators. These ways are reflected in their preferred funding modalities. There are three principal ways in which
money is provided to outside investigators to conduct clinical trials (Malone, 1985):

- **Grants.** Individual research project grants (R01s) and program project grants (P01s) are awarded to institutions on behalf of principal investigators to carry out specific projects or long-term research programs, respectively. Investigators submit grant proposals, which compete with each other for funding. Once an institution receives a grant, NIH does not supervise the conduct of the trial; its influence is restricted to its potential to fund future grants.

- **Contracts.** Research contracts are awarded to institutions for scientific inquiry directed toward particular topics specified by the awarding Institute. Most typically, the contracts are competitively awarded following investigators' responses to an RFP to carry out specific work. Contract performance is closely monitored by NIH to insure compliance with the project goals.

- **Cooperative Agreements.** These are a newly developed form of NIH sponsorship that are midway between grants and contracts. They are similar to grants in that they are awarded to support research, but are like contracts in that the NIH Institute takes a major role in the design and performance of project activities. Institute involvement is less than in regular contracts, however.

NIH Institutes differ widely in their use of these three mechanisms for supporting clinical trials. NHLBI and NIDR rely almost exclusively on contracts to conduct their clinical trials. NCI and NEI are turning more and more to cooperative agreements. NIADDK and NINCDs, however, strongly prefer the grant mechanism. Each believes that its mode of functioning is the best for the particular problems that it faces.

The investigators we interviewed largely favored grants over the other mechanisms. They believed that grants gave the initiative for proposing clinical trials to the investigators themselves, who were better aware and had a better idea of relevance than did NIH support staff. There was general agreement among investigators that they are better than NIH staff at creating a good study design. With a grant, the investigator has flexibility, independence, and autonomy; with a contract, matters depended too much on the particular NIH supervisor. One respondent preferred grants because once the budget is approved, it is set; with cooperative agreements or contracts, budgets can be modified, interfering with efficient research planning. Another
respondent commented that the skills of allocating monies and conducting scientific research are very different, and that although NIH might be good at the former, it is not as good as the investigators it supports at the latter.

The other side of the picture emerged from interviews with NIH personnel. These officials were largely concerned with controlling clinical trials so that the right issues would be investigated in methodologically appropriate ways. Contracts, in which the institutional objectives are spelled out in a precise manner, work well in this regard but incur a heavy institutional cost in design and supervision. For this reason, there is an increasing move to cooperative agreements, where the Institute and teams of investigators work as a team. Groups with proven records of productivity arrange general research programs with the Institute for funding and then design specific studies that are approved by the Institute before initiation. In this way, according to NIH personnel, the Institutes have the control they need with less of the direct supervision and design costs. Additionally, the Institute takes less of a risk with its funds; it is supporting people with a record of quality.

Even Institutes that largely fund through grants were concerned with control. One respondent from such an Institute noted that they negotiate intensively with prospective grantees and demand revisions to the grants before funding. Moreover, once an RCT is funded, the Institute plays an active and continuous role through its independent monitoring committee. In this way, it maintains the degree of control that other Institutes obtain through contracting.

How does the particular funding modality affect the feasibility of a clinical trial? If an investigator believes that a clinical trial is relevant for a current practice, that person must look to Institutional funding policies to see if it is feasible to obtain funding for the trial. For contracts, this can be a largely pro forma matter, as the awareness of the problem and its relevance have been determined by the Institute before its decision to issue an RFP. But even here, interest is not guaranteed. Institutional review of proposals may lead to reconsideration and the conclusion that the possible trials suggested do not serve the interest of the Institute, resulting in the contract being deferred or withdrawn. For grants, programmatic interest can be expressed formally through the peer review and advisory council deliberations that determine the fate of a proposal (Henley, 1977a,b,c) or through informal negotiations between applicant and institution. For example, NEI will add or subtract points from the peer review priority score depending on whether the grant fits into its formal plan. As that plan is readily available to all would-be investigators (National Advisory Eye Council, 1982), the bulk of grant applications received do fit into the NEI agenda.
We may once again turn to our case studies to examine how the interaction between Institute and investigator influenced the initiation of the clinical trial. The breast cancer surgery trial evolved out of an interaction between NCI and the NSABP that resembles current cooperative agreements. Although the trial was largely funded as a contract, the design was worked out between the two agencies, with NSABP, one of the best-regarded cooperative clinical groups, taking most of the initiative. The EC/IC trial, although funded as a grant, was the result of intensive negotiations between the Institute and investigators. NINCDS showed a clear interest in the trial and gave the principal investigator a small pilot grant to demonstrate the feasibility of the study. The IPPB trial was a straightforward reflection of Institutional priorities following its Sugarloaf Conference. Contracts to do specifically targeted research made it clear that the trial would be feasible. Finally, the CCU proposed trial seems to have foundered in part because it did not fit NHLBI programmatic interests. Although there was recognition that the trial could be useful, no incentive was offered the investigator to revise the proposal or to scale down the proposed trial to demonstrate feasibility. Moreover, it is possible to speculate that NHLBI, which strongly prefers the contract modality for large-scale clinical trials, was reluctant to fund such a trial through the grant modality.
VIII. ISSUES IN INITIATING CLINICAL TRIALS

In the course of our investigations, we found that a small number of issues of potential policy importance arose repeatedly. The three most prominent of these issues are summarized below.

1. Does the NIH have a uniform process for deciding whether to initiate clinical trials of current practices?
2. Is there an adequate system for identifying questionable current medical practices?
3. How should clinical trials on questions of the cost-effectiveness of medical practices be sponsored?

DOES NIH HAVE A UNIFORM DECISIONMAKING PROCESS?

It became rapidly evident from the lack of evidence in the literature, from our interviews, and from our case studies that no single model could describe how NIH decides to sponsor a clinical trial of a current medical practice. Procedures and proclivities vary from one Institute to another, and even within particular Institutes the process varies from one time to another and from one area or type of medical research to another.

Nonetheless, our decisionmaking model (Fig. 3) contains at least the elements of the decisionmaking process of each Institute that we examined in our interviews and case studies. Although there were major differences in the extent to which an Institute became aware of questionable medical practices, the dimensions of relevance were important to each Institute; and feasibility, in all of its dimensions, appeared to be a major consideration in determining whether a trial was funded.

Everyone we interviewed noted that before a trial is undertaken there must be a clear determination that the question to be examined is relevant. But the criteria for relevance were always stated in rather vague terms, and even at an individual level there was little attempt to arrive at systematic rankings of the relevance of proposed subjects for trials. Certainly, there has been no attempt at defining NIH-wide criteria for determining the relevance or feasibility of a proposed trial.

The apparent diversity in NIH decisionmaking procedures might be due to the widely varying degrees of formality with which the separate Institutes set priorities for research. Some have formal programmatic
goals and direct applicants for funds to those goals. Others have advisory committees that help determine what direction should be taken. And still others rely on the implicit knowledge of the Study Sections in assigning priority scores to grant applications. But apparently for all Institutes, the awareness that some current practices should be studied arises not from any systematic survey of current practice but indirectly from other, less formal evaluations by practitioners and investigators of the state of medicine. The means by which these evaluations take place seem to be highly dependent on the particular personalities involved and on informal relationships among investigators and between investigators and the NIH.

Correspondingly, Institutes also differ in the way that they interact with investigators. Some wait for investigators to propose studies or to respond to RFPs for specific kinds of research. Others (NCI is the best example) prefer to work with established cooperative groups, deciding periodically how each group's energies can best be employed. Although it might be predicted that Institutes with strong programmatic goals might seek out investigators, and those with implicit goals wait for investigators to take the lead, that was not the case. For example, NEI, which has explicit and published programmatic goals (National Advisory Eye Council, 1982), relies on unsolicited grant applications to have those goals met.

We also found variations in the poignancy of tradeoffs faced by administrators in different Institutes. In some cases, it was generally believed that the limitation on doing RCTs of current practices was not so much a shortage of funds as a lack of suitable topics for trials or a lack of personnel (both inside and outside NIH) to organize and conduct the trials. In those cases, it was believed that important RCTs generally got funded. In other cases, administrators seemed to perceive every decision to do a trial as a heart-rending decision not to do some other trial or to forgo funding other valuable nonclinical research.

Finally, we found important differences in the degree to which outside forces (Congress or such organizations as the American Cancer Society) influence the direction of research pursued by particular Institutes.

This diversity within NIH is not necessarily undesirable. The state of science and the nature of disease processes relevant to particular Institutes vary considerably. In neurology, for example, there are few treatments to test, whereas in oncology there are large numbers of chemotherapeutic agents to be tested in combination. The former situation mandates bench research instead of clinical trials, and the latter calls for careful selection among candidate RCTs because of the limited patient population available for testing. Thus, it is neither
surprising nor inappropriate that NINCDS and NCI have differing philosophies and procedures regarding clinical trials of current practices.

With all of these differences within the Institutes and their surrounding environments, it is also neither surprising nor inappropriate that there is no single model across NIH of the decisionmaking process that determines whether to initiate a clinical trial of a current medical practice.

IDENTIFYING QUESTIONABLE CURRENT MEDICAL PRACTICES

There is no uniform way in which current practices are questioned by NIH or other potential research support agencies. There is, however, a consensus (although not a universal one) that in the present system, a great many current practices that probably should be investigated are not questioned.

Part of the problem could be the influence on awareness of the feasibility constraints imposed by institutional factors. When an Institute has a focused programmatic interest that is known to investigators, there are disincentives to search in areas that, even if they produce worthwhile projects, are not likely to result in research funding. With respect to NIH, the most frequently cited observation in this regard is its dedication to basic biomedical science. But the observation may be overstated. Remington (1982) argues persuasively for broad-based research support from NIH. Both applied and basic research should be supported because one can never know which findings of basic research will have applications. Both targeted (through contracts and cooperative agreements) and investigator-initiated (through grants) research are necessary because there are varied sources of insight into problems. Diagnosis, prevention, and treatment are all necessary because modern medical practice has both a need to know about and a need to act upon diseases. Therefore, argues Remington, choosing to support any one of these research aims to the exclusion of the others loses a major part of the picture. We found no objection to Remington’s arguments from our respondents.

Our investigations lead us to conclude that if either efficacy and safety, biomedical theory, or quality of life considerations cause a practice to be considered for a clinical trial, then the mechanisms of relevance and feasibility regarding the practice appear to function well. The problem as we see it is that there is no systematic search for suspicious current practices. One can be left with the impression that candidate practices for reappraisal receive attention only by chance and the talent of the medical community.
Even the FDA, an agency whose mission includes examination of current drugs, has only a limited monitoring function. Although FDA requires that drug companies report to it any observed adverse reactions to their products, the agency has no control over the use of a drug for purposes beyond those for which the drug was originally approved. That is, FDA continues to monitor the safety of currently used drugs, but it has no systematic way to monitor the efficacy of those drugs for new purposes. In this way, drugs can come to be used without formal tests of their efficacy.

There is therefore a need for some new means of cooperation and coordination among government agencies for the evaluation of current medical practices. Clearly, interest in discrediting unsafe, ineffective, or cost-ineffective practices extends well beyond NIH and FDA. Indeed, concern for these matters is probably greater in such agencies as HCFA that pay for "standard" medical care. Means should be found for cooperative identification of practices to be studied, for cooperative design of studies to guarantee that the needs of different agencies are met, and for cooperative funding of studies.

There is a need for systematic, fairly inexpensive methods to search for questionable medical practices. Such methods might include generation and examination of registries, monitoring of practice patterns to detect natural experiments that arise from differences in practice patterns, and regular examination of large medical datasets collected for other purposes (e.g., Medicare records). These datasets can be used to detect systematic differences in the costs, patterns of care, or outcomes of treatment in different settings. For the most part, these survey methods could help identify candidate practices for more formal research. To supplement these passive methods, Institutes could constitute special committees, composed of researchers, specialty practitioners in the relevant fields, and generalists, to convene and identify the conditions under which certain targeted practices should be used. If this group were unable to reach consensus on such conditions, either because of polarization or because of a common state of ignorance, that would signal the practice as one requiring investigation. More "health services research" making use of big but inexpensive datasets might yield some important results in the way of identifying current medical practices for reevaluation and possible elimination.

Systematically questioning current medical practices can be performed by many agencies. It is not necessary that the entity that produces a list of questionable practices be responsible for performing the

---

1Awareness of this need arises not only out of our examination of the initiation of RCTs, but from a wide variety of sources. See, e.g., Brook and Lohr (1985) and Wennberg (1984).
ensuing research. Regular conferences (such as the Princeton Conference held annually by NINCDS) sponsored by NIH can and do provide awareness of questionable practices; databases within Institutes (such as the SEER registry of NCI) can perform a similar function. Health care technology assessment, whether by NIH (through the Consensus Development Conferences of the Office for Medical Applications of Research, for example), NCHSR, OTA, the Institute of Medicine of the National Academy of Sciences, or nongovernmental organizations such as the American College of Physicians can evaluate current practices. But at present, there is no coordinated effort that utilizes what information is already available, much less plans for better data gathering. Proposals such as that offered recently by Wennberg (1984) should be seriously entertained, and every agency with a mission of investigating health care should be linked in some form of awareness network.

SPONSORING CLINICAL TRIALS TO STUDY COST-EFFECTIVENESS

Our examination of the role of the cost-effectiveness of medical practices as a motivation for a clinical trial leaves us a bit confused. On the one hand, much has been written encouraging trials to study what is obviously an important question. Few, if any, observers doubt the need to contain health care costs. Everyone advocates seeking more efficient approaches to health services. RCTs are a demonstrated means of obtaining information about the cost-effectiveness of medical practices. Why, then, has cost-effectiveness had so little effect on clinical trial decisionmaking?

The simplest reason is that this concern is just too recent to have been reflected yet. It takes time, after all, for worthwhile topics for clinical trials to be identified and suitable proposals to be designed. All of our respondents were of the opinion that cost-effectiveness would be a more important issue in the future (see also Brook et al., 1984; Frederickson, 1980).

There is considerable discussion about who should sponsor trials of cost-effectiveness. At least part of the explanation for the infrequency of clinical trials of the cost-effectiveness of medical care is that there is simply no government agency in a position to carry out such investigations. As some of our respondents put it, cost-effectiveness as a research issue falls through the cracks in the Establishment. Among the candidates suggested have been NIH, HCFA, NCHSR, a new

---

2This issue is important and complicated, but an extended analysis of it is beyond the scope of the contract supporting this report.
agency created expressly for that purpose, the private sector, and combinations of the above.

But even if an agency with a mandate to study cost-effectiveness did exist, there would be barriers to this type of research. The most difficult of these grows out of the basic noncomparability of costs and benefits in the health services field. No one would argue that when confronted with two equally effective treatment strategies, we should not favor the less expensive. But alternative treatment strategies are rarely seen as equally effective; one almost always carries risks or benefits that the other does not. We all agree in principle that the additional benefits of the more expensive procedure may not be sufficient to justify its additional costs, but how in practical cases are we to determine how much benefit justifies how much cost? Such questions have no straightforward answers and in some circumstances are better left unasked. But if it is true that considerations of cost-effectiveness often lead us to questions that are analytically intractable and perhaps ethically disturbing, what agency or research organization would not prefer to stick to simple questions of medical efficacy?

Finally, there is a practical obstacle to cost-effectiveness research. The traditional sites of major clinical trials have been major research medical centers. Community hospitals often do not have the patient base, the control over medical practice, the facilities for data collection, or the financial leeway to conduct major clinical trials. It is widely recognized that the style of medicine practiced in major medical centers often differs markedly from that practiced in local community hospitals or in private physician practices. The former enjoy financial support that is not dependent solely on patient care; they have staff who are rewarded to a large degree on the basis of the research they do rather than purely for the patients they treat; staff in medical centers have ready access to new techniques and equipment. In all these ways major medical centers differ from community hospitals and private physician practices. But it is in these latter hospitals and practices that the bulk of health care in the United States is provided. What is cost-effective in a university medical center may or may not be cost-effective in a community practice, and trials conducted in the former may have limited applicability to the latter.

Thus, there are institutional, political, bureaucratic, philosophical, and practical obstacles to RCTs directed toward questions of cost-effectiveness. Although there is no doubt about the need for information regarding these questions, there seems to be disagreement about the appropriateness of obtaining this information through large-scale clinical trials. Even if RCTs of cost-effectiveness are desirable, they are currently difficult to implement. If more RCTs of cost-effectiveness
issues are desired, then we require at the very least that some government agency be given both a clear mandate and adequate funding to carry them out and that some means be found to conduct them in settings where the bulk of U.S. health care is delivered.
Appendix A

LIST OF RESPONDENTS

Dr. Don Archibald
Cooperative Studies Program
VA Medical Center, West Haven, Connecticut

Dr. Jean-Pierre Boissel
Unité de Pharmacologie Clinique de l'Hôpital Neuro-Cardiologique
Lyons, France

Dr. Gustaf Brynolfsson
Planning and Rationalization Institute of the Health Services
Stockholm, Sweden

Dr. Thomas Chalmers
Mount Sinai Hospital
New York, New York

Dr. Greg Curt
National Cancer Institute

Mr. Fred Ederer
National Eye Institute

Dr. Stan Finkelstein
Sloan School of Management
Massachusetts Institute of Technology

Dr. Robert Flamant
Institut Gustave-Roussy
Villejuif, France

Dr. Lawrence Friedman
National Heart, Lung and Blood Institute

Dr. Michael Friedman
National Cancer Institute

Dr. Curt Furberg
National Heart, Lung and Blood Institute

Dr. Lawrence Geiter
Centers for Disease Control
Atlanta, Georgia

77
Ms. Hellen Gelband  
Office of Technology Assessment

Dr. Ronald Geller  
National Eye Institute

Dr. Israel Goldberg  
National Eye Institute

Dr. Murray Goldstein  
National Institute of Neurological and Communicative Disorders and Stroke

Dr. Robert Gordon  
National Institutes of Health

Dr. Barbara Hawkins  
Johns Hopkins School of Hygiene and Public Health

Dr. Roger Herdman  
Office of Technology Assessment

Dr. Stephen Heyse  
National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases

Dr. Gladys Hirschman  
National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases

Dr. Teh-wei Hu  
Institute for Policy Research and Evaluation  
Pennsylvania State University

Dr. Kenneth James  
Cooperative Studies Program  
VA Medical Center, Palo Alto, California

Dr. Egon Johnsson  
Planning and Rationalization Institute of the Health Services  
Stockholm, Sweden

Dr. Christian Klimt  
Maryland Medical Research Institute

Dr. Carl Leventhal  
National Institute of Neurological and Communicative Disorders and Stroke

Dr. Guy McKhann  
Department of Neurology  
Johns Hopkins Hospital
Dr. Paul Meier  
Department of Statistics  
University of Chicago  

Dr. Curtis Meinert  
Department of Epidemiology  
Johns Hopkins School of Hygiene and Public Health  

Dr. David deMets  
Department of Statistics  
University of Wisconsin  

Dr. Teresa Radebaugh  
National Institute on Aging  

Dr. Mans Rosen  
Planning and Rationalization Institute of the Health Services  
Stockholm, Sweden  

Dr. Anna Sandberg  
National Institute of Arthritis, Diabetes, and Digestive and  
Kidney Diseases  

Dr. Richard Simon  
National Cancer Institute  

Dr. Edward Sondik  
National Cancer Institute  

Dr. Michael Walker  
National Institute of Neurological and Communicative Disorders and Stroke  

Dr. Norman Weissman  
National Center for Health Services Research and Health Care  
Technology Assessment  

Dr. Robert Wittes  
National Cancer Institute  

Dr. Salim Yusuf  
National Heart, Lung and Blood Institute
Appendix B

SUMMARIES OF CASE STUDIES

EXTRACRANIAL-INTRACRANIAL
ARTERIAL ANASTOMOSIS

This case study (Anderson and Kahan, 1985) reviews the inception of a grant from the National Institute of Neurological and Communicative Disorders and Stroke to conduct a randomized clinical trial of the effect of extracranial/intracranial anastomosis on stroke incidence. We examine the investigators’ decision to apply for the grant and the Institute’s decision to fund the application.

The investigators’ decision to propose the study was based on their view that an RCT of the EC/IC bypass technique was relevant to stroke prevention because stroke is a major disease, because EC/IC is of potential importance in preventing stroke, and because a trial had the potential to affect medical practice. The study was feasible because it was possible to design an appropriate study and to obtain the interdisciplinary and interinstitutional cooperation necessary to obtain the resources required for data collection and analysis, the support of multiple clinical centers for patient recruitment, and an agency willing to fund the study.

Interviews with the investigators and reviews of relevant documents permitted the following conclusions to be drawn regarding the investigators’ perceptions of the relevance and feasibility of the trial:

Relevance

- Stroke has a major influence on health in terms of both its incidence and its severity.
- Stroke prevention was the most workable approach to lessening the effect of this disease, and the EC/IC procedure might benefit a considerable subset of individuals at risk for stroke.
- A trial of EC/IC anastomosis was timely; its results could affect medical practice because the technology for the procedure was stable but the benefits were not yet accepted as certain.
Feasibility

- The clinical group at the University of Western Ontario and the methodology group at McMaster University had established an ongoing relationship that made the design of complicated RCTs possible.
- The effect of the EC/IC procedure could be adequately assessed in a randomized trial of 1000 patients followed for five years.
- The investigators had high status and strong reputations and had the influence necessary to obtain the required support from other clinical centers.
- NINCDS had shown interest in funding an appropriately designed study.

The NINCDS decision to fund the EC/IC trial was made in the context of the NIH grant review process. This process is divided into two major stages: (1) peer review by a Study Section, and (2) review by the Council of the relevant institute. The NIH is limited to grants of no more than five years in duration; because the EC/IC trial was planned to last eight years it was reviewed twice, once in 1977 and again in 1981/82. The 1977 Study Section review was very positive on all aspects of the scientific merit of the EC/IC grant application and specifically referred to its potential effect on the use of the procedure. The 1977 NINCDS Council meeting funded the study straightforwardly. The 1981/82 review was less positive, with some discussion of potential methodological problems and disagreement on the potential effect of the results of the trial on medical practice. But NINCDS believed the renewal should be funded because the study was important and because failure to renew would mean they would receive little information for their substantial investment in the first five years of the trial.

INTENSIVE CARE FOR ACUTE MYOCARDIAL INFARCTION

This case study (Hammons and Kahan, 1985) describes the development of a grant application submitted to the National Institutes of Health proposing a randomized controlled trial to assess the comparative effectiveness of care in a coronary care unit and care in a monitored hospital bed for uncomplicated acute myocardial infarction. The Clinical Trials Review Committee of the National Heart, Lung, and Blood Institute reviewed the application and recommended disapproval.
Information for the preparation of this case study was gathered from a review of the literature pertinent to the effectiveness of intensive care for acute myocardial infarction and to clinical trials, from the written proposal for the study and the written Summary Statement detailing the results of the review of the proposal, and from interviews with the principal investigator who developed the proposal and officials of NHLBI, some of whom were involved in the review of the proposed study.

This case study led to the following conclusions:

- The effectiveness of intensive care for uncomplicated acute myocardial infarction has never been rigorously evaluated, despite the importance and expense of this practice.
- Although the proposed study was well-conceived by able investigators and CTRC and NHLBI officials considered the problem to be important, several flaws in the design of the study and incomplete preparation by the investigators led to disapproval of the proposal.
- The literature, the principal investigator, officials of NHLBI, and the members of the CTRC raised several points suggesting that clinical trials of current medical practices may be difficult to carry out. Randomized clinical trials of current practices that are widely accepted by the medical community, such as intensive care for uncomplicated acute myocardial infarction, will be especially difficult. These threats to the feasibility of such a trial, combined with the funding priorities of the National Institutes of Health, make it difficult for a grant application to undertake such a study to be funded.

Many believe that rigorous evaluation of established as well as emerging clinical practices is needed to improve medical practice and to assure that scarce resources are spent on effective practices. This case study suggests that applications to do such evaluations are likely to encounter several impediments to development and funding. Policy measures to increase technology assessment in medicine must take note of these special problems.

TREATMENT OF BREAST CANCER

This case study (Hillman and Kahan, 1985) describes and analyzes the events and considerations resulting in the initiation by the National Cancer Institute (NCI) and the National Surgical Adjuvant Breast and Bowel Project (NSABP) of the B-06 clinical trial of breast-preserving surgery for breast cancer.
The B-06 trial is a three-arm trial comparing the relative efficacy of total mastectomy vs. segmental mastectomy (lumpectomy) with and without radiation as treatment for stages I and II breast cancer. Patient recruitment began in April 1976; the NCI agreed to the trial in January 1977. Thus, this case study largely details the period from 1975–1977, when the protocol was being designed and considered. Nonetheless, because of the relationship of the trial to previous research, and because publication of the trial's results occurred only in 1985, portions of the case study deal with phenomena both before and after the initiation period.

Data for the preparation of this case study were accrued from several sources: a review of the literature pertinent to the B-06 trial and clinical trials in general, extensive interviews with individuals involved in the initiation decision and in the operations of NCI and the NSABP today, and attendance at the spring 1985 meeting of the NSABP. These sources described several influences germane to the initiation of the trial:

- The progress of biological and clinical sciences. New findings challenged the conventional view of breast cancer and hence its accepted surgical treatment. The B-06 trial is a direct descendant of previous work performed by the laboratory of the principal investigator, Dr. Bernard Fisher, and the NSABP. The tension between the biological and clinical foci of the trial is an interesting aspect of the story of its initiation.
- The funding mechanism and administrative history of the trial and the relationship between NCI and the NSABP. There was considerable flux in NCI's organization and approach to funding clinical trials during the period of the trial's consideration. These changes worked to the advantage of the NSABP and the B-06 trial. Although the Cooperative Agreement/Cooperative Clinical Research mechanism used to fund clinical trials today was not in force at that time, many features of the trial's history are similar to that mechanism.
- The importance of influential individuals. Reviewers and NCI officials raised both scientific and ethical objections to the trial; the advocacies of Dr. Fisher and the then Director of the Division of Cancer Treatment, Dr. Vincent DeVita, were important in overriding the objections and permitting the trial to proceed.
- The concerns of participating physicians and ethical issues pertinent to the trial. Initially, patient recruitment into the trial was slowed by physicians' concerns over the divergence from traditional surgical treatment and fears for their relationships
with their patients. These difficulties were overcome by changing the study design to prerandomization—allowing physicians to know the result of the randomization before they decided whether to enter a patient into the trial. The design and ethical issues raised by conducting the study in this fashion are germane to the conduct of clinical trials in general.

- The potential effect of the trial upon physicians' practices and patients' behavior. The hope among those advocating the trial was that, if the trial proved breast-preserving surgery equivalent to more radical surgery, practicing physicians would change to the lesser procedure; in consequence, women would be less fearful of seeking physician consultation and would present at an earlier stage of their disease.

This case study details these and related factors and analyzes their influences within the context of clinical trials in general, providing a concrete example of NIH clinical trials decisionmaking process.

**INTERMITTENT POSITIVE PRESSURE BREATHING**

This case study (McGlynn and Kahan, 1985) describes the events leading up to the decision by the National Heart and Lung Institute\(^1\) of the National Institutes of Health to issue a request for proposals to conduct a multicenter randomized clinical trial comparing the efficacy of intermittent positive pressure breathing with compressor nebulizer therapy in treating patients with chronic obstructive pulmonary disease (COPD).

In 1974, the American Thoracic Society, under a grant from NHLI, sponsored the Sugarloaf Conference to examine the scientific basis of various respiratory therapies used in the treatment of COPD. The purpose of the conference was to bring together experts in the field of pulmonary disease to review the literature concerning the efficacy of respiratory therapy for COPD, consider additional data necessary to assess the therapies adequately, and disseminate the findings so investigators would be stimulated to initiate needed research.

The Sugarloaf Conference concluded that most of the information concerning the efficacy of IPPB was drawn from studies with serious methodological flaws. The consensus was that IPPB had not been demonstrated to be an effective therapy, but its widespread use and expense compared with other therapies warranted a controlled clinical

\(^1\)Now the National Heart, Lung, and Blood Institute.
trial. The conferees recommended that the highest priority for IPPB research be given to a study of its long-term effects.

NHLI recognized the need for a multicenter trial to ensure an adequate patient panel for this study and decided to use its contract mechanism to fund the study. The contract approach permitted central control to guarantee standardization and comparability across sites. The RFP was issued in December of 1975. Five clinical centers were funded: Baylor College of Medicine, Loma Linda University, University of Oklahoma Health Sciences Center, University of California at San Francisco, and University of Manitoba.

Although the investigators were primarily interested in learning more about the disease process associated with COPD, they recognized that some focus was necessary to ensure standardization. Evaluating the efficacy of IPPB therapy provided that focus. Most of the investigators believed that IPPB was an ineffective therapy and that the by-product data related to the disease itself would be the most important outcome of the study.

The IPPB trial illustrates the sort of clinical trial of current medical practice that NHLI considers within its mission. It advanced scientific understanding of the structure and function of the respiratory system in patients with stable COPD, using a therapeutic intervention to standardize data collection and patient recruitment.
BIBLIOGRAPHY


Friedman, Lawrence M., Curt D. Furberg, and David L. DeMets, Fundamentals of Clinical Trials, John Wright, PSG Inc., Boston, 1981.


National Center for Health Services Research, Health Technology Assessment Reports, Department of Health and Human Services, DHHS Publication No. (PHS) 84–3372, Rockville, Maryland, 1983.


