

IDEAS AS REFORMS: THERAPEUTIC EXPERIMENTS  
AND MEDICAL PRACTICE, 1900-1980

by

HARRY MILTON MARKS

M.A. History, University of Wisconsin (1972)  
B.A. History, Hofstra University (1968)

SUBMITTED TO THE DEPARTMENT OF  
POLITICAL SCIENCE  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN POLITICAL SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

© Harry Milton Marks, 1987

The author hereby grants to M.I.T. permission to reproduce and to  
distribute copies of this thesis in whole or in part.

Signature of Author \_\_\_\_\_  
Department of Political Science  
February, 1987

Certified by \_\_\_\_\_  
Professor Harvey M. Sapolsky  
Thesis Supervisor

Accepted By \_\_\_\_\_  
Professor Daniel Metlay  
Department of Political Science

Accepted By \_\_\_\_\_  
Professor Peter Buck  
Program in Science, Technology  
and Society

Accepted By \_\_\_\_\_  
Professor Harvey M. Sapolsky  
Chairman  
Graduate Program Committee  
Political Science Department

MASSACHUSETTS INSTITUTE  
OF TECHNOLOGY

ARCHIVES

JUN 16 1987

LIBRARIES

IDEAS AS REFORMS: THERAPEUTIC EXPERIMENTS  
AND MEDICAL PRACTICE, 1900-1980

by

HARRY MILTON MARKS

Submitted to the Department of Political Science  
on February 24, 1987 in partial fulfillment of the  
requirements for the degree of Doctor of Philosophy in  
Political Science

ABSTRACT

In recent decades, academic physicians and health policy analysts have espoused the use of randomized experiments (RCTs) for obtaining reliable, scientific data about the benefits of medical treatments. Since the early 1900's, the controlled therapeutic experiment has played a dual role as an arbiter of scientific validity and a tool for regulating medical practice. Such experiments have been the chosen instrument of a small cadre of medical researchers who have sought to constrain and direct the average practitioner's use of medical treatments. How and why these reformers thought improvements in the quality of therapeutic research would naturally benefit clinical practice, and the steps they took to bring these improvements about, are the subjects of this thesis.

The thesis examines the activities and scientific practice of professional reformers on the Council of Pharmacy and Chemistry of the American Medical Association (AMA) in the 1910's and 1920's; the influence of these reformers on the Federal Food, Drug and Cosmetic Act of 1938; and the intellectual and practical problems faced by the organizers and government sponsors of controlled therapeutic trials up to the 1950's. I then examine postwar developments, including the role of statistics and statisticians in experimental design, controversies over the design and interpretation of RCTs, and current proposals to use RCTs as a guide in allocating medical resources.

My argument is first, that our public policies and practices for regulating therapeutic drugs during the first half of this century were developed and shaped by private institutions, most notably the AMA's Council on Pharmacy and Chemistry, and the community of academic physicians represented in the National Research Council's Division of Medical Science. Second, that in focusing their efforts on the scientific problem of improving the evidence on which judgments about the risks and benefits of medical treatments are based, therapeutic reformers have avoided the more intractable political problems presented by attempts to regulate or manage the therapeutic decisions of practicing physicians. Third, that the problem of determining what constitutes a "good" clinical decision is as much a problem for political theory as for statistics and/or "clinical" judgement.

The problem of reforming therapeutic practices is treated here as an instance of the more general problem of the relation between science and democracy. In the concluding chapter, I examine contemporary proposals to rely on randomized experiments in determining public policies towards the use of the medical practices and technologies. I argue that if we are to regard the use of RCTs in evaluating medical technologies as more than an administrative convenience, the theoretical claims that RCTs provide an authoritative basis for evaluating and allocating medical goods must be assessed. I analyze the relation between method and science implicit in contemporary proposals for therapeutic reform, and argue that a purely methodological theory of science, as is offered by advocates of RCTs, cannot provide a basis for either scientific judgments or public policy. I conclude with a series of arguments for a more participatory, more public, approach to setting policies towards medical practices, as opposed to relying on more corporatist arrangements which delegate authority and responsibility to the medical community and the private sector.

The methodology of the thesis is largely historical: I have used manuscript collections to identify the individuals and organizations active in therapeutic reform, their activities and their ideology. Collections consulted include the archives of the AMA, records of the Food and Drug Administration in the National Archives and on deposit at the Federal Records Center, records of the National Research Council, the National Institutes of Health, and a variety of private manuscript collections. These materials have been supplemented with interviews and with analysis of the published materials.

## TABLE OF CONTENTS

Acknowledgements.	5
Introduction.	9
Chapter One. A Rational Therapeutics.	21
Chapter Two. Regulating Medicine. The 1938 Drug Act.	51
Chapter Three. Memories of Underdevelopment. The Decline and Rise of Cooperative Research.	92
Chapter Four. A Tale of Two Studies.	130
Chapter Five. Managing Chance. Statistics and Therapeutic Experimentation After World War II	152
Chapter Six. Anatomy of a Controversy. The University Group Diabetes Program Study	187
Chapter Seven. Science and Politics.	235
Appendix. A Note on Sources.	276

Acknowledgements

This project had its origins a decade ago when Marc J. Roberts innocently asked me to edit a Harvard Business School case on the tolbutamide controversy. The subject of the case piqued my curiosity, which was further provoked by my realization that neither of the social scientists who worked on the case had bothered to get their medical facts right, or pay much attention to the scientific issues involved. Professor Roberts has continued to show a lively interest in the various versions of the project and the argument, and offer thoughtful commentary along the way.

That a project critical of the current gospel regarding randomized clinical trials could thrive at an institution which has done so much to promote that gospel is a tribute to the intellectual integrity of those at the Harvard School of Public Health who have supported this work, both morally and materially. I am indebted to Harvey V. Fineberg and Howard Hiatt for encouraging and supporting this project. Frederick Mosteller and John Bailar III have responded with intellectual generosity above and beyond the call of scholarly duty to this interloper in statistics. For their willingness to mix warm encouragement for the enterprise with thoughtful criticisms of the arguments entertained along the way, I am especially appreciative. The Department of Social Medicine and Health Policy, at Harvard Medical School, and its chair, Leon Eisenberg, have been equally generous in offering me a home in which to undertake the bulk of this heretic endeavour.

Historians encountered along the way have been generous in sharing their expertise and their critical reactions to portions of this work. I am indebted to James Harvey Young for guiding me through the tangled path of the Food and Drug Administration records, and for offering his comments on an earlier version of chapter two. Peter Temin has been extremely responsive to inquiries and conjectures from someone who disagrees with him on most points of fact and interpretation regarding the evolution of regulatory policy towards drugs. John Swann has shared his knowledge of the drug industry and of its relations with pharmacology, and offered some valuable correctives to an earlier draft of chapter three.

David Willis provided his offices as editor of the Milbank Memorial Fund Quarterly to get an earlier version of this material extensively criticized by a veritable Who's Who of biostatistics; Russell Maulitz and Diana Long ensured that a good many medical historians would have a crack on this material, by inviting me to present a paper on therapeutic research at their conference on the history of internal medicine at the College of Physicians of Philadelphia.

No author as dependent as the present one on manuscript sources fails to accumulate a large collection of debts to their guardians. For their help and patience in pursuing leads, false and true, I am indebted especially to Gerald Deighton and Eras Rodas of the Food and Drug Administration's Freedom of Information Office; to Ernest Bird of the Federal Records Center at Suitland, Maryland; to Lee Johnson of the Scientific and Technical Records Branch of the National Archives; to

Altha South, of the Judicial, Fiscal and Social Branch of the National Archives; to Marguerite Falucco and Terry Austin of the American Medical Association archives; to Janice Goldblum of the National Academy of Sciences; and for numerous instances of assistance, to Richard J. Wolfe of the Francis Countway Library of Medicine at Harvard Medical School. Along with the help of these professionals, I am indebted to the many individuals who allowed me to interview them and in some cases to consult their professional correspondence and records concerning the events described herein. I owe special thanks to Thomas C. Chalmers of Mt. Sinai Medical Center who was generous in sharing his correspondence and papers with someone whom he knew in advance disagreed with him on numerous issues.

Scholars must eat, and the generous grants from the National Center for Health Services Research [HS 05151] and the Health Services Improvement Fund, Inc. of New York enabled me to worry about the esoteric problems of therapeutic experimentation and medical practice instead of meeting next month's mortgage check.

Harvey Sapolsky has been the ideal thesis supervisor, knowing just when to push, when to kick and when to stay out of the way. For welcoming me as a student to the Department of Political Science, and for discretely overlooking those moments when I relied more on Habermas than on James Q. Wilson, he has my sincere appreciation.

I doubt that this project would have arrived even at its present incomplete stage were it not for the intellectual and moral support of two friends, Peter Buck and Steve Thomas. The time they have spent in recent months, listening patiently to endless fragments of argument

over the phone, and offering the appropriate murmurings of quiet encouragement and incisive criticism in just the right proportions, surely ranks them among the lamed voy. To Steve Thomas, my debt is doubled, for without him to direct me through the tracks of contemporary Anglo-American political theory, this refugee from European marxism would never have found his way to the arguments and books on liberal democratic theory which have informed the conclusion of this thesis.

I would like to dedicate this thesis to two people who have influenced it, and its author. To Barbara G. Rosenkrantz and Edward T. Gargan, for teaching me about the practice of history and for reminding me that we are always living in it.



### Introduction

Bensalem will have an intellectual history consisting of the progress of science. It will have a social history consisting of the impact of science. It will, however, have no political history. The state is withered away.<sup>1</sup>

"Modern" medicine, "scientific" medicine: the terms are virtually synonymous. The modernity of twentieth century medicine consists of its reliance on the physical and biological sciences. Yet the association is deceptive, so familiar that it passes without further investigation. What does it mean, what should it mean, to call medicine a science? Is medicine dependent on science for its tools, its knowledge, or its methods? Is medicine scientific because physicians use an advanced technology, the x-ray, because they rely on a knowledge of pathology and radiology to interpret the technological data, or because of the rigor with which the technology and knowledge are used to shape clinical decisions?

From Boerhaave through Flexner, it was generally believed that establishing medicine as a science meant grounding medical practice in one or more of the scientific disciplines which study the functioning of biological organisms: biochemistry, physiology, genetics and the like. Scientific medicine was a matter of applying at the bedside knowledge produced elsewhere, a conception of medical science which

---

<sup>1</sup> Samuel H. Beer, "Two Models of Public Opinion: Bacon's 'New Logic' and Diotima's 'Tale of Love'," Political Theory (May, 1974), 166.

still thrives today.<sup>2</sup> In this century, another interpretation of the project has been put forth: clinical medicine was, or could be, every bit as scientific as the research laboratory, if "scientific method" were directly applied to judging the results of therapeutic decisions. Paramount among the means for placing medical practice on a scientific basis has been the controlled experiment.

Since the early 1900s, the controlled therapeutic experiment has played a dual role as an arbiter of scientific validity and a tool for regulating medical practice. Such experiments have been the chosen instrument of a small cadre of medical reformers who have sought to constrain and direct the average practitioner's use of medical treatments. How and why these reformers thought improvements in the quality of therapeutic research would naturally benefit clinical practice, and the steps they took to bring these improvements about, are the subject of this thesis.

The narrative begins in 1906, with the founding of the Council on Pharmacy and Chemistry by the American Medical Association (AMA), a small group of chemists and research physicians convened to provide practising physicians with authoritative evaluations of new commercial

---

<sup>2</sup> Claude Bernard's An Introduction to the Study of Experimental Medicine (New York: Dover Publications, 1957), as introduced by L.J. Henderson in 1927 provides the locus classicus for twentieth century proponents of this view. The introduction to a recent article in Science, citing T.H. Huxley on the potential for chemistry and physics to transform biology and medicine, indicates that the tradition is alive and well in the medical community. See George K. Radda, "The Use of NMR Spectroscopy for the Understanding of Disease," Science 223 (August 8, 1986), 640. [The author is aptly titled Professor of Molecular Cardiology at the University of Oxford.]

drug products. Present-day commentators on health policy may wonder why the story does not begin more recently, with the widespread introduction of statistical concepts and approaches to experimentation after World War II.<sup>3</sup> To the contemporary researcher, the modern therapeutic experiment, employing randomized assignment of patients to treatment, "blinded" assessment of treatment results, and sophisticated mathematical techniques in experimental design, analysis and interpretation, represents the apotheosis of the scientific method in clinical research.<sup>4</sup> The development of modern experimental procedures *per se*, however, is not the subject of this thesis. What I am concerned with here is the expectation that qualitative improvements in experimental method will bring about corresponding improvements in the quality of medical practice. Current views about therapeutic experimentation, on this account, constitute the concluding episode in the development of an ideology equating scientific with social progress.<sup>5</sup>

---

<sup>3</sup> In a recent letter to *Science*, the statistician Paul Meier may be found reminding an otherwise knowledgeable journalist that the tradition of large scale cooperative trials of medical treatments did not, after all, commence in the 1960's as the journalist had implied but went back a long way, to the late 1940's and early 1950's to the studies discussed in chapter three of this thesis. Paul Meier, "Correspondence," *Science* 204 (April 27, 1979), 367.

<sup>4</sup> For introductions to the literature on randomized controlled trials, see Thomas Louis, Frederick Mosteller and Butnam McPeck, "Timely Topics in Statistical Methods for Clinical Trials," *Annual Review of Biophysics and Engineering* 11 (1982), 81-104; Sonja M. McKinley, "Experimentation in Human Populations," *Milbank Memorial Fund Quarterly. Health and Society* 59 (1981), 308-323; and Stuart Shapiro and Thomas Louis, eds., *Clinical Trials: Issues and Approaches* (New York: Marcel Dekker, 1982).

<sup>5</sup> The historian, conversely, may wonder why the story does not begin earlier, with accounts of the 18th century surgeon James Lind's experiments in preventing scurvy, or Pierre Louis 19th century evaluations of blood letting, both frequently cited in historical accounts of the

In the circles to which this thesis is addressed, ideology is a tainted term. Only the leaders of sectarian religions and political movements possess ideologies; men of science have rational beliefs. In what respects the beliefs contemporaries hold about therapeutic experiments are "rational" is a question I defer until the concluding chapter, where I will consider the claims made on their behalf by statisticians, physicians and policy analysts. Meanwhile, I intend to examine the beliefs about experimentation to which therapeutic reformers in this century ascribed, and in the vernacular of the social sciences such beliefs constitute an "ideology": a set of assumptions about how the world works, about how people (in this case physicians) in it behave, and about how best to change their behavior.<sup>6</sup>

---

experimental tradition. Historians have generally followed statisticians in presenting a tale of forgotten ancestors and notable conceptual breakthroughs, a hagiographical tradition which results in tracing clinical experimentation to the Book of Daniel in the bible. Like other origin theories, such accounts are mythic rather than properly historical. However often such predecessors as the 18th century surgeon, James Lind, and the 19th century clinician, Pierre Louis are cited by latter-day experimenters, their examples were re-discoveries after the fact, historical symbols rather than historical influences. Such myths are interesting in their own right, but they are not the subject of the present book. See J.P. Bull, "The Historical Development of Clinical Therapeutic Trials," Journal of Chronic Diseases 10 (1959), 218-246; Abraham M. Lilienfeld, "Ceteris Paribus: The Evolution of the Clinical Trial," Bulletin of the History of Medicine 56 (Spring, 1982), 1-18.

<sup>6</sup> George Lichtheim's "The Concept of Ideology" [in his The Concept of Ideology and Other Essays (New York: Vintage Books, 1967), 3-46], remains a useful historical introduction to the changing meaning of the term in social and political analysis. The work of Jurgen Habermas, while rarely focused on the concept of ideology per se, is fundamentally concerned with examining the theoretical and practical relations between knowledge and interests which give birth to the problem of ideology. See Knowledge and Human Interests (Boston: Beacon Press, 1971) and the essays in Theory and Practice (Boston: Beacon Press, 1973). Habermas' contributions represent the principal theoretical advances to the discussion of ideology since the work of Lukacs and

All ideologies have their blindspots, assumptions so crucial to the worldview held that they are rarely articulated and even less frequently given critical scrutiny. In the present instance, the sensitive area is the connection between improvements in the quality of therapeutic evaluation and changes in therapeutic practice. The physicians and scientists discussed in this book have all been critical of the way in which physicians employ drugs and all have believed that if physicians knew more about the effects of drugs in specific clinical conditions, they would use drugs more appropriately and hence more effectively. They have all favored certain improvements in experimental procedures, such as the use of untreated controls, which would make it possible to rely more on the therapeutic claims issued by medical authorities and drug manufacturers alike. Where they have differed is in the relative importance they attached to the improvement of methods, institutions and character in ameliorating therapeutic practice.

What distinguished therapeutic reformers in the first third of this century from their successors was their willingness to explicitly acknowledge a connection between science, morality and the scientific practice of medicine. Men of science were said to have different,

---

Mannheim in the 1920's which closes Lichtheim's account. On the incorporation of the sociology of knowledge in analyses of the natural sciences, Steven Yearley offers a thoughtful account of the state of play in his "The Relationship Between Epistemological and Sociological Cognitive Interests: Some Ambiguities Underlying the Use of Interest Theory in the Study of Scientific Knowledge," Studies in the Philosophy and History of Science 13 (1982), 353-388. Yearley's objections to "explaining" adherence to scientific theories or methods in terms of the classical theory of interests have influenced the methodological stance taken here.

higher, motives than men of commerce; research physicians were expected to pursue the truth in their laboratories and clinics, and practising physicians to follow those truths, as established by the experts, in treating their patients. The work of the AMA's Council of Pharmacy and Chemistry, described in chapter one, was premised on a conviction that practising physicians would acknowledge both the scientific authority of the more experienced, more knowledgeable expert in selecting treatments, and the moral obligation of employing the most carefully validated remedies for his clients. For this generation, beliefs about character provided the missing link between experiment and practice.

The Council's approach to reform provided the basis not only for professional efforts at improving the quality of medical practice, but the intellectual framework in terms of which regulatory legislation was interpreted by the individuals in charge of the Food and Drug Administration (FDA). In implementing the Federal Food, Drug and Cosmetic Act of 1938, an effort described in chapter two, FDA officials relied heavily on the community of expert physicians in formulating their judgments about the merits of specific drugs. The crucial dilemma for both therapeutic reformers and their allies in government was how to handle the physician who failed to use drugs in accordance with the recommendations established in the studies so precisely conducted by experts. While both professional reformers and FDA regulators hoped that most physicians would practice the best medicine possible with the tools at hand, by the late 1930s neither group wholly expected that all physicians would do so. The task of changing the therapeutic practices of the less discriminating physicians was left to educational institutions

and organs of scientific authority such as the AMA's Council on Pharmacy and Chemistry.

Producing the knowledge which was meant to guide the individual physician in the use of novel drugs presented no less of a problem for reformers. As the number of such drugs increased in the 1930s and 1940s, so did the difficulties of obtaining dependable information about how new drugs compared with existing therapies, and how best to use them. Chapters three and four describe the intellectual and social development of the tradition of cooperative studies, organized efforts to enlist specialists from around the country to follow standardized guidelines in studying new compounds, so that the results of treatment in large numbers of patients could be rapidly accumulated and reliably compared. By combining the observations of many experienced physicians, the organizers of such studies hoped to overcome the limitations inherent in studies conducted in clinics dominated by one or two individuals. But in completing their studies according to plan they encountered a difficulty similar to that which had hampered reformers in directing the therapeutic practice of individual physicians: the reluctance of individual researchers to follow agreed upon procedures for selecting patients, administering treatments and evaluating the results.

The cooperative studies tradition left an ambiguous legacy to researchers in the postwar era. While those planning and funding research remained sympathetic to cooperative evaluations of new drugs, the earlier confidence that some combination of bureaucratic organization and moral exhortation could ensure the successful completion of a study

according to plan was gone. Methodological reforms inherited the faith previously invested in organization and character. Henceforth, procedures such as the use of untreated controls and "blinded" assessments of therapeutic outcomes were to be the arbiters of therapeutic merit and the guarantors of scientific integrity. An emphasis on such procedures, more than the application of associated mathematical techniques, was the principal contribution of the statisticians whose role in therapeutic research is analyzed in chapter five.

By improving the quality of experimental studies, methodological reformers in the postwar era hoped thereby to wean practising physicians away from an overconfidence in therapeutic innovations or in established, but inadequately tested, remedies. But they entered their campaign for therapeutic reform armed only with the truth, or rather, what they claimed were more reliable means for getting at it. Sharing in the general postwar conviction that scientific progress begets social change, they lacked even the rudimentary organizational and moral theories of social change on which previous generations of reformers had grounded their hopes for improving therapeutic practice. Accordingly, they had no means for addressing, much less answering, the question of how the results of experiments conducted in accordance with the recommendations of statistical consultants were to be translated into clinical practice.

In chapter six, I examine how reformers' claims to provide an authoritative basis for therapeutic practice were challenged when the FDA sought to modify recommendations regarding the use of oral hypoglycemic drugs on the basis of a decade long randomized controlled trial sponsored by the National Institutes of Health. This case study of the



controversy over the University Group Diabetes Program provides an opportunity for exploring the political and intellectual dilemmas faced by proponents of controlled clinical experiments in the current period. The concluding chapter considers the claims currently being made regarding the use of controlled experimentation as an instrument for professional reform, in the light of this previous history.

The insistence of therapeutic reformers that physicians practice the best medicine possible may be seen as an effort of the profession's scientific leadership to remake the profession in their own image. It can equally be regarded as a defensive strategy, an attempt to stave off outside intervention by saving medicine from its own worst aspects. The historical sociology of the medical profession within the twentieth century has yet to be written, however, and until it is, we will not know enough about the relative position of the scientific community in each era to reliably interpret their behavior. What can be analyzed are the representations made by reformers in the name of science, and the manner in which they sought to accommodate their beliefs about the scientific practice of medicine to prevailing circumstances.

The events and institutions discussed in this account are properly described as pre-political; the institutions of government which are generally thought to be the subject of political analysis play at most a cameo role. The choice is deliberate, my argument being that public policy towards the regulation of therapeutic practice has been shaped almost entirely by the beliefs of so-called "private" groups and organizations, most notably the community of professional reformers

whose activities are described in chapters one and two.<sup>7</sup> The influence of private organizations on the formulation of public policy is hardly a new phenomenon in American politics. There is an almost equally long tradition of questioning the legitimacy of delegating social authority to private groups.<sup>8</sup> The bulk of this literature, however, concerns itself with arenas directly related to economic policy or to the management of interclass relations.<sup>9</sup> For most

<sup>7</sup> Of the formal institutions traditionally the subject of political analysis, only the judiciary has had a substantive and independent effect on the scientific standards applied in drug regulation. [See especially the discussion of judicial interpretations of the 1906 Food and Drug Act in chapter two. With the development of the Administrative Procedures Act, however, the judiciary retreated to the stance of reviewing principally the administrative procedures by which these scientific standards were promulgated and applied, rather than the standards themselves. A fuller historical analysis of the way in which the law has affected the conduct of contemporary drug legislation would be most welcome, however.

<sup>8</sup> Awareness of this influence among American social scientists seems to come in waves; prior to the recent flourishing of interest in neo-Marxist and corporatist analyses, attention seemed to peak in the 1950's, with analyses of the character of various programs from the 1930's and 1940's. See especially Grant McConnell, The Decline of Agrarian Democracy (Berkeley: University of California Press, 1953) and Philip Selznick, TVA and the Grass Roots. A Study in the Sociology of Formal Organization (Berkeley: University of California Press, 1949). For the corporatist literature, see note 9, below.

<sup>9</sup> For introductions to the corporatist literature, in which the emphasis on economic policy and class relations are clear, see the essays by Alessandro Pizzorno and Philippe C. Schmitter in Suzanne Berger, ed. Organizing Interests in Western Europe. Pluralism, Corporatism, and the Transformation of Politics (New York: Cambridge University Press, 1981). The recent collection edited by Wolfgang Streeck and Phillippe C. Schmitter provides a fair sample of the kinds of policies examined under this framework; see their Private Interest Government. Beyond Market and State (Beverly Hills: Sage Publications, Inc., 1985). The formal delegation of political authority to private organizations is a rarer phenomenon in the United States, where the organization of private interests is more likely to be consultative, as in the Council on Wage and Price Stability formed during the Nixon administration. For a discussion of American politics which takes account of recent writings on European corporatism, see Lawrence

theorists, the rightness or wrongness of giving private organizations a role either in formulating or implementing public policy rests on pragmatic criteria: do the groups effectively represent the crucial interests and constituencies affected by the policies in question?<sup>10</sup> But the claims of medicine to social authority do not rest principally on notions of interest representation. Nor, as even its critics acknowledge, do they depend entirely on medicine's ability to deliver the goods. Rather, they depend on medicine's claims to scientific status.

Medicine's critics take the profession's scientific aspirations as both illusory and delusive: medicine is not, they argue, a science and its claims in this regard are chimerical.<sup>11</sup> What such demystifying criticisms ignore is the authority of science within the profession itself. Like similar concepts in political theory—equality, justice, democracy—the notion of science is at one and the same time central to the medical community's view of itself, and sufficiently

---

D. Brown, New Policies, New Politics: Government's Response to Government Growth (Washington, D.C.: The Brookings Institution, 1993).

<sup>10</sup> See especially the discussion by Claus Offe, "The Attribution of Public Status to Interest Groups: Observations on the West German Case," in Berger, Organizing Interests (n. 9), 123-158. Brown's essay, cited above (n. 9), is similarly concerned with the dysfunctional aspects of spinning off government functions to non-governmental or quasi-governmental organizations such as Health Systems Agencies. Discussions of an older theory in which it is public opinion, not private interests, which are represented may be found in Charles S. Maier, "'Fictitious Bonds ... of Wealth and Law': On the Theory and Practice of Interest Representation," in Berger, Organizing Interests, 27-62 and "Five Types of Politics," part one of Samuel Beer's British Politics in the Collectivist Age (New York: Alfred A. Knopf, 1966), 3-102.

<sup>11</sup> See E. Richard Brown, Rockefeller Medicine Men (Berkeley: University Of California Press, 1979); John Ehrenreich, "Introduction," in John Ehrenreich, ed. The Cultural Crisis of Modern Medicine (New York: Monthly Review Press, 1978), 1-38.

polyvalent so that there are numerous claimants to the title, and numerous programs for its instauration. An evaluation of the concept calls for both an explication of the normative claims made in its name, and a critical examination of the practices through which those claims are meant to be realized. The historical analysis of therapeutic reform, in that sense, is intended to inform the concluding theoretical discussion of the place controlled experiments can have in health policy.

Contemporary reformers have claimed a superiority for the use of randomized controlled trials on the grounds that "...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."<sup>12</sup> Such claims may be evaluated by examining the logical and epistemological status of statistical inference. A purely methodological theory cannot in itself, however, provide a scientific basis for medical practice or health policy. Something more is needed: an account of the processes by which competing scientific judgments are to be adjudicated, and assimilated into practice and policy. Such an account calls for an examination, not only of the possible aims of medical science but the institutions of medical reform as well. It is time now to proceed to that task.

---

<sup>12</sup> John Bailar III, "Introduction," in Stuart Shapiro and Thomas A. Louis, eds. Clinical Trials: Issues and Approaches (New York: Marcel Dekker, 1983), 1.

Chapter One. A Rational Therapeutics

On August 15, 1922, Elizabeth Hughes arrived at the Toronto clinic of Fred Banting. A fourteen year old diabetic being treated with a starvation diet of 900 calories, Hughes had been losing weight steadily all Spring. On arrival, she weighed forty-five pounds and had all the signs of terminal diabetes: "hair brittle and thin, abdomen prominent, shoulders drooped, muscles extremely wasted, subcutaneous tissues almost completely absorbed. She was scarcely able to walk on account of weakness." Banting immediately began treatment with an experimental drug. Within two weeks, she was able to tolerate a normal diet and by November, she returned home to resume a normal life.

Elizabeth Hughes died on April 15, 1981, at the age of seventy-three.<sup>1</sup>

Banting's experimental drug was insulin, the means for rescuing thousands of diabetics from almost certain death. Insulin was the most successful accomplishment of what one historian has termed a "therapeutic revolution": the introduction of a seemingly interminable series of potent therapeutic agents.<sup>2</sup> Few of these were as effective as

<sup>1</sup> The account of Hughes' condition, and that of other diabetics treated with Banting and Best's wonder drug, comes from Michael Bliss' The Discovery of Insulin (Chicago: The University of Chicago Press, 1982), 144, 151-165, 244.

<sup>2</sup> The term is Charles Rosenberg's, from "The Therapeutic Revolution: Medicine, Meaning and Social Change in Nineteenth-Century America," in Morris J. Vogel and Charles E. Rosenberg, eds. The Therapeutic Revolution. Essays in the Social History of American Medicine (Philadelphia: University of Pennsylvania Press, 1979), 3-25. Harry Dowling, Fighting Infection. Conquests of the Twentieth Century (Cambridge: Harvard University Press, 1977) provides a useful overview of therapeutic practices and advances for infectious diseases. Michael Bliss's otherwise excellent monograph on insulin, cited (n. 1) above, is weakest when it comes to putting the development, introduction and commercialization of insulin into comparative perspective.

Banting and Best's insulin and few recoveries as dramatic as Elizabeth Hughes' resurrection. However puny and inconsequential these weapons seem in retrospect, contemporary physicians found themselves armed for the first time with the means to treat even terminally ill patients with success.<sup>3</sup>

The majority of physicians accepted the laboratory's bounty without question. For the profession's scientific leadership, however, the prospect of potent and effective drugs created a novel intellectual and political problem.<sup>4</sup> Whereas earlier generations sought to dampen

<sup>3</sup> For two perspectives on these "advances" from the vantage of the late 1970's, see Lewis Thomas, The Youngest Science. Notes of a Medicine Watcher (New York: Viking Press, 1983), 12-18, 40-43 and Paul Beeson, "Changes in Medical Therapy During the Past Half Century," Medicine 59 (1980), 79-99. Both of these eminent clinical scientists view this period as a therapeutic dark ages redeemed from error by the scientific accomplishments of a later epoch. Their Whiggisness is not shared by sceptics like Thomas McKeown, who are equally doubtful of the significance of medicine's latter day therapeutic accomplishments, as measured by their demographic consequences. See McKeown's The Role of Modern Medicine. Dream or Mirage? (Princeton: Princeton University Press, 1980).

<sup>4</sup> By the middle of the nineteenth century, European inspired clinicians in the United States, in an effort to place therapeutics on an empirical basis, were advocating experimental clinical evaluations of medical treatments. In their skepticism about the majority of available therapies and in their advocacy of experimental approaches to testing therapeutic claims, these reformers anticipated later developments. Several features distinguish them from subsequent advocates of a "rational therapeutics." First, they expected therapeutic knowledge to be validated at the bedside and laboratory studies made little or no contribution to their perspective. Second, despite an effort to distinguish themselves from therapeutic nihilists, who doubted on a priori grounds the virtues of most treatments, these physicians acquired a reputation for therapeutic nihilism, i.e. for doubting the merits of most therapies. Thirdly, they represented a minority current within American medicine, and their attempts at organized therapeutic reform were largely unsuccessful.

For a general introduction to the views of these reformers, and some examples of their work on specific therapies, see James H. Cassedy, American Medicine and Statistical Thinking, 1800-1860 (Cambridge: Harvard University Press, 1984), 60-91, esp. 73-77; John Harley Warner provides a thoughtful account of the French influence on American physicians,

the unwarranted enthusiasm of laymen and physicians for useless or inadequate nostrums, the new generation had the more difficult task of ensuring that only effective drugs were chosen from the diverse repertoire of products being touted. For this task, the therapeutic skepticism of an earlier era was unsuitable: "an overskeptical mind is as undesired as an overcredulous one."<sup>5</sup> Circumstances called for a new attitude to replace the "therapeutic nihilism" of 19th century skeptics: "rational therapeutics".

The notion of a rational therapeutics referred first to the use of therapeutic agents whose mechanisms of action were scientifically established prior to their introduction into clinical practice. A rational, as opposed to an empirical remedy, was one whose effects were demonstrable in the laboratory and ideally, one which acted on the cause, not the symptoms, of disease.<sup>6</sup> At the same time, rational

---

which stresses the importance of an empiricist epistemology to the latter, in "The Selective Transport of Medical Knowledge: Antebellum American Physicians and Parisian Medical Therapeutics," Bulletin of the History of Medicine 59 (Summer, 1985), 213-231. As described by Martin S. Pernick, the advocates of a "conservative medicine" come closest to the proponents of rational therapeutics in their desire to distinguish between the worthwhile and worthless in therapeutics. See his A Calculus of Suffering. Pain, Professionalism and Anesthesia in Nineteenth Century America (New York: Columbia University Press, 1985), esp. 21-31, 93-124.

<sup>5</sup> James P. Herrick, "The Relation of the Association to Therapeutics," Transactions of the Association of American Physicians 38 (1923), 5. See also Morris Fishbein, "Scientific Therapy and Pharmaceutical Research," JAMA 84 (May 16, 1925), 1518.

<sup>6</sup> The emphasis on the specificity of cure may be found in S.J. Meltzer, "The Present Status of Therapeutics and the Significance of Salvarsan," JAMA 56 (June 10, 1911), 1709-1711. Such emphasis could lead to the neglect of perfectly serviceable drugs, as was the case with aspirin in treating arthritis, according to Jane and James Goodwin. Aspirin's value was doubted because of the belief that no analgesic could possibly affect the disease's root causes, which were

therapeutics referred to the conduct of clinical practice. In rational practice, the dosage and uses of a drug were in accordance with what was known about its' pharmacological activity and effects. It made no sense to use even an effective drug at doses which were sub-therapeutic or under clinical circumstances where it could not possibly benefit.<sup>7</sup>

The movement for a rational therapeutics accordingly had two aspects. On the one hand, it encompassed efforts to control the introduction and promotion of new drugs by their manufacturers. On the other, it referred to attempts to inculcate a scientific and critical attitude toward therapeutics in the medical profession itself. The two reforms were inter-dependent. Restricting the number of drugs to a handful of proven remedies would lessen the confusion of practitioners and thereby contribute to more "rational" therapeutic practice. At the same time, the success of efforts to reform the industry presumed the

---

thought to be infectious, dietary or metabolic. See their "Failure to Recognize Efficacious Treatments: A History of Salicylate Therapy in Rheumatoid Arthritis," Perspectives in Biology and Medicine 25 (Autumn, 1981), 78-92. Nonetheless, it was generally accepted that symptomatic treatment was "rational," when no specific cure existed, and when a demonstrable experimental basis for symptomatic treatment had been established. See J.H. Means and A.L. Barach, "The Symptomatic Treatment of Pneumonia," JAMA 77 (October 15, 1921), 1217.

<sup>7</sup> Both aspects are clearly identified in Robert A. Hatcher, "The Duty of the Medical Profession Toward the Council on Pharmacy and Chemistry," JAMA 67 (November 14, 1916), 1340; Jacob Diner, "Rational Drug Therapy," JAMA 72 (Jan 25, 1919), 264-5; L.G. Rowntree, "The Role and Development of Drug Therapy," JAMA 77 (October 1, 1921), 1061-1065; Harry Gold, "Recent Advances in Drug Therapy," International Clinics (December, 1930), 89-90; and Theodore Koppanyi, "Applied Pharmacodynamics: Rational Therapeutics," Medical Annals of the District of Columbia 4 (May 1935), 127-132. Pharmacologically oriented writers were somewhat more inclined to stress the importance of understanding a drug's metabolic fate; some reformers were sceptical that the average practitioner could or would acquire "a full scientific appreciation of the mode of [a] drug's action." Fishbein, "Scientific Therapy and Pharmaceutical Research," (n. 5) 1519.



support of a reformed profession, capable of recognizing the merits of using only carefully screened products. The present chapter describes the efforts of the American Medical Association's Council on Pharmacy and Chemistry to inculcate a critical attitude towards the selection and use of new drugs among practicing physicians in the early decades of the twentieth century. The most visible sign of medical science's accomplishments, the new therapeutics held the potential either to discredit medicine or make it praiseworthy. As representatives of the profession's scientific leadership, members of the Council assumed the task of ensuring that the therapeutic potential of new laboratory products would be realized at the bedside. The Council neglected none of the available tools for reform—legislation, publicity, education—in their work, but they relied most on the expectation that physicians would seek to emulate the attitudes and judgments of medicine's scientific elite.

In seeking to change their colleagues' behavior, these reformers relied on intellectual means; first and foremost, they sought to change the way physicians thought about therapeutics. In particular, reformers hoped that physicians would adopt their own, experimental, attitude towards therapeutic claims. Their chosen instrument for reformation, accordingly, was the idea of an experiment. Among active researchers, appeals to experimentation were clear enough. Experimental studies were the ideal means for producing and evaluating beliefs about the causes and treatment of disease. Most physicians, however, were not engaged in research. Reformers nonetheless expected that practicing physicians

would adopt the experimenter's methodical and provisional approach towards such knowledge. Depending upon the context, appeals to experimentation might refer to the production of new knowledge, a task for experts alone, or to an attitude about that knowledge. In either instance, an experimental approach to therapeutics was intended to provide the basis for reforming therapeutic practice. This ambiguity in reformers' idea of an experiment, and its consequences for subsequent efforts at improving therapeutic practice are the subjects of this chapter.

#### Reforming the Profession

From its founding in 1847, the leaders of the American Medical Association (AMA) had been interested in therapeutic reform. The promotion of patent medicines, the appropriate use of therapeutic innovations, and the elimination of so-called sectarian schools of therapeutics, such as homeopathy, were all objects of AMA concern. The intellectual means subsequently employed to arrive at such therapeutic assessments were already available to these nineteenth century reformers. But the fledgling organization lacked either the political resources or the cultural authority to direct therapeutic practices in a divided profession. To steer clear of divisions of opinion within its own membership, the AMA refrained from taking an official position on the merits of nineteenth century innovations such as anesthesia. Although proposals for reforming therapeutics surfaced periodically during the

remainder of the century, it was not until the first decade of the twentieth century that the organization began to act.<sup>8</sup>

On February 3, 1905, the AMA established the Council on Pharmacy and Chemistry.<sup>9</sup> Composed of individuals selected for their interest and eminence in pharmacological research, the Council placed its greatest emphasis on regulating drugs, and the representations made on their behalf. Of ten initial members, seven were professors of pharmacology, pharmacy or chemistry, while two worked in government laboratories responsible for the evaluation of vaccines and drugs. In a period when influence in the medical community was still measured

<sup>8</sup> For accounts of the AMA's nineteenth century activities, see Paul Starr, The Social Transformation of American Medicine (New York: Basic Books, Inc., 1982), 127-129 and Pernick, A Calculus of Suffering (n. 4), 26-29, 41, 69-70. Both accounts agree that the AMA failed to do much about regulating therapeutics before the twentieth century, but Starr places slightly greater emphasis on the AMA's inability to command the necessary political resources to enforce their views on the profession as a whole and society at large, while Pernick emphasizes divisions of opinion within the organization itself which led to a lack of political will. In explaining the AMA's about-face in these matters after 1900, Starr accordingly stresses the support of groups outside the AMA and the elimination or cooptation of competing authorities within the profession as the key to subsequent success.

<sup>9</sup> A creation of the Board of Trustees, the Council represented the views of the profession's national leadership rather than the local medical interests whose base was in the AMA's newly organized House of Delegates. On the structure of the AMA, see Starr, Social Transformation (n. 8), 109-110. On the decision to create the Council, see James G. Burrow, "The Prescription-Drug Policies of the American Medical Association in the Progressive Era," in John Blake, ed. Safekeeping the Public (Baltimore: The Johns Hopkins University Press, 1970), 112-122. Burrow's account is far and away the best treatment of the early Council. For additional information, see Austin Smith, "The Council on Pharmacy and Chemistry and the Chemical Laboratory," in Morris Fishbein, ed. A History of the American Medical Association 1847 to 1947 (Philadelphia: W.B. Saunders Company, 1947), 865-886 and Harry Dowling, Medicines For Man. The Development, Regulation and Use of Prescription Drugs (New York: Alfred A. Knopf, 1970), 155-185. Both these accounts are better at reporting the early Council's programmatic intentions than at analyzing its accomplishments.

first by local, and then by national reputation, the Council's composition placed it at a disadvantage. Although six of the ten held medical degrees, none was regarded as engaged in the practice of medicine, and only one, George H. Simmons, editor of the Journal of the American Medical Association (JAMA) was widely known in the profession.<sup>10</sup>

To compensate for its lack of reputation, Simmons provided the Council with access to the pages of JAMA. There a regular column, aptly named "The Propaganda for Reform," provided "intelligent physicians" with an "unprejudiced examination" of information concerning items "worthy of [their] patronage." Any manufacturer seeking recognition of his product by the Council, and thereby the profession, was required to make its composition known, to justify the therapeutic claims made on its behalf and to avoid exaggerated or misleading advertising. Justly deserving products were accorded recognition in

---

<sup>10</sup> Arthur Cushny (M.D.), Lewis Diehl (Ph.M.), C.S.N. Hallberg (M.D., Ph.G.), Robert A. Hatcher (M.D., Ph.G.), W.A. Puckner (Ph.G.), J.O. Schlotterbeck (Ph.G., Ph.D.) and Torald Sollman (M.D.) were professors of pharmacology, pharmacy or chemistry. Two additional members, Lyman F. Kebler (Ph.G., M.S., M.D.) and M.I. Wilbert (Ph.M) worked for the Department of Agriculture and the U.S. Hygienic Laboratory, respectively. These men were soon joined by Samuel Sadtler (Ph.D), J.H. Long (M.S., Sc.D.), Julius Stieglitz (Ph.D), F.G. Novy (M.D., Sc.D.) and Harvey W. Wiley (M.D., Ph.D.). Sadtler, Long and Stieglitz were professors of chemistry; Novy was professor of bacteriology; and Wiley, physician and chemist, was head of the Bureau of Chemistry in the U.S. Department of Agriculture, responsible for administering the 1906 Pure Food and Drug Act. Within the academic community, Novy, Sollman and Stieglitz were almost certainly well known, and Cushny's reputation was not national but international. What any of these names might mean to the practitioner in Louisville, Kentucky, is, however, another matter. For a partial list of early membership see Smith, "The Council," (n. 9) 866-869, from which the above biographical information is taken.

an annual compilation issued by the Council, New and Nonofficial Remedies.<sup>11</sup>

Like many other Progressive era reforms, the Council's program was an attempt to moderate the excesses of capitalism: "Honest advertising is a necessary feature of civilization—at least it is not an unmitigated evil. Fraudulent advertisements are one of the curses of civilization."<sup>12</sup> Reforming capitalism, in this instance, was but the means to an end—reforming medicine. However appropriate in other walks of life, in therapeutics the profit motive exercised a baneful influence. Seeking to benefit from the enhanced aura of medical science, drug manufacturers couched their claims in the pseudo-scientific jargon of the day. Increased expenditures in drug advertising, unchecked since the Civil War, took what reformers regarded as an ominous turn around 1900, as firms concentrated their propaganda on the profession itself.<sup>13</sup>

---

<sup>11</sup> For a list of the Council's initial criteria see Council on Pharmacy and Chemistry, "Preliminary Announcement," JAMA 44 (March 44, 1905), 720-721. The first edition of NNR was in 1907.

<sup>12</sup> Torald Sollman, The Broader Aims of the Council on Pharmacy of the American Medical Association (Chicago: AMA, 1908), 21.

<sup>13</sup> On the growth of advertising for patent medicines and the growing size and consolidation of the drug industry, see James Harvey Young, The Medical Messiahs. A Social History of Health Quackery in Twentieth Century America (Princeton: Princeton University Press, 1967), 21-25. Historians have recently begun to extend Young's observation (26) that drug companies reoriented both the rhetoric and the immediate targets of their advertising in concentrating their marketing efforts on physicians. See Rima D. Apple, "'To Be Used Only Under the Direction of a Physician': Commercial Infant Feeding and Medical Practice," Bulletin of the History of Medicine 54 (Fall 1980), 402-417; and Starr, Social Transformation, (n. 8) 132-134. Starr's account, in particular, ignores Young's caveat that direct advertising to physicians did not necessarily imply a shift in the actual channels of drug distribution. Longitudinal data on the proportion of drug sales marketed directly to physicians would be extremely useful in distinguishing between the claim that the medical profession aspired to control drug distribution

Unchecked, commercialism threatened to undermine the scientific basis for a "rational therapeutics", "debauching our medical journals" and "tainting our textbooks."<sup>14</sup> The Council's reports were a means for science to fight back.

Their initial targets were the manufacturers of "secret" remedies: variations on standard compounds whose uselessness (or potential toxicity) was concealed behind names invoking occult powers (Bioplasm) or suggesting cut-rate narcotics (Sal-codeia).<sup>15</sup> Forcing manufacturers to reveal the contents of these drugs would enable physicians to discard redundant or impotent drugs. But the truth-telling principle applied equally to medicines "that have merit and that would be used even if the simple unvarnished truth were told about them". The campaign against secret remedies was merely a "preliminary to a larger and broader aim, the general reformation of what is debased and debasing in the present status of therapy."<sup>16</sup>

---

and the claim that they succeeded in doing so much before 1950.

For a fuller account of the incorporation of the new medical science into the tactics and objectives of pharmaceutical firms, see Jonathan Liebenau, Medical Science and Medical Industry, 1890-1929. A Study of Pharmaceutical Manufacturing in Philadelphia, esp. 104-114 and 202-222. Ph.D. dissertation. University of Pennsylvania, 1981. Given the concentration of pioneering firms in the middle Atlantic region, Liebenau's thesis is less parochial than it might seem from the subtitle.

<sup>14</sup> George H. Simmons, "The Commercial Domination of Therapeutics and the Movement for Reform," JAMA 48 (May 18, 1907), 1645.

<sup>15</sup> "The Secret Nostrum Vs. the Ethical Proprietary," JAMA 44 (March 4, 1944). On the ingenuity of cure peddlars in naming their products, see Young, Medical Messiahs (n. 13), passim.

<sup>16</sup> Sollman, Broader Aims of the Council (n. 12), 3. See also the discussion by Julius Stieglitz of chemical compounds which were potentially worthwhile, but for which exaggerated therapeutic claims were made and whose quality varied from lot to lot and manufacturer to manufacturer. "The Problem of the Synthetic Chemical Compound," JAMA 46

For the products with which the AMA Council was concerned, the physician was the ultimate consumer. A rational therapeutics required not only a scientific assessment of the accomplishments and limitations of specific drugs, but practitioners capable of recognizing and acknowledging those limitations in their practice. Even the carefully screened compounds appearing in the official pharmacopeia's did not "constitute an advance" when "used as uncritically, with the same exaggerated expectations, as are the proprietary articles...." <sup>17</sup>

The Council benefited from the support of "the better element of the medical profession," which endorsed its program. But the Council's authority with manufacturers rested with its ability to reshape the attitudes of the profession at large.<sup>18</sup> The reformers' program of publicity made sense only if the medical public acted on the information provided:

In whatever we look at it, the responsibility always returns to the individual physician—he is the man who carries the arms. It comes to him not so much on the day when he votes for a resolution, but every day, every time when he writes a prescription. Whenever he picks up a prescription blank, he is not only directing the treatment of his patient; he is also directing the proprietary business, in all its ramifications; and beyond this, he is directing the future of therapeutics. With each prescription, he renders a decision whether truth or falsehood shall prevail; whether therapeutics shall be scientific or unscientific; whether the abuse of indiscriminate self-medication shall continue or not. <sup>19</sup>

---

(May 5, 1906), 1341-1342.

<sup>17</sup> Sollman, Broader Aims of the Council (n. 12), 33.

<sup>18</sup> David L. Edsall, "The Work of the Council on Pharmacy and Chemistry," JAMA 55 (November 12, 1910), 701.

<sup>19</sup> Sollman, Broader Aims of the Council (n. 12), 46.

As with other AMA sponsored reforms, in physician licensure and medical education, new legislation abetted the movement for a rational therapeutics. Federal officials responsible for administering the Biologics Act of 1902 and the Pure Food and Drug Act of 1906 held seats on the Council and worked closely with it.<sup>20</sup> Likewise, the Council accepted the support of the consumer movement in their efforts to

---

<sup>20</sup> Harvey Wiley, the crusading chief of the federal Bureau of Chemistry, responsible for administering the Pure Food and Drug Act of 1906, and John Anderson and George McCoy, responsible for testing vaccines and serums under the federal Biologics Act of 1902, were each Council members. [See Smith, "The Council," (n. 9) 868-869; for more discussion of the links between therapeutic reformers and federal officials, see chapter two.]

Historians in recent years have generally cast a suspicious eye on the connections between the AMA and federal authorities, reading these ties as one more evidence of the profession's attempt to enhance their monopoly power over the practice of medicine. Such readings do not take into account either the continued reservations of the AMA toward its lay allies, or the extent to which private organizations more generally played a role in shaping and administering public policy. The appointment of officials like Wiley or McCoy to the AMA's Council on Pharmacy and Chemistry should be no more surprising than the willingness of private philanthropic foundations to subsidize the salaries of government officials and programs. However one interprets the miscegenation of "public" and "private" spheres in this period, historians of medicine might be advised to cease and desist in singling the AMA's relations for special interpretation. [On the subrogation of federal actions by private philanthropy, see Barry D. Karl and Stanley N. Katz, "The American Private Philanthropic Foundations and the Public Sphere 189-1930," *Minerva* 19 (Summer, 1981), esp. 238-243, 256-263. In an earlier article, Karl documents the private role in financing and staffing federal agencies, boards and commissions under the administrations of Theodore Roosevelt, Woodrow Wilson and Herbert Hoover. See his "Philanthropy, Policy Planning and the Bureaucratization of the Democratic Ideal," *Daedalus* (1976), esp. 138, 146-147.] More to the point, in the present instance, is that efforts of therapeutic reformers were directed as much inwards, towards self-regulation, as towards altering the behavior of either consumers or producers of drugs with the aid of legislation.



chasten industry, and welcomed the cooperation of progressive drug firms.<sup>21</sup> But the Council's initial program depended less on legislative and institutional remedies than on the moral conversion of the individual physician.<sup>22</sup> The physician who benefited from the enhanced status of medicine had an obligation to exercise his new powers responsibly and intelligently.<sup>23</sup>

Ultimately, it was the medical profession and not the pharmaceutical trade which the reformers found lacking:

We cannot blame manufacturing chemists for finding new things or advertising them as cleverly as possible. That they and the nostrum vendor are surprisingly successful in selling their wares is largely our fault.<sup>24</sup>

If physicians were equipped to make competent judgments on the merits of new drugs, the desire of each manufacturer to produce and sell its own unique compounds would present few problems: "Unfortunately, however, the physician's training is likely to be such that he can not

<sup>21</sup> For the response of various firms to the Council's initial announcement, see "Expressions on the Announcement," JAMA 44 (March 25, 1905), 971. On the support of consumer advocates for the campaign against quackery, see Young, Medical Messiahs, (n. 13) 29-32. Paul Starr reports that the AMA distributed over 150,000 copies of "The Great American [drug] Fraud," a muckraking account of the drug industry published by Collier's Weekly. See Social Transformation (n. 8), 31.

<sup>22</sup> On the Council's reservations about the 1906 drug law and its recognition of the need to go beyond either federal or state legislation to accomplish its aims, see especially Burrow, "Prescription Drug Policies," (n. 8) 115-117.

<sup>23</sup> Hatcher, "The Duty of the Medical Profession," (n. 7) 1339, 1341; Edsall, "The Work of the Council," (n. 18) 1703-1704.

<sup>24</sup> N.S. Davis, "Effect of Proprietary Literature on Medical Men," JAMA 46 (May 5, 1906), 1339. See also Simmons, "The Commercial Domination of Therapeutics," (n. 14) 1646.

distinguish the rank fraud from the efficacious remedy, honestly made and sold." 25

At the time the Council was formed, the teaching of therapeutics was nearly as bad off as its practice. Only a handful of schools provided instruction beyond materia medica, the learning by rote of a repertoire of medications and their uses. The new therapeutics altered somewhat the drugs covered, but the atheoretical, didactic character of the instruction remained unchanged.<sup>26</sup> As might be expected, several of the Council's founding members were extensively involved in efforts at educational reform.<sup>27</sup> But educating the medical public was more than a matter of changing classroom instruction. Members of the Council expected no less than an intellectual transformation in the profession: cultivation of an experimental orientation toward therapeutic practice.

---

<sup>25</sup> W.A. Puckner, "The Nostrum From the Point of View of the Pharmacist," JAMA 46 (May 5, 1906), 1340.

<sup>26</sup> On the teaching of pharmacology, see David L. Cowen, "Materia Medica and Pharmacology," in Ronald L. Numbers, The Education of American Physicians. Historical Essays (Berkeley: University of California Press, 1980), 105-110. Cowen's discussion of the period from 1900-1930 is extremely cursory, however.

<sup>27</sup> Three of the nine scientists singled out for mention by Cowen were founding members of the Council (Sollman, Hatcher and Cushny) while several others (S.J. Meltzer, A.N. Richards, John J. Abel) were prominent in efforts to establish the practice of therapeutics on a rational basis. Cowen, "Materia Medica," ibid., 112.

The Idea of An Experiment

Considered in itself, the experimental method is nothing but reasoning by whose help we methodically submit our ideas to experience—the experience of facts.<sup>28</sup>

In the last quarter of the 19th century, medicine experienced a scientific revolution. The revolution occurred not in the practice of medicine but in what came to be known as the medical sciences: bacteriology, physiology, physiological chemistry and pharmacology, each of which in turn acquired the means to produce and manipulate in the laboratory the phenomena of disease.<sup>29</sup> To a small community of elite physicians, European trained and inspired, the future of medicine would be shaped in the laboratory. But only a few of these developments proved of immediate consequence for medical practice. Their practical yield was less consequential than their ideological effect.<sup>30</sup> By 1900,

---

<sup>28</sup> Claude Bernard, An Introduction to the Study of Experimental Medicine (New York: Dover Publications, 1957), 2.

<sup>29</sup> For a charming analysis of an exemplary episode, see Bruno Latour's account of how the technological conquest of the anthrax bacillus secured an ideological victory for the new Pasteurian laboratory science. "Give Me a Laboratory and I Will Raise the World," in Karin D. Knorr-Cetina and Michael Mulkay, Science Observed. Perspectives on the Social Study of Science (London: Sage Publications, Ltd, 1983), 141-170.

<sup>30</sup> Practising physicians, for example, responded much more enthusiastically to the laboratory production of diphtheria anti-toxin, a treatment, than they had a few years earlier to the use of laboratory diagnostic tests in establishing the presence of the disease. Barbara Rosenkrantz, "Cart Before Horse: Theory, Practice and Professional Image in American Public Health, 1870-1920," Journal of the History of Medicine (January, 1974), 69-71. See also Gerald L. Geison, "Divided We Stand: Physiologists and Clinicians in the American Context," in Vogel and Rosenberg, The Therapeutic Revolution (n. 2), 66-90. The more important contributions were frequently not the breakthrough discoveries on which historians of science are prone to focus attention but the more mundane accomplishments which benefited daily practice, such as William Welch's careful experiments on the effects of different techniques of wound incisions and suturing on infection rates.

a new generation of medical researchers and educators were converts to the gospel of experimental truth: "Experiment is the only certain way of progress." 31

First and foremost, the belief in experimental method denoted confidence in the virtues of a specific cast of mind, an intelligence capable of clear reasoning and unprejudiced judgment. The application of experimental reasoning to the facts of the clinic was no less a scientific accomplishment than a laboratory experiment:

The patient's history, conditions, symptoms, form [the physician's] data. Thereupon he, too, frames his working hypothesis, now called a diagnosis. It suggests a line of action. Is he right or wrong? Has he actually amassed all the significant facts? Does his working hypothesis properly put them together? The sick man's progress is nature's comment and criticism. [....] The progress of science and the scientific or intelligent practice of medicine employ, therefore, exactly the same technique. 32

Experimental knowledge, in this sense, was within the reach of all physicians: "The essential conditions are ... in the mental equipment of the investigator. What is needed is, first of all, a frank dislike for cant...." 33 Sound research was as much a matter of "attitude" as of "technic": the novice investigator must learn to accept agnosticism "in regard to what is not proved." 34 Readiness to believe in the merits of new treatments stemmed from a failure of intellect and character: "The

---

31 Sollman, Broader Aims of the Council (n. 12), 47.

32 Abraham Flexner, Medical Education in the United States and Canada (New York: Carnegie Foundation for the Advancement of Teaching, 1910), 55.

33 Torald Sollman, "Experimental Therapeutics," JAMA 58 (Jan 27, 1912), 244.

34 Torald Sollman, "The Evaluation of Therapeutic Remedies in the Hospital," JAMA 94 (April 26, 1930), 1279-1280.

man who makes an empirical discovery, who believes it to be a fact, has the moral obligation to establish by exact observation that it is a fact and not just a figment of his imagination."<sup>35</sup>

Considered as an attitude, experimentalism was potentially accessible to all. Considered as a technology, however, experiments were the domain of the few—experts with the training, intelligence and resources to produce and interpret their findings.<sup>36</sup> The dual character of experimental knowledge was most acutely felt in areas closest to the actual conduct of medical practice, such as therapeutics. By its very nature, therapeutic experimentation was an esoteric subject. Experience alone provided knowledge about the vagaries of specific diseases necessary to complement more general training in research techniques.<sup>37</sup> The expertise required to design an experiment on treating heart failure was not the same as that needed to judge the worth of a new bactericidal compound.

The value of an experiment depended on the degree to which an experimenter anticipated potential sources of error: "It is the purpose of an experimental science to replace accident by design."<sup>38</sup> Therapeutic reformers took their counsel from Francis Bacon, who advised men of science to be ever on the alert for self-deception. The laboratory

<sup>35</sup> Sollman, "Experimental Therapeutics," (n. 33), 242.

<sup>36</sup> Frederick P. Gay, "Specialization and Research in the Medical Sciences," Science 45 (January 12, 1917), 25-33.

<sup>37</sup> Harry Gold, "Recent Advances in Drug Therapy," International Clinics (December, 1930), 97-102; Walter Houston, "Standards in Therapeutics," International Clinics (June, 1933), 191.

<sup>38</sup> Frederick P. Gay, "Immunology. A Medical Science Developed Through Animal Experimentation," JAMA 56 (February 25, 1911), 579.

sciences provided the model for their work: "... the laboratory worker plans a series of experiments, and he endeavours to eliminate errors by repetition, and by controlling the various factors which might influence his results." <sup>39</sup> Experimentation was an aid to methodical self-doubt.

Clinical evaluation, no less than pre-clinical testing, needed to be conducted according "to the canons of other scientific experimentation. Otherwise, its scientific usefulness is nil, and even its practical usefulness is, at best, doubtful."<sup>40</sup> Placing therapeutics on an experimental basis meant more than subjecting a series of patients to treatment:

Experiments may be framed so loosely, the observations may be so superficial, the analysis of results so careless, the deductions so illogical, that the experiment has no permanent value—it is not an experiment in the precise sense of the word.<sup>41</sup>

Like the laboratory study, the clinical investigation must be planned and regulated: "The results in ten well controlled cases are of more value than the haphazard impressions from a thousand cases."<sup>42</sup>

A "well-controlled" experiment was not necessarily one with an untreated series of patients, but one in which an knowledgeable and experienced investigator had anticipated the "multitude of factors" which might effect the outcome: patient selection, dosage, laboratory tech-

<sup>39</sup> Albion Walter Hewlett, "The Cooperation Between Pharmacology and Therapeutics," JAMA 59 (October 6, 1917), 1123.

<sup>40</sup> Sollman, "Experimental Therapeutics," (n. 33), 244.

<sup>41</sup> Ibid, 242.

<sup>42</sup> Rowntree, "The Role and Development of Drug Therapy," (n. 7), 1064. See also Torald Sollman's remarks in "Therapeutic Research," JAMA 58 (May 4, 1912), 1390.

nique, and the natural history of the disease. Employing a series of untreated cases was one aid to interpreting experimental results, but not necessarily the only means, or the best. Where the use of untreated controls led investigators to neglect "the individuality of cases," they were of "limited or doubtful" value.<sup>43</sup>

Replicating the mastery of a laboratory experiment was difficult to accomplish in the physician's office.<sup>44</sup> The hospital accordingly formed an essential adjunct to the conduct of "well-controlled" experiments:

In a case of hypertension it is not usually feasible—though it is quite justifiable—to have a patient make ten or more office visits without any treatment in order to ascertain the spontaneous variations in that patients' blood-pressure, yet unless that is done one can rarely draw any conclusions about the action of a drug upon the blood-pressure with any assurance that the change was not entirely independent of the drug.<sup>45</sup>

Hospitals afforded the opportunity not only to record observations but, if needed, to regulate the behavior of experimental subjects. Where the activity of patients constituted an aspect of the experimental conditions, this too could be "controlled".<sup>46</sup>

<sup>43</sup> For advocacy of using untreated cases, see Joseph L. Miller, "How May the Science of Therapeutics Be Advanced?" JAMA 59 (Sept 21 1912), 915. For some of the objections to untreated controls, see Torald Sollman, "The Crucial Test of Therapeutic Evidence," JAMA 69 (July 21, 1917), 199. The majority of references to "controlled" studies refer to more than use of an untreated control series, even where that practice is advocated.

<sup>44</sup> Sollman, "Evaluation of Therapeutic Remedies in the Hospital," (n. 34), 1278.

<sup>45</sup> Gold, "Recent Advances in Drug Therapy," (n. 37), 96. See also John F. Anderson, "Some Unhealthy Tendencies in Therapeutics," JAMA 63 (July 14, 1914), 2.

<sup>46</sup> Sollman, "Evaluation of Therapeutic Remedies in the Hospital," (n. 34), 1279.

The special requirements of facilities, equipment and expertise placed therapeutic investigation beyond the means of most physicians. But controlled experiments remained a standard by which other evidence was found wanting. Advocates of an experimental philosophy saw no contradiction in believing that experimental knowledge was within the reach of all physicians while maintaining that proper therapeutic experimentation demanded special talents and resources. If the average physician was no longer able to participate in the production of experimental truth, he still belonged to a community bound to be guided by that truth in its actions. Even the physician who could not produce new therapeutic knowledge was obligated to accept the authority of those who did.

The Work of the Council: Knowledge and Virtue

Nearly all abuses arise because  
someone profits thereby. <sup>47</sup>

In theory, establishing a rational therapeutics meant providing an experimentally based chain of evidence linking laboratory and bedside. In the practical work of the Council many of the necessary links were missing and others were weaker than desired. As a consequence, the Council's deliberations reflect a curious mixture of judgments about the quality of evidence and opinions about the motives of the men who provided it. Where evidence of therapeutic value was equivocal, evidence of character aided the decision. The products of firms which

---

<sup>47</sup> Robert Hatcher to Torald Sollman, November 25, 1936. Torald Sollman papers. Archives, Cleveland Health Sciences Library, Cleveland, Ohio. [Hereafter Sollman papers].



had proven reliable in the past were scrutinized less carefully than those of habitual offenders.<sup>48</sup> The most trustworthy data were offered by those who lacked economic motives entirely: the "high-minded men" and "institutions" of clinical science.<sup>49</sup>

Where secure experimental evidence existed, the Council had few difficulties in arriving at a decision. For many of their assessments, the chemistry laboratory proved adequate. Drugs whose principal ingredients were found to be inert or whose active ingredients varied wildly from lot to lot could be readily dismissed.<sup>50</sup> The intrepid chemist could even discover honesty, or at least, its absence:

"...misstatements as to therapeutic efficiency...may be due to ignorance or excusable prejudice; but misstatements as to composi-

<sup>48</sup> As David Edsall noted shortly after his appointment to the Council, distinguishing between the trustworthy and the untrustworthy firm was no simple matter, and nothing keeps a "reputable" firm from hiring disreputable individuals or itself taking advantages of the Council's lack of scrutiny. See Council on Pharmacy and Chemistry, Bulletin 7 (April 2, 1908), 152-153. I am extremely grateful to the American Medical Association for permitting me to consult their collection of the Bulletins, which are the minutes of the Council's weekly meetings, and to Ms. Terry Austin and Ms. Marguerite Falucco for assisting me with the collection. [Hereafter cited as Bulletin.]

<sup>49</sup> The term is Torald Sollman's, from "The Evaluation of Therapeutic Remedies in the Hospital," (n. 34), 1279.

<sup>50</sup> Torald Sollman, "Yesterday, Today and Tomorrow. The Activities of the Council on Pharmacy and Chemistry," JAMA 61 (July 12, 1912), 5-6; Paul Nicholas Leech, "Chemistry in the Service of Pharmaceutical Medicine," JAMA 85 (July 11, 1915), 139-140. Another object of Council scorn was the so-called irrational mixture, a proprietary combination of drugs which contained either too little of one drug or too much of another to obtain the desired effects. On the whole, the members of the Council favored the view of pharmacologists that when multiple drugs were to be used, the physician ought to prescribe them individually, but the Council nonetheless had difficulties in setting a blanket policy on such mixtures. For discussion, see Bulletin 1 (March 9, 1905), 15; Bulletin 1 (March 16, 1905), 24, 26; Bulletin 1 (March 23, 1905), 31-33.

tion cannot be due to these causes, but only to downright dishonesty and intentional fraud." 51

If it had been possible to resolve all therapeutic questions in the chemical laboratory, the members of the Council might have been content. But manufacturers continued to pose claims which that laboratory could not address. Take the case of glandular extracts, a hotbed of innovative activity in the early 1900's. A package might well consist of what the manufacturer said it did: red bone marrow extract. That claim could be tested in the chemistry laboratory. But did red bone marrow do what physicians thought it might? And if the Council approved red bone marrow extract, what about ovarian extract, parotid gland extracts and a small platoon of other products waiting in the wings? 52 For these questions, the chemical laboratory would not suffice.

Animal experimentation provided the next step in judging a drug's merit. By 1900, the practice of evaluating vaccines and anti-infective agents on experimentally infected animals was already well established. 53 The practice soon generalized to other kinds of drugs. The animal study was the exemplar of the well-controlled experiment. For animals, unlike sick patients, conditions could be varied at will: the selection of subjects, the range of doses and even the intensity of pathology were

---

51 Bulletin 3 (March 20, 1906), 99. On the distinction between the motives of industry and the honest, but sometimes mistaken, professional, see also Rowntree, "The Role and Development of Drug Therapy," (n. 7), 1064.

52 In the case of red bone marrow, the Council recommended its approval on the grounds that it was "honestly exploited." Bulletin 4 (August 9, 1906), 97; Bulletin 4 (August 23, 1906), 127.

53 Horatio C. Wood, Jr., "The Limits and Purpose of Bio-Assay," Transactions, Section on Pharmacology & Therapeutics. A.M.A. (1912), 309-310.

manipulable on the experimenter's wish. Research on animals provided rapid and extensive information about a drug's safety, or lack thereof. For compounds which eventually proved useful, such studies would aid in determining the range of effective dosages and circumstances under which the drug might be used.<sup>54</sup> Their essential virtue, however, was in providing experimenters with the means to guard against the pitfalls of clinical experimentation: mice were far less likely than humans to recover solely on the basis of their keepers' kind attentions and encouraging wishes.<sup>55</sup>

However essential a tool, animal experimentation, too, had its limitations. Some drugs presented to the Council could not be reliably tested on animals.<sup>56</sup> A growing awareness of the differences between laboratory animals and humans made it necessary to interpret data from such studies with caution.<sup>57</sup> For clinicians, as well as for an older generation of pharmacologists, the testing of drugs on humans remained the ultimate court.<sup>58</sup> Their skepticism regarding the laboratory made

---

<sup>54</sup> On the advantages of animal studies, see Hewlett, "Cooperation Between Pharmacology and Therapeutics," (n. 39), 1123.

<sup>55</sup> "Report of the Council on Pharmacy and Chemistry. Cactus Grandiflorus," JAMA 54 (March 12, 1910), 889.

<sup>56</sup> See, for example, the discussion of the anti-pyretic agent A-S-Phen and Sollman's opinion that "animal experimentation is not very satisfactory for determining antipyretic action." Bulletin 7 (February 20, 1908), 86.

<sup>57</sup> George B. Wallace, "The Influence of Pathologic Conditions on the Actions of Drugs," JAMA 59 (September 9, 1912), 839-841.

<sup>58</sup> Horatio C. Wood, Jr. "Pharmacological Superstitions," JAMA 66 (April 6, 1916), 1068.

it difficult for the Council to dismiss a compound out of hand on the basis of animal studies.<sup>59</sup>

Clinical studies constituted the "weakest link" in therapeutic research.<sup>60</sup> Strengthening it meant keeping clinical evaluation out of the "average" practitioner's hands: "...the approximate value of a new drug should be determined exhaustively, on patients as well as animals, before it is advertised to the profession."<sup>61</sup> In the hopes of obtaining evaluation by experienced clinical investigators, promising drugs which passed the Council's initial laboratory screening were deemed "experimental". Once the drugs were granted "provisional" approval, however, physicians no longer regarded them as experimental. They were used by "incompetent observers" with little interest in producing evidence the Council might find acceptable.<sup>62</sup> Controlling the quality of experimental studies proved as difficult as controlling the quality of therapeutic practice more generally.

<sup>59</sup> On the necessity of conducting animal studies, despite the difficulties of extrapolating between species, see Sollman, "Experimental Therapeutics," (n. 33), 243-244; Hewlett, "Cooperation Between Pharmacology and Therapeutics," (n. 39), 1123-1124. Council member Robert Hatcher was responsible for demonstrating the variability of absorption and excretion of different drugs in different species. Rather than making Hatcher a skeptic about the role of animal studies, it made him insist on studying drug effects on an even wider range of species. See Robert A. Hatcher and Cary Eggleston, "Studies on the Absorption of Drugs," JAMA 63 (August 8, 1914), 469-473.

<sup>60</sup> Arthur Cushny to Torald Sollman, July 9, 1909, Sollman papers.

<sup>61</sup> Sollman, "Yesterday, Today and Tomorrow," (n. 50), 6.

<sup>62</sup> Bulletin 21 (June 17, 1915), 354; Bulletin 22 (July 29, 1915), 68. The problem of when to grant experimental approval continued to recur through the end of the decade: Bulletin 32 (February 23, 1920), 104; Bulletin 32 (March 16, 1920), 142.

Lacking access to patients, and already overburdened by their laboratory work, the Council's members were in no position to provide their own clinical data. From the outset, they acknowledged the need for a pool of capable clinical observers who would "be in a position" to help them in assessing therapeutic claims.<sup>63</sup> The difficulty was in finding "suitable" individuals. Initially, the Council doubted the possibility of finding intellectual qualified and politically sympathetic clinicians.<sup>64</sup> Eventually, they came to rely on the opinions of a small group of physicians expert on the disease or type of drug in question.<sup>65</sup>

Knowledgeable though they might be, these consultants rarely had data to guide them. The Council's minutes are replete with reports from referees who doubt the value of a compound but lack the evidence to prove it worthless. Where "honest differences" of opinion existed among responsible observers, the Council was powerless to act, even when several

---

<sup>63</sup> Bulletin 1 (June 15, 1905), 147.

<sup>64</sup> See the discussions in Bulletin 6 (October 24, 1907), 184; Bulletin 6 (October 31, 1907), 194; Bulletin 6 (November 7, 1907), 202; Bulletin 6 (November 14, 1907), 214-215. A staff of clinical consultants to advise the Council was selected early the following year, followed by the appointment of two clinical men, David Edsall and Joseph Capps, to the Council, at the suggestion of the AMA Board of Trustees. See Bulletin 7 (January 23, 1908), 46, 49-50 and Bulletin 7 (February 7, 1908).

<sup>65</sup> Unfortunately, the AMA has not preserved the working correspondence of the Council, making it difficult in most instances to identify particular individuals; the reports of referees in the minutes of the Council make it clear, however, that the Council relied on a small pool of unusually well informed observers who shared the values and assumptions of rational therapeutics espoused by Council members. The few cases where referees are identified support this contention: see the discussion of syphilis remedies [Bulletin 35 (May 24, 1922), 245-246] where the consultants are Harold Cole and Udo Wile, prominent syphilologists who subsequently organize one of the first major cooperative investigations of syphilis treatment. On Cole and Wile, see chapter three.

members had strong reservations about a compound.<sup>66</sup> At best, they could negotiate with the manufacturer to tone down the therapeutic claims somewhat.

Publicly, the Council members continue to propagandize for more careful clinical experimentation. Privately, they relied in large part on the opinions and recommendations of trusted colleagues. In the absence of reliable evidence, the Council turned to reliable men. The opinions of experts were an imperfect substitute for facts, but their allegiance to science meant that they would only err. They would not deceive:

In their entrance through scientific channels exaggerated therapeutic claims are made at times, as the result of a lack of critical judgment and adequate controls. But in their introduction through commercial channels, financial consideration and lack of true appreciation of the fundamental problems preclude unbiased observations.<sup>67</sup>

Belief in the integrity of scientists formed no part of the Council's official pronouncements on experimental method but where truth could not be established, a dedication to truth might suffice in the interim. The assurance of reformers in the contribution of experimental method to medicine was premised on a relentless confidence in the future: truth will out.

---

<sup>66</sup> Bulletin 21 (May 5, 1915), 298; Bulletin 22 (July 29, 1915), 431; Bulletin 31 (March 31, 1920), 141; Bulletin 32 (September 8, 1920), 159.

<sup>67</sup> Rowntree, "The Role and Development of Drug Therapy," (n. 7), 1064.

The Legacy of the Council. 1900-1930

It may seem strange that I should put forward three sentiments, namely, interest in an indefinite community, recognition of the possibility of this interest being made supreme, and hope in the unlimited continuance of intellectual activity, as indispensable requirements of logic.<sup>68</sup>

Historians have puzzled over the seemingly deluded faith in experts which characterized early 20th century reform movements. The aspirations of science so far exceeded its visible accomplishments that any confidence in the authority of science seems misplaced, a product either of conspiracy or false consciousness.<sup>69</sup> Certainly the fruits of the Council's labors through the 1920's seem meagre, especially in the area of therapeutic experimentation, where the majority of physicians continued to rely on their own limited clinical impressions or the uncontrolled testimonials of others. Yet for reformers, the shortcomings of actual therapeutic experimentation, and its failure to guide the conduct of practicing physicians were a cause for disappointment but not doubt.

Their unquenched optimism rested on the equation, curious to us but

---

<sup>68</sup> C.S. Peirce, "On the Doctrine of Chances, With Later Reflections," in Justus Butler, ed. Philosophical Writings of Peirce (New York: Dover Publications, Inc, 1955), 163.

<sup>69</sup> See, for example, E. Richard Brown, Rockefeller Medicine Men. Medicine and Capitalism in America (Berkeley: University of California Press, 1979). Paul Starr's account of the "cultural authority" of the physician, while warily avoiding either extreme of interpretation, likewise steers clear of examining the profession's scientific claims and conduct in any detail, while remaining dubious that such an examination could yield much of use. See Social Transformation (n. 8), esp. 134-140.

recognizable to their contemporaries, between science and morality.<sup>70</sup> The conduct of therapeutic experimentation was meant as an exercise in morality as much as intelligence. The reliance on certain experimental procedures—adequate controls and "blind" testing—were but the visible signs of an inner conviction—a commitment to the "merciless search for errors".<sup>71</sup> So long as a small community of like minded individuals endorsed the experimental standards of the Council, reformers remained confident of eventual success. If some physicians were motivated by an interest in knowledge, then so all might be someday.

Along with a belief in science's morality went suspicions of the immorality of commerce. A loyalty to theories was regrettable but excusable; a loyalty to products and firms was more damning. Commercially sponsored therapeutic research was guilty of bias and unreliability, until proven innocent. Yet reformers' congenital distrust of commercial sponsors restricted an already limited field of opportunities for financing therapeutic research.<sup>72</sup>

---

<sup>70</sup> On the late nineteenth century background to such views, see the essays by David A. Hollinger, "Inquiry and Uplift: Late Nineteenth-Century American Academics and the Moral Efficacy of Scientific Practice," and Thomas L. Haskell, "Professionalism versus Capitalism: R.H. Tawney, Emile Durkheim and C.S. Peirce on the Disinterestedness of Professional Communities," both in Thomas L. Haskell, The Authority of Experts. Studies in History and Theory (Bloomington: Indiana University Press, 1984), 142-156, 180-225; and the introduction to Charles Rosenberg's No Other Gods. On Science and American Social Thought (Baltimore: The Johns Hopkins University Press, 1978), 1-21.

<sup>71</sup> The phrase is Torald Sollman's, from "Evaluation of Therapeutic Remedies in the Hospital," (n. 34), 1280.

<sup>72</sup> For more on the problems of financing therapeutic research prior to World War II, see chapter three.



The reformers' view of expertise left them ambivalent towards the practicing physician. What distinguished the expert was not that he possessed truth, but that he renounced false complacency. Recognizing the limits of one's knowledge was of equal importance with having a supply of it in the first instance. Putatively men of science, practicing physicians were eligible and expected to join its community. Anticipating the best of doctors, reformers were nonetheless unsurprised when they found the worst. Men not found in a state of grace are, characteristically, sinners. Physicians who ignored the Council's precepts demonstrated only their own shortcomings.

Undoubtedly, the language of religion would sound foreign to the ear of therapeutic reformers. But it captures the ambivalence towards the physician evinced in the dialect of morality which they did employ. By the 1930's even that form of expression was becoming uncouth in the corridors of science, but the attitudes it captured continued to thrive. Physicians were supposed to be guided by the dictates of experimental evidence although, like other fallen men, they might be expected to stray from the indicated path.

Reformers' expectations of the medical profession constrained their scope of action: to tell physicians what they must do, rather than instruct them in what they should do, went beyond the limits of the Council's authority. Vis-a-vis manufacturers, the Council had some direct powers: they could forbid non-compliant firms access to the

advertising pages of JAMA and other cooperating journals.<sup>73</sup> With non-compliant physicians, they were armed only with persuasion. Lectures and exhortations were the only public outlet for their private frustration with the practitioner. Revision of the federal drug legislation in 1938 provided reformers with an opportunity to extend the authority of the experiment over companies and practicing physicians alike.

---

<sup>73</sup> The Council's influence over firms seemed, at times, just as shaky as its claims on the medical profession, and it then became a question of who it was most politic to attack. Some Council members, more frustrated with the profession's lack of response than with industry's, thought attacks on firms "whose cooperation one seeks" unwise. See Bulletin 31 (March 20, 1920), 91 and Bulletin 31 (March 27, 1920), 110.

Chapter Two. Regulating Medicine.  
The 1938 Drug Act

It is a well recognized scientific fact that a drug may be useful under certain conditions, useless under other conditions, or detrimental under still other conditions, and in coming to a decision with respect to the safety of a drug it is necessary to take into account the use to which the drug is put. [....] Safety as applied to a drug in the Act is unquestionably a relative term. Absolute safety is very difficult to obtain.<sup>1</sup>

What we are seeking to do here is to put the courts on notice ...that it is the purpose of Congress to have consideration given therapeutic claims.<sup>2</sup>

On June 25, 1938, after five years of failed attempts, Congress enacted legislation authorizing the U.S. Food and Drug Administration to review the safety and composition of new drugs. Passed in the wake of an episode of mass poisoning from an untested drug, the Federal Food, Drug and Cosmetic Act of 1938 granted the FDA authority to act before, not after, the hazards of a drug were revealed through distribution on the open market.<sup>3</sup> The new law required drug firms to demon-

<sup>1</sup> J.J. Durrett, "Some of the Implications of Section 505(b) (1) of the Food, Drug and Cosmetic Act," American Drug Manufacturers Association, Proceedings. 28th Annual Meeting (1939), 98.

<sup>2</sup> W.G. Campbell [FDA], in Congressional testimony. U.S. Senate. Committee on Commerce. Hearings on S. 2800. Food, Drugs and Cosmetics. February 27 to March 3, 1934, 557. 73rd Congress, Second Session.

<sup>3</sup> Public Law 717, Federal Food, Drug and Cosmetic Act, 75th Congress, Third Session. The best source on the legislative history of the 1938 law remains David F. Cavers, "The Food, Drug and Cosmetic Act of 1938: Its Legislative History and Its Substantive Provisions," Law and Contemporary Problems 6 (Winter, 1939), 2-42. See also James Harvey Young, The Medical Messiahs. A Social History of Health Quackery in Twentieth Century America (Princeton: Princeton University Press, 1967), 158-190. Charles O. Jackson's Food and Drug Legislation in the New Deal (Princeton: Princeton University Press, 1970) describes the lobbying process in excruciating detail; it is rather short on analysis, however.

strate that drugs were "safe for use under the conditions prescribed, recommended or suggested in the proposed labeling thereof," wording which reflected the view of therapeutic reformers that no drug is absolutely safe, and that many drugs are, in fact, quite dangerous.<sup>4</sup>

The new law left it up to manufacturers to demonstrate that their products were safe to use as recommended. The FDA's job, according to the statute, was to establish the tests and criteria by which manufacturers' statements about drug safety were to be judged. In determining a drug's safety, FDA officials applied a utilitarian calculus: a "safe" drug was one whose proposed use would benefit more patients than it harmed. A drug which was unsafe for treating colds might be safe for treating pneumonia or influenza. In the view of FDA officials, such assessments required an evaluation of therapeutic merit. Even an inert, but ineffective drug allowed on the market would do harm if it kept the patient from treatment with a more effective, albeit more toxic drug. To formulate judgments about therapeutic merit, the FDA relied on a system of expert consultants not unlike that developed by the Council on Pharmacy and Chemistry. Decisions in difficult cases were arrived at by a consultative process in which the opinions and values of clinical specialists as well as their data were elicited and deliberated.

The claim that considerations of drug efficacy played an important role in the 1938 law will come as a surprise to readers of recent

---

<sup>4</sup> Federal Food, Drug and Cosmetic Act, Chapter 5, Section 505 (d). This clause sets the standards for new drugs; sec 502 (j) authorized the secretary to treat as misbranded any drug which was dangerous to health when used in the dosage and conditions recommended in the labeling.

historical accounts of drug regulation. The 1938 law is regarded as allowing the FDA to regulate the safety of new drugs, and no more.<sup>5</sup> The narrow construction such interpretations place on the notion of drug safety is an artifact which, in large part, results from reading subsequent regulatory and judicial history into the original legislation. Before 1946, there is little evidence that anyone, even members of the regulated industry, questioned the FDA's interpretation of the Act. In the interim, the FDA operated on the presumption that the determination of drug safety necessarily entailed making judgments about therapeutic merit.<sup>6</sup> To advocates of a rational therapeutics, assessing a drug's dangers and identifying its limitations were part of the same task.<sup>7</sup> Passage of the 1938 Act marked the end of a long effort to put these convictions to work in federal drug regulation. It was not the assumptions about drug safety but the power to act on them which form the novelty of the 1938 legislation.

---

<sup>5</sup> See Peter Temin, Taking Your Medicine. Drug Regulation in the United States (Cambridge: Harvard University Press, 1980), 38-57.

<sup>6</sup> Only after challenges to specific FDA rulings were upheld by the Courts did the FDA began looking to Congress for additional legal authority to make therapeutic efficacy an explicit, rather than implicit aspect of drug safety.

<sup>7</sup> A.J. Carlson to Mr. Virgil Chapman, February 21, 1938. Record Group 88. Accession no. 52A-89. Box 124. Folder 062 (Food and Drug Act A. March-April). Washington-National Records Center [hereafter cited as W-NRC]. Chapman was the sponsor of new drug legislation in the House of Representatives. See also W.G. Campbell [FDA], in Congressional testimony. U.S. Senate. Committee on Commerce. Hearings on S. 2800. Food, Drugs and Cosmetics. February 27 to March 3, 1934, 550-552, 73rd Congress, Second Session; and Joseph A. Capps, "Irrational Tendencies in Modern Therapy," JAMA 83 (July 5, 1924), 1-3.

Drug Regulation. 1900-1938.

The history of drug regulation in this century reflects a central tenet of therapeutic reformers: the more potent the drug and the more serious the disease for which it was intended, the greater the importance of regulating the therapeutic claims made on its behalf. Federal efforts at drug regulation in this century began with the Biological Control Act of 1902, which mandated the Public Health Service's Hygienic Laboratory to regulate the interstate commerce of "viruses, serums, toxins and analogous products". The Act authorized the Hygienic Laboratory to license manufacturers to market specific products, to set standards and test for the potency of approved items, to inspect manufacturers' facilities before and after licensing, and, to a limited extent, evaluate manufacturers' claims concerning the therapeutic value of their products.<sup>8</sup> Questions of therapeutic merit which could be resolved in the laboratory were actionable: a license need not be granted to inactive or sub-therapeutic products. But beyond laboratory

---

<sup>8</sup> The Hygienic Laboratory was the precursor of the National Institutes of Health which took over its regulatory functions in 1930. Responsibility for biologics remained with the Public Health Service until 1972, when it was transferred to the FDA. In its early years, the Hygienic Laboratory supervised the production not only of biologic products such as serums and vaccines but as of 1919, the chemically produced anti-syphilis drug such as arsphenamine. On their regulatory activities, see Laurence F. Schmeckbeier, The Public Health Service. Its History, Activities and Organization (Baltimore: The Johns Hopkins Press, 1923), 27, 129-133; Ramunas Kondratas, "The Biologics Control Act of 1902," in James Harvey Young, ed. The Early Years of Federal Food and Drug Control (Madison: American Institute of the History of Pharmacy, 1982), 8-27 and Jonathan Liebenau, Medical Science and Medical Industry, 1890-1929: A Study of Pharmaceutical Manufacturing in Philadelphia, 254-258, 269-270. Ph.D. Thesis. University of Pennsylvania, 1981. Liebenau correctly stresses the extent to which the activities of the H.L. under the 1902 legislation were a more significant model for subsequent approaches to drug development and evaluation than the 1906 food and drug legislation.

tests, the Hygienic Laboratory had no legal authority to determine therapeutic benefit. Like the Council on Pharmacy and Chemistry, it found proof positive that a product lacked clinical value the most difficult of claims to establish, and it proceeded to tread cautiously in this area.<sup>9</sup>

The 1906 Food and Drug Act, while covering a greater range of products, granted even fewer powers. The Act gave the federal government no right to screen drugs prior to sale. It merely authorized the Bureau of Chemistry to seize "adulterated" or "misbranded" products.<sup>10</sup> From the reformers' point of view, the law had many defects: only claims which physically accompanied the product label were actionable; claims made in advertising the product weren't. Producers who elected not to specify the contents of their products on their labeling could

<sup>9</sup> Under the 1902 law, the Hygienic Laboratory regularly tested new products for their "therapeutic or prophylactic value" in the laboratory. See Schmeckbeier, The Public Health Service, 130. On several occasions, HL officials emphasized the difficulties of establishing a clear case for disallowing products whose therapeutic value was, in their view, doubtful. G.W. McCoy, "Official Methods of Control of Remedial Agents for Human Use," JAMA 74 (June 5, 1920), 1554. See also "Report of the Committee on Sera and Vaccines," Council on Pharmacy and Chemistry, Bulletin (August 20, 1908), 88-89, American Medical Association archives. Victor H. Kramer attributes the hesitations of the HL to act in certain cases to the personal philosophy of George McCoy, the director, who was reluctant to dictate to the medical profession in an arena characterized by strong differences of opinion. These views did not prevent McCoy from acting against certain products, most notably vaccines intended for the treatment of tuberculosis. See Victor H. Kramer, The National Institute of Health. A Study in Public Administration (New Haven: Quinpiack Press, Inc., 1937), 31-37. My thanks to Dr. Raumanas Kondratas for calling this reference to my attention.

<sup>10</sup> The 1906 Act was administered by the Bureau of Chemistry in the Department of Agriculture until 1927, when its regulatory functions were established in a separate agency: the Food, Drug and Insecticide Administration (after 1930, the FDA). See Young, Medical Messiahs (n. 3), 98. The diversity of nomenclature obscures a strong continuity of personnel. See Temin, Taking Your Medicine (n. 5), 40.

not be prosecuted for misrepresenting them, nor could producers whose labeling acknowledged the presence of a small number of mandated dangerous substances. Its greatest flaw was that it was remedial, but not preventive: the Bureau could act only after a drug was distributed, and harm done.<sup>11</sup>

Even under the limited authority of the 1906 Act, the Bureau of Chemistry tried to regulate therapeutic efficacy, by confiscating as "misbranded" products which made grossly inflated or misleading therapeutic claims. From the regulators' perspective, even a non-toxic but ineffective drug represented "a definite public health menace ... because its use may cause delay in resorting to rational methods of treatment."<sup>12</sup> As interpreted by the courts, however, the law required that the Bureau demonstrate not merely the falsity of a manufacturer's claims but that he knew the claims to be false. Such demonstrations were costly, time consuming, and hampered by the legal difficulties of

---

<sup>11</sup> The limits of the labeling provision, and its failure to contain comparable abuses in free-standing advertising, were singled out for criticism by Bureau Chemist Carl L. Alsberg in an otherwise positive ten year review of the Act. See Gustavus A. Weber, The Food, Drug and Insecticide Administration. Its History, Activities and Organization (Baltimore: The Johns Hopkins Press, 1928), 27. While the 1906 law required that certain narcotic or otherwise and hazardous components be identified when used, it was no crime to employ them. The requirement, moreover, did not cover a longer list of equally dangerous substances. See Young, Medical Messiahs (n. 3), 54.

<sup>12</sup> Report of the Food and Drug Administration, 1930, 11; see also Report of the Food and Drug Administration, 1931, 15-16.



proving fraudulent intent.<sup>13</sup> The Bureau's aspirations to regulate therapeutic claims foundered on the courts' reluctance to accept scientific and medical authority as prima facie grounds for establishing a standard of knowledge in this area. The courts did not recognize the distinction, embraced by the Bureau, between the opinions of experts, backed by laboratory studies of drug action and a critical review of the clinical literature, and the testimony of physicians and patients, grounded on uncontrolled personal experience. Proving fraud under these conditions was onerous.<sup>14</sup>

---

<sup>13</sup> According to Peter Temin's reading, Justice Holmes' 1911 decision in *U.S. vs. Johnson* (221 U.S. 448) instituted the requirement that the Bureau prove fraudulent intent, and further restricted the scope of the Act by excluding false therapeutic claims as grounds for fraud because they rested on the unreliable and uncertain authority of medical opinion. Subsequent legislative amendments and court hearings only reinforced Holmes' reading. Temin, Taking Your Medicine (n. 5), 32-34. For a further discussion of Temin's account see note 14, *infra*.

James Harvey Young offers a more intricate and ambiguous legal history for the 1906 Act, which places greater emphasis on the opportunity costs legal precedent imposed on the Bureau's regulatory strategy, and the consequent emphasis on obtaining the industry's voluntary compliance. See his Medical Messiahs (n. 3), Chapters 1, 3 and 5, especially 53-54, 56-59, 92-96, 99-101, 111-112. Even a successful prosecution did not guarantee permanent protection: a manufacturer who chose to revise the claims for his product could safely begin sales again (Young: 12, 104-106).

<sup>14</sup> Three distinct issues arise in reading the judicial history of the 1906 Act: 1) Was the Bureau of Chemistry required to prove the fraudulent "intent" of manufacturers as well as the falsity of their claims? 2) Were false therapeutic claims, as opposed to false claims about drug content, covered under the Act? 3) Did reliable medical opinion, as represented, for example, by experts of the Public Health Service, have any special standing in the Courts, or was medical opinion too uncertain an authority to justify reliance on it in official administrative decisions? Historians, following the lead set by the courts, have generally conflated the three issues.

The initial ruling which excluded scientific assessments of therapeutic efficacy from consideration dealt not with the 1906 Drug Act but an earlier set of decisions regarding the Postal Service. In 1900, the Postmaster General attempted to seize the advertizing of a modern faith healer who offered to transmit "magnetic healing" long distance over the ether to his eager subscribers. The Supreme Court

The implications of judicial precedent were not lost on reformers. In 1933, the Roosevelt administration provided FDA officials with a long-sought opportunity to extend their powers. When Senator Royal Copeland introduced legislation to replace the 1906 law, FDA officials urged Congress to "put the courts on notice" that evaluating therapeutic

---

ruled that the P.G. had not proved its charge of fraud merely because some people may doubt the value of magnetic healing. In any area where medical opinions may differ, the Court argued "there is no exact standard of absolute truth by which to approve ... [an] assertion false and a fraud." Young, Medical Messiahs (n. 3), 70-71.

The principal effect of the McNulty Doctrine, as the decision came to be known, was on the regulatory activity of the Postmaster General and the Federal Trade Commission [Young, 85-87, 125-126]. In 1911, however, Justice Holmes cited the McNulty Case in ruling against the Bureau of Chemistry. Holmes' decision was based entirely on the grammatical exposition of the 1906 Act, which confined the Bureau of Chemistry to acting on false statements concerning the composition of products. While he cites the McNulty doctrine in a closing paragraph, Holmes does so while explicitly ruling out reliance on "broader" constitutional issue such as the uncertainty of medical opinion, an approach which is in keeping with his generally restrictive approach to judicial opinion. (Justice Hughes, in contrast, goes on much greater length about the questions of Congressional "intent" and the reliability of medical opinion.) [See U.S. vs. Johnson, 221 U.S. 488 (1911).] Holmes's decision did affect the Bureau's effort's to regulate therapeutic claims, but there is little evidence supporting Peter Temin's contention that the decision influenced the language of the 1912 Sherley amendments [Temin, Taking Your Medicine (n. 5), 33-34]. In fact, the opposite appears to be true: "There is ... a wide field in medicine within which the curative or therapeutic effects is as well known and as definitely determined as the law of gravitation. Within that field, apart from any question of opinion, the fact that a so-called remedy is absolutely worthless and its label false and fraudulent is easily susceptible of proof." House Report 1138, 62nd Congress, 2nd Session, 1912, cited in Ashley Sellars and Nathan D. Grundstein, Administrative Procedure and Practice In the Department of Agriculture Under the Federal Food, Drug and Cosmetic Act of 1938 (Washington: USDA, 1940), Part I, 14.

claims was part of their job.<sup>15</sup> The subject of extensive Congressional hearings, the proposed legislation underwent numerous revisions in response to industry and consumer criticisms. But little objection was made to the principle that a new drug law address the truth of therapeutic claims. To the extent the subject was discussed at all, attention focused on the choice of means for accomplishing this end.

The problem, as construed by the legislators, was to avoid excessive delegation of federal authority which might bring about unfavorable judicial review of the statute. The difference between an enforceable law and an unconstitutional one depended on the way in which judgments about therapeutic merit were arrived at. Reformers had proposed that powers be vested in a scientific board of experts to advise the FDA. Their proposal foundered in disagreements over the appropriate membership for this body. In subsequent drafts, the idea of a formal consultative body was abandoned, along with any language which evoked the bar of "medical opinion" to which judicial wrath had

---

<sup>15</sup> W.G. Campbell [FDA], in Congressional testimony. U.S. Senate. Committee on Commerce. Hearings on S. 2800. Food, Drugs and Cosmetics. February 27 to March 3, 1934, 557. 73rd Congress, Second Session. Although the bill Copeland introduced in 1933 was prepared by the administration, Copeland himself had earlier attempted to extend the scope of federal authority regarding drugs by amending the 1902 Biologics Act to regulate therapeutic claims. His attempts were rebuffed by Public Health Service officials and the director of the Hygienic Laboratory. Royal S. Copeland to H.S. Cumming, February 12, 1924; G.W. McCoy to A.H. Stimson, January 29, 1924. RG 90. Public Health Service General Files. 1924-1935. Box 70. National Archives; hereafter NA. The Hygienic Laboratory's reticence in these matters contrasts with that of FDA officials, who had made an earlier attempt, in 1912 to revise the drug statute on the model of the 1902 Biologics Act. See Sellars and Grundstein, Administrative Procedure and Practice (n. 14), Part I, 82-83.

been drawn.<sup>16</sup> The new law placed its confidence instead in the FDA, which was to decide "by all methods reasonably applicable" whether a new drug was "safe for use" under the conditions proposed by the producer.<sup>17</sup>

---

<sup>16</sup> On the proposal for a body of experts to advise the FDA, see A.J. Carlson to Mr. Virgil Chapman, February 21, 1938. RG 88. 52A-89. Box 124. F 062 (Food and Drug Act A. March-April). W-NRC; and Stephan Wilson, Food and Drug Regulation (Washington: American Council on Public Affairs, 1942), 105.

The testimony of W.G. Campbell [FDA] and the ensuing discussion by Senators Copeland and Herbert emphasize the importance of finding a legally defensible procedure for incorporating the views of experts. See U.S. Senate. Committee on Commerce. Hearings on S. 2800. Food, Drugs and Cosmetics. February 27-March 3, 1934, 548-558. 73rd Congress, Second Session. The AMA proposed that "competent medical witnesses" or the authority of "reputable" journals be specified and defined in the legislation, to avoid the ambiguity in terms like "substantial medical opinion." Testimony of William C. Woodward, U.S. Senate, Committee on Commerce, Hearings on S. 2800. February 27-March 3, 1934, 376. 73rd Congress, Second Session. See also Bureau of Legal Medicine and Legislation, AMA, "Food, Drug, Therapeutic Device and Cosmetic Legislation Pending in Congress," JAMA 105 (December 21, 1935), 2055-2062. Even representatives of industry were agreed in the importance of recognizing "substantial medical or scientific opinion" in determining the truth of drug claims. See testimony of James F. Hoge, Proprietary Manufacturers Association, Hearings Before a Subcommittee of the Committee on Commerce on S. 5. March, 1935 in U.S. Food and Drug Administration, A Legislative History of the Federal Food, Drug and Cosmetic Act and Its Amendments (Washington: FDA, n.d), v. 3, 325.

<sup>17</sup> Public Law 717, Federal Food and Drug Act, 75th Congress, Third Session. Section 505 (d). The actual legal authority was nominally vested in the administrator of the Federal Security Agency, under whose jurisdiction the Food and Drug Administration operated.

Implementation: "Dangerous" Drugs

...potent drugs, containing large amounts of acetanilid and bromide, should not be sold over the ice cream counter, as if they were just another fizz concoction in the same class as ice cream soda.<sup>18</sup>

The average layman little realizes ... the serious consequences which may follow indiscriminate and ignorant use of potent ethical preparations whose administration the manufacturer has tried to limit to professional supervision.<sup>19</sup>

The 1938 Act required the FDA to see that drugs were "safe for use ...under the conditions recommended." But according to clinical specialists, any drug was potentially unsafe:

I do not know what the word 'dangerous' means in the way he uses it. I am sure it is at least grossly undesirable for people to take indefinitely 9 grains of acetanilid or 20 grains of bromide.<sup>20</sup>

In a regulatory context, such ambiguity might lead to legal challenges, and possibly, defeats. One of the first tasks facing FDA officials was to translate the opinions of reformers into defensible policies.

The majority of drugs presented to the FDA posed few problems. The rank fraud and the known poison were easy to identify. As with the AMA Council, either the laboratory or the library sufficed in many cases to evaluate minor variations on relatively innocuous compounds.

<sup>18</sup> Theodore G. Klumpp, "The Work of the Federal Food and Drug Administration," Journal of the Medical Association of Georgia 28 (July 1939), 280.

<sup>19</sup> Memorandum Submitted by Winthrop Chemical Company... With Respect to Proposed Regulations for the Enforcement of the Federal Food, Drug and Cosmetic Act, 7. RG 88. 52A-89. Box 144. F. 603 (Proposed Regulations), Book 3, W to Z. W-NRC.

<sup>20</sup> George Minot to Sama Weiss, October 3, 1939. Sama Weiss papers, Francis M. Countway Library of Medicine, Harvard Medical School.

By requiring adequate documentation of a drug's composition and extensive screening on animals for toxicological effects, the FDA readily disposed of most applications. The more difficult problem was posed by drugs whose potency was not in doubt, but whose benefit depended on the precautions taken in their use.

The men in charge of the Drug Division, Drs. Theodore Klumpp and J.J. Durrett, came to the FDA from backgrounds in academic medicine, and were familiar with the contention of reformers that many worthwhile drugs were causing problems because of "inappropriate" use.<sup>21</sup> They soon put their contacts to work to document the hazards of such use.<sup>22</sup> Among the drugs on which they focused attention was a powerful new anti-infective agent, sulfanilamide.

The first of a new class of compounds known as sulfonamides, sulfanilamide was introduced here in 1936. Almost overnight, it

---

<sup>21</sup> Durrett, a graduate of Harvard Medical School in 1914, had spent most of his career in government service. He had been responsible for drug control at the U.S. Department of Agriculture since 1928. Shortly after passage of the new law, Durrett was kicked upstairs, to serve as a technical advisor to W.G. Campbell on administration of the new law. His post as chief of the Drug Division at FDA was taken over by Theodore Klumpp. Klumpp, a more recent graduate of Harvard (1928) had come to the FDA in 1936 from a teaching post at Yale Medical School. See Jacques Cattell, ed. American Men of Science. 10th edition. (Tempe, Arizona: Jacques Cattell Press, 1960-1962), I: 1021; "Changes in the Food and Drug Administration," JAMA 111 (September 17, 1938), 1116; Who's Who in America, 1982-1983. 42nd edition. (Chicago: Marquis Publishing, 1982), I: 1838.

<sup>22</sup> Dr. George Dobbs and Dr. James Q. Gant, Memorandum of Interview, Dr. DuBois, October 4, 1938. RG 88. 52A-89. Box 137. F. 512.1.10-512.6. W-NRC; George Minot to Sama Weiss, October 3, 1939. Sama Weiss papers, Francis M. Countway Library of Medicine, HMS. In addition to prominent clinicians, surveyed for their experience with particular drugs, hospital superintendants at large, prestigious institutions were asked to report adverse drug reactions on specific drugs. For a list of individuals surveyed in March, 1939, see RG 88. 58A-277. Box 38. F 511.07-512. W-NRC.

established a reputation as a "truly remarkable" drug. Its dramatic success in treating advanced streptococcal infections were followed by evidence of its value in meningococcal, gonococcal and other infections.<sup>23</sup> Each report was accompanied by a wave of publicity advertising the wonder drug's merits.<sup>24</sup>

Among clinical investigators, enthusiasm for sulfanilamide's healing potential was soon tempered by knowledge of the drug's toxicity: initial reports of mild reactions were supplemented by the fall of 1937 with studies which associated its use with anemias, depressed white cell counts and a variety of other serious blood disorders.<sup>25</sup> Sulfa-

---

<sup>23</sup> A handful of American investigators tried out samples of the drug in 1935, but "serious laboratory and clinical investigations" here did not began until 1936. Perrin H. Long and Eleanor A. Bliss, The Clinical and Experimental Use of Sulfanilamide, Sulfapyridine and Allied Compounds (New York: The MacMillan Co., 1939), 9. For a full discussion of the range of conditions treatable with sulfanilamide, and a critical evaluation of the evidence upon which such claims were based as of 1939, see Long and Bliss, 147-229. For general background on the development and early use of sulfanilamide, see Harry Dowling, Fighting Infection. Conquests of the Twentieth Century (Cambridge: Harvard University Press, 1977), Chapter 8.

<sup>24</sup> James Harvey Young, "Sulfanilamide and Diethylene Glycol," in John Parascandola and James C. Whorton, eds. Chemistry and Modern Society (Washington: American Chemical Society, 1983), 107. One of the first clinical uses of the drug in this country was on Franklin Roosevelt, Jr., whose dramatic recovery was widely reported in the press. In addition to the sources cited by Young, see "Prontosil," Time (December 28, 1936), 21; "Again, Sulfanilamide," Time (August 30, 1937).

<sup>25</sup> Perrin H. Long and Eleanor Bliss, "Para-Amino-Benzene-Sulfonamide and Its Derivatives. Experimental and Clinical Observations On Their Use in the Treatment of Beta-Hemolytic Streptococcic Infection: A Preliminary Report," JAMA 108 (January 2, 1937), 37. For early rumours of acute hemolytic anemia and agranulocytosis, see Council on Pharmacy and Chemistry, "Report of the Council. Sulfanilamide and Related Compounds." JAMA 108 (May 29, 1937), 1889. Subsequent reports, more adequately documented, include A.M. Harvey and C.A. Janeway, "The Development of Acute Hemolytic Anemia During the Administration of Sulfanilamide," JAMA 109 (July 3, 1937), 12-16; S.E. Kohn, "Acute

nilamide joined a long list of drugs whose value depended on the intelligence and skill with which it was used. When employed to treat the life-threatening conditions for which it was effective, under conditions which allowed for the close monitoring of drug reactions, sulfanilamide's contribution was welcome. But until more was known about the drug's toxic properties and the manner in which the body used the drug, caution was advised and its indiscriminate use discouraged.

"It is not," the AMA Council urged, "a panacea."<sup>26</sup>

By the Spring of 1938, evidence of sulfanilamide's potential toxicity was well established. Whether or not the more severe reactions associated with the drug were due to the "idiosyncrasies" of individual patients was still under investigation. But the FDA found little disagreement among the experts they consulted that sulfanilamide, employed without awareness of its potential to cause such reactions, was a dangerous drug. Patients being treated with sulfanilamide, authorities agreed, were best handled under close medical supervision.<sup>27</sup> On August 26, 1938, the FDA announced that marketing of sulfanilamide

---

Hemolytic Anemia During Treatment with Sulfanilamide," JAMA 109 (September 25, 1937), 1005-1006; Perrin H. Long and Eleanor A. Bliss, "Clinical Use of Sulphanilamide and Its Derivatives in the Treatment of Infectious Disease," Annals of Internal Medicine 11 (October 1937), 584-585.

<sup>26</sup> Council on Pharmacy and Chemistry, "Report of the Council. Sulfanilamide and Related Compounds." JAMA 108 (May 29, 1937), 1889. The Council's cautions were echoed by the other investigators reporting on the drug. See references in note 25, above.

<sup>27</sup> On the issue of individual idiosyncrasy, see "Sulfanilamide and the Leukocytes," JAMA 110 (January 29, 1938), 372-373, and the studies cited therein. The most detailed work was being done at Johns Hopkins. See E.K. Marshall, Jr., W.C. Cutting, Kendall Emerson, Jr., "The Toxicity of Sulfanilamide," JAMA 110 (January 22, 1938), 252-257.



which allowed for "indiscriminate use by the general public" would be "actionable" under the new law.<sup>28</sup> Manufacturers would be liable to prosecution, unless they attached to packages of sulfanilamide a "warning so conspicuous as to certainly attract attention" that the drug was dangerous unless used under "appropriate medical supervision".<sup>29</sup>

The new law said nothing about the FDA's right to control the way drugs were used; its authority was vested in the right to regulate what manufacturers said about a drug. But the idea of restricting the sale of sulfanilamide to physicians did not trouble manufacturers.<sup>30</sup> What

<sup>28</sup> The notice on sulfanilamide was soon followed in September by similar announcements on aminopyrine and cinchophen, drugs whose toxic properties had long been noted in the literature. See Charles Wesley Dunn and Vincent A. Kleinfeld, Federal Food, Drug and Cosmetic Act. Judicial and Administrative Record, 1939-1949. (New York: Commerce Clearing House, 1949), Trade Correspondences 1, 2, 3. It is important to note that decisions were made on a case by case basis. The dangers of some drugs proved especially difficult to establish, owing to differences of medical opinion about the relative risk-benefit ratios. The use of thyroid substances for weight reduction were one such difficult case. See the correspondence in RG 88. 58A-277. Box 22. F. 512.1.10-512.6-31, especially Fuller Allbright to Theodore G. Klumpp, January 28, 1939. See also Theodore G. Klumpp to Ephraim Shorr, October 22, 1938. RG 88. 52A-89. Box 137. F. 512.1 and Theodore G. Klumpp to Arthur G. Sullivan, June 4, 1940. RG 88. 59A-2736. Box 27. F. 511.06. W-NRC. For a general account of the thyroid campaign, see Young, Medical Messiahs (n. 3), 210-215.

<sup>29</sup> Dunn and Kleinfeld, Federal Food, Drug and Cosmetic Act (n. 28), 562-563. Trade Correspondence-4 [undated, but presumably October 1938].

<sup>30</sup> In at least one draft of the proposed new law, industry spokesmen had themselves proposed that products sold to physicians be exempted from the bill's labeling requirements. James F. Hoge, A Bill for a Food, Drug and Cosmetic Act [1935]. RG 88. Food and Drug Administration. Office of the Commissioner. Legislation, 1927-1940. Box 10. NA. After the bill's passage, one manufacturer objected, not to the requirement that certain drugs be used under medical supervision but to the possibility that they might not be so used. According to Winthrop Chemical Company, the proposed labeling requirements were tantamount to granting a "correspondence course in medicine" to laymen who were unqualified to understand, much less evaluate, the dangers of using such drugs. Memorandum Submitted by Winthrop Chemical Co., Inc., November 25, 1938. RG 88. 52A-89. Box 144. F 603. Proposed Regulations. Book 3. W to Z. W-NRC.

bothered them was the possibility of being held accountable for the acts of others: the distributors, licensees and purchasers of their products.<sup>31</sup> The manufacturers proposed a compromise: require detailed labeling except in cases where the product was distributed only to professionals, where companies would be allowed to label the actual product "for professional use only". Such labeling would substitute for the detailed warnings otherwise called for.<sup>32</sup> With minor modifications, the FDA accepted the industry proposal.<sup>33</sup> That "solved" the problem of "indiscriminate" use by the lay public.<sup>34</sup> Regulating unintelligent practices by the medical profession proved to be another problem.

---

<sup>31</sup> "Testimony of James F. Hoge," Hearings on Proposed Regulations, November 17 & 18, 1938, pp. 70-96. RG 88, 52A-89, Box 144. F 603. Book 1. W-NRC. See also Hoge's remarks as reported in "Drug Law Hearing Shows Labeling Opposition," Oil Paint and Drug Reporter 134 (November 21, 1938), 32A. A prominent industry concern was the requirement that all relevant information be placed on the drug package: the logistic and economic difficulties of reprinting thousands of existing labels were troublesome.

<sup>32</sup> Memorandum Submitted by Winthrop Chemical Co, Inc., November 25, 1938. RG 88. 52A-89. Box 144 F. 603. Proposed Regulations. Book 3. W to Z; Charles Wesley Dunn [Genl Counsel, American Pharmaceutical Manufacturers' Association] To Members. Bulletin No. 261. March 3, 1939. RG 88. 58A-277. Box 22. F. 512.1 (General Correspondence January-March). W-NRC.

<sup>33</sup> "Professional" labeling would be permitted and the companies would not be liable if the products were misused. They would, however, lose the right to continue shipping the drug under the "professional" exemption. The question of liability concerned to be of concern to manufacturers, however. See the discussion in J.J. Durrett, "Some of the Implications of Section 505(b) (1) of the Food, Drug and Cosmetic Act," American Drug Manufacturers Association, Proceedings, 28th Annual Meeting (1939), 105-112.

<sup>34</sup> On the difficulties of policing the marketing of drugs with a restricted labeling, both before and after the formal promulgation of the regulations, see O. Olsen to George P. Larrick, December 27, 1938. RG 88. 58A-277. Box 22. F. 512.1. W-NRC; J.J. Durrett to Assistant General Counsel [FDA], October 24, 1940. RG 88. 59A-2736. Box 24. F 505.1-508.2. W-NRC. Compliance was highly variable, from year to year and depending on the drug in question. Food and Drug Administration, Annual Report 1941, 17; Annual Report 1942-1943, 39-40.

Over the course of the 1920's, therapeutic reformers had continued their researches into the pharmacological principles which underlay effective therapeutic practice. The growth of knowledge about the proper use of familiar drugs such as digitalis and arsphenamine enhanced awareness of the dangers inherent in inappropriate use.<sup>35</sup> The Elixir Sulfanilamide episode only reinforced this concern. By 1937, more than 100 firms were marketing variants of sulfanilamide. One such company, S.E. Massengill, hit upon the ingenious idea of manufacturing sulfanilamide in syrup form. The substance they chose as a buffer, ethyl diglycerol, was an antifreeze additive whose toxic properties were well known, except to Massengill's chemist. Distribution on the open market led to 106 deaths, and the subsequent passage of the 1938 drug act.<sup>36</sup>

The job of tracking down the victims and reporting on the circumstances of their deaths fell to the FDA. While putting the new law to work, the FDA's Klumpp was also writing up an analysis of the tragedy which led to its passage:

I think you will be interested in some of the implications that arise from the observations recorded. I refer particularly to the fact that of 105 deaths associated with the consumption of the drug and, to the best of our knowledge, attributable to the drug, in a hundred instances the drug was administered on a physician's

---

<sup>35</sup> On digitalis, see the remarks by Henry Christian in "Abstract of Discussion," JAMA 75 (Aug 14, 1920), 465.

<sup>36</sup> Young, "Sulfanilamide and Diethylene Glycol," (n. 24) 105-126.

prescription. The physician's diagnoses ...are also of some interest.<sup>37</sup>

Among the questionable conditions for which the drug was prescribed, Klumpp noted "Bright's disease, bichloride of mercury poisoning, renal colic and backache," none with the remotest connection to the infectious diseases for which sulfanilamide was known to work. Moreover, in "most cases" the laboratory tests necessary to monitor patients on sulfanilamide "were not made."<sup>38</sup> If Klumpp and his colleagues needed any additional grounds for believing that physicians too needed more guidance in selecting drugs, the Elixir Sulfanilamide episode provided it.<sup>39</sup>

---

<sup>37</sup> Theodore G. Klumpp to John P. Peters, December 29, 1939. RG 88. 58A-277. Box 38. F. 511.07-512. W-NRC. Klumpp's friend, "Jack" Peters, was at the time engaged in a fairly serious political campaign in support of national health insurance.

<sup>38</sup> Herbert O. Calvery and Theodore G. Klumpp, "The Toxicity for Human Beings of Diethylene Glycol with Sulfanilamide," Southern Medical Journal 32 (November, 1939), 1106-1107.

<sup>39</sup> For additional expressions of FDA scepticism regarding physicians' knowledge of drugs, see the remarks of J.J. Durrett, addressing members of the drug industry: "Not all physicians know all of the established dangers which drugs have and by reiteration of the dangers, even to physicians, you are going to be served because the use of drugs that result in harm is going to be in great measure reflected back on you...." J.J. Durrett, "Some of the Implications of Section 505(b) (1) of the Food, Drug and Cosmetic Act," American Drug Manufacturers Association, Proceedings. 28th Annual Meeting (1939), 104.

New Drugs. The Case of Sulfapyridine

If the drug that killed one person in 10,000 was of only minor use therapeutically it might still be judged to be unsafe, whereas the drug which killed one in a thousand persons if it had marked and undisputed therapeutic value, such as the drug under question, it would still be a safe and valuable drug....<sup>40</sup>

The clinical information submitted with the application is usually the ultimate basis upon which a decision regarding the product is reached.... The appraisal of the clinical work is not easy.<sup>41</sup>

Notwithstanding their problems, the promise of the sulfonamides was substantial. By 1937, numerous chemical compounds related to sulfanilamide were under development. Such variants, researchers counseled, must be tested carefully in controlled laboratory and clinical settings prior to their introduction into clinical practice.<sup>42</sup>

Therapeutic reformers had issued such warnings before, but now they had the opportunity to put the force of law behind them. The new legislation not only enabled the FDA to regulate the hazards of existing drugs, it authorized them to rule on the safety of novel compounds before their introduction into general use. The announcement of a new sulfonamide product in the Spring of 1938 gave the FDA an

<sup>40</sup> J.J. Durrett, Memorandum of Interview with Perrin H. Long and E. Kinnerly Marshall [Johns Hopkins], December 5, 1938. RG 88. 69A-2099. Box 1. NDA 90. vol. 1. W-NRC.

<sup>41</sup> Walton Van Winkle, Jr. to Paul Dunbar and Robert Herwick, January 30, 1946. RG 88. 59A-2736. Box 220. F 505.1074-509. W-NRC.

<sup>42</sup> References to variant compounds under development may be found in "Sulfanilamide—A Warning," JAMA 109 (October 2, 1937), 1128. Lionel Whitby makes reference to over 1000 "such compounds" by the following year: "Chemotherapy of Bacterial Infections," The Lancet ii (November 12, 1938), 1095.

opportunity to define the standards they would use in evaluating new drugs.

In May, 1938, shortly before enactment of the new law, British investigators reported on a new sulfonamide compound, 2-para-aminobenzene sulfamide pyridine (sulfapyridine), which appeared to be extremely effective in a variety of experimental infections. Claims that the drug showed low toxicity in mice led to speculation that it might prove to be a safer, as well as more beneficial drug than sulfanilamide.<sup>43</sup> The prospect of a drug which might be beneficial, when used safely, made sulfapyridine an ideal test case for putting the new law to work.

On October 7, 1938, representatives of Merck & Co. submitted their application for sulfapyridine to the FDA. Along with copies of the published British literature and preliminary clinical reports from investigators in this country, the Merck representatives noted that "great demand had [already] arisen for the article" owing to advance publicity. Review by the AMA's Council on Pharmacy and Chemistry, they added, was imminent.<sup>44</sup> The FDA was not persuaded: the toxicological

---

<sup>43</sup> "Dose for dose, the drug appears more efficient than sulphanilamide, and, in low dose, it is very definitely superior." Lionel Whitby, "Chemotherapy of Pneumococcal and Other Infections With 2-(p-Aminobenzenesulphonamido) Pyridine," The Lancet i (May 28 1938), 1212. The initial British clinical studies were reported by G.M. Evans and Wilfrid F. Gaisford, "Treatment of Pneumonia with 2-(p-aminobenzene-sulphonamido) Pyridine," The Lancet ii (July 2, 1938), 14-19. On comparative toxicity relative to sulfanilamide, see letter of J.B. Ravdin [University of Pennsylvania] to D.F. Robertson [Merck & Co.], September 3, 1938. RG 88. 69A-2099. Box 1. NDA 90. vol. 1. W-NRC.

<sup>44</sup> J.J. Durrett, Memorandum of Interview with Joseph Rosin and R.E. Gruber, October 7, 1938; Joseph Rosin to Henry Wallace, October 7, 1938. Preliminary clinical data from A.R. Dochez, Columbia College of Physicians and Surgeons, and J.B. Ravdin, University of Pennsylvania, were presented with the application. Though finding the drug useful in pneumonia and systemic infections, Dochez noted that the

data submitted was quite "meagre," the manner in which humans metabolized the drug was undocumented, and clinical reports from investigators in this country were few in number. The firm would have to present more data before the application could be acted on.<sup>45</sup>

On scientific and economic grounds, the most promising use of sulfapyridine appeared to be for pneumonia. Existing serum treatments of pneumonia were expensive, required special facilities and experience to use effectively, and only worked on certain strains of pneumococci. Sulfanilamide was of limited use in treating pneumonias. The idea of a cheap, effective drug for pneumonia which could be used readily by the average physician had great appeal.<sup>46</sup> The clinical investigators to whom Merck entrusted sulfapyridine therefore focused their attention on pneumonia cases.

number of cases "are still too small to draw any conclusions based on statistics." Dochez to D.F. Robertson, September 26, 1938. [All references RG 88. 69A-2099. Box 1. NDA 90. vol. 1. W-NRC.]

<sup>45</sup> W.G. Campbell to Joseph Rosin, Merck and Company, Inc., October 28, 1938. Perrin H. Long, who was to act as the FDA's advisor on this drug, had already expressed some scepticism about the "very enthusiastic reports" appearing in the British literature, and advised getting "really good detailed clinical work" from "reputable hospitals" in the U.S. Perrin H. Long to Theodore Klumpp, September 28, 1938. [Both references RG 88. 69A-2099. Box 1. NDA 90. vol 1. W-NRC.]

<sup>46</sup> J.J. Durrett, Memorandum of Interview with Perrin H. Long and E. Kinnerly Marshall [Johns Hopkins], December 5, 1938. RG 88, 69A-2099, Box 1, NDA 90, vol. 1. W-NRC. On the technical difficulties facing the physician who attempted to use serum effectively, and one organized attempt to overcome these barriers, see Frederick T. Lord and Roderick Heffron, Lobar Pneumonia and Serum Therapy, With Special Reference to the Massachusetts Pneumonia Study (New York: The Commonwealth Fund, 1936) and, more generally, Harry Dowling, "The Rise and Fall of Pneumonia-Control Programs," Journal of Infectious Diseases 127 (February, 1973), 201-206.

During the fall, the FDA continued to accumulate clinical and laboratory data favorable to sulfapyridine. The agency did not confine itself to material submitted through Merck but actively solicited the opinions and reports of investigators known to be using the drug.<sup>47</sup> By December, the initial deficiencies in the animal toxicological data had been remedied.<sup>48</sup> The FDA had decided to approve sulfapyridine, but had not yet decided under what restrictions. Determining sulfapyridine's "safety" depended as much on its value in treating specific clinical conditions as on any toxicological data. Safety, according to this philosophy, was "a relative term and its exact meaning for each preparation would, on the basis of the facts, have to be determined."<sup>49</sup>

Investigators agreed that the drug was likely to prove of considerable benefit in treating pneumonias. Questions remained, however, about when and how it was safest to use. Clearly, "the drug was not killing many people" but at the same time, there was "no uniformity of opinion with respect to the harm which the drug might be capable of

---

<sup>47</sup> Harrison F. Flippin to J.M. Carlisle, November 28, 1938; J.J. Durrett, Memorandum of Interview with A.R. Dochez, Dr. Cook, November 28, 1938; J.J. Durrett to M.A. Blankenhorn and to W.C. Davison, December 9, 1938. [All RG 88. 69A-2099. Box 1. NDA 90. vol. 1. W-NRC.]

<sup>48</sup> Herbert E. Stokinger, Ph.D. Absorption, Acetylation and Excretion of 2-sulfanilamide (Dagenan, M & B 693), November 11, 1938 and Hans Molitor, Preliminary Report . Toxicity of 2-sulphanilyl-amino-pyridine, November 28, 1938. RG 88. 69A-2099. Box 1. NDA 90. vol. 1. W-NRC. Stokinger was a biochemist in Michael Heidelberger's lab at Columbia; Molitor worked for Merck.

<sup>49</sup> J.J. Durrett, Memorandum of Interview with Perrin H. Long and E. Kinnerly Marshall [Johns Hopkins], December 5, 1938. RG 88. 69A-2099. Box 1. NDA 90. vol. 1. W-NRC. Marshall, a supporter of the 1938 act, expressed himself pleasantly surprised with the FDA's philosophy.



doing from one investigator to another."<sup>50</sup> Some researchers reported extensive vomiting and nausea among their patients; others did not. The differences might be due either to variations in manufacturing routine, or to differences in the dosages used by different investigators.<sup>51</sup> Before approving the drug, the FDA would require additional clinical data, a demand to which the manufacturer readily agreed.<sup>52</sup>

Some of the clinical investigators following the drug's progress were not as patient. With the next pneumonia season imminent, and the data favorable to sulfapyridine mounting, several researchers began to urge release to the medical profession at large, if not to the lay public.<sup>53</sup> Others placed less confidence in their peers: even the

<sup>50</sup> J.J. Durrett, Memorandum of interview with Dr. Joseph Rosin [Merck] and D.W. Richards [Columbia University], December 1, 1938; in Durrett's view, "there was no way to dispute the value of this drug." See J.J. Durrett, Memorandum of Interview with Perrin H. Long and E. Kinnerly Marshall [Johns Hopkins], December 5, 1938. [All references RG 88. 69A-2099. Box 1. NDA 90. vol. 1. W-NRC.]

<sup>51</sup> J.J. Durrett, Memorandum of Interview with Perrin H. Long and E. Kinnerly Marshall [Johns Hopkins], December 5, 1938. RG 88. 69A-2099. Box 1. NDA 90. vol. 1. W-NRC. By the end of the month, it appeared as if manufacturing variations did not account for the variation in such reactions: J.J. Durrett to Perrin H. Long, December 22, 1938. RG 88. 69A-2099. Box 1. NDA 90. v. 2. W-NRC.

<sup>52</sup> On being informed of the FDA's position, the firm's representative expressed their "sympathy with it." See J.J. Durrett, Memorandum of interview with Dr. Joseph Rosin [Merck] and D.W. Richards [Columbia University], December 1, 1938. RG 88. 69A-2099. Box 1. NDA 90. vol. 1. W-NRC. By this time, an additional application had been filed by Calco Chemical Co. RG 88. 69A-2099. Box 2. NDA 160. W-NRC. Three additional firms filed in January: see RG 88. 69A-2099. Box 3. NDAs 422 (E.R. Squibb), and 469 (Sharp & Dohme), and 68A-1292. Box 1. NDA 476 (Abbott Laboratories).

<sup>53</sup> J.J. Durrett, Memorandum of Interview with E.K. Marshall and Perrin H. Long, December 30, 1938. RG 88. 69A-2099. Box 1. NDA 90. vol. 2. W-NRC. By the end of the month, Marshall and Long had reversed their position of December 5, and become advocates for immediate release.

medical profession might abuse the drug if it was released before more was known about appropriate dosage.<sup>54</sup> Some feared that sulfapyridine's premature release might lead physicians to abandon a proven remedy, serum therapy, even for those cases where it was best suited.<sup>55</sup>

Faced with conflicting advice, the FDA took its dilemma to the leaders in the field. In a series of interviews, they polled the group of investigators who had been working with sulfapyridine.<sup>56</sup> The question posed by the regulators was the one which had been addressed to them: should the drug be released now for use by physicians, or should its distribution be limited to qualified investigators while additional research continued? The answers they got depended largely on how the question, and the legislation, was read.

Proponents of delay were concerned about having the drug released to general practitioners before sufficient information was available on its safe and optimal use. A few months delay would have several advantages: 1) additional data on the use of sulfapyridine would be accumu-

---

<sup>54</sup> The question of finding appropriate dosage was compounded by the difficulties of maintaining effective therapeutic concentrations, absent knowledge of how the drug was absorbed and excreted, and by conflicting data on toxicity and absorption from animal studies. See the discussion cited in note 57, infra.

<sup>55</sup> Some, but not all, of this opposition came from advocates of serum therapy, including Rufus Cole, who had initially developed the serum treatment. For Cole's position on sulfapyridine, see John L. Rice [N.Y.C. Commissioner of Health] to Thomas Parran, [U.S. Surgeon-General], December 22, 1938, and the accompanying resolution of N.Y. State's Advisory Committee on Pneumonia Control. RG 88. 69A-2099. Box 1. NDA 90. vol. 2. W-NRC.

<sup>56</sup> W.G. Campbell to M.A. Blankenhorn, January 5, 1939. For a list of those surveyed, see W.G. Campbell to Paul Leech, [AMA Council on Pharmacy and Chemistry], January 5, 1939. RG 88. 69A-2099. Box 1. NDA 90. vol. 2. W-NRC.

lated, enabling physicians to assess its merits in specific types of pneumonias; 2) by the end of the pneumonia season, sufficient reports of the drug's side effects would be in print to chasten and instruct physicians' use of the drug; 3) more would be known about safe and effective dosage, an issue which was growing increasingly complex.<sup>57</sup> Some researchers feared losing the opportunity to study the drug on sufficient numbers of patients if it was released immediately:

"Putting this drug on the market right now will make it difficult to get the data which will make it possible to evaluate the action of this drug alone and in combination with serum...." <sup>58</sup>

By contrast, advocates of early release thought that the FDA already had enough data to act. The ultimate determination of the new drug's toxicity, like that of sulfanilamide, would be a long time in

<sup>57</sup> Walter Grady Reddick to W.G. Campbell, January 13, 1939; Charles McKhann to W.G. Campbell, January 7, 1939. A.R. Dochez, one of the more careful and reflective investigators, while he did not oppose immediate distribution, noted two intellectual/practical problems with the current state of knowledge: 1) because individuals reacted to the drug differently, and acetylation rates between individuals were highly variable, controlling dosage required measuring actual serum concentrations directly at frequent intervals, rather than simply regulating the amount administered by units; 2) the East Coast was experiencing a wave of non-pneumococcal pneumonia on which the drug was useless, and at present "we find ourselves considerably confused from a clinical and laboratory standpoint, with what kind of pneumonia we are dealing and a decision whether or not to use the drug becomes difficult." A.R. Dochez to W.G. Campbell, January 11, 1939. For other reports of variability in acetylation/absorption of drug see the remarks of Jesse M. Bullowa and H.E. Stokinger, reported at the New York Academy of Medicine meeting January 17, 1939. C.A. Hermann [Chief Eastern District] to Food and Drug Administration, January 18, 1939. [All references RG 88. 69A-2099. Box 1. NDA 90. vol. 2. W-NRC.]

<sup>58</sup> John T. Cain and R.W. Weilerstein, Memorandum of Interview with Norman Plummer and Dr. Henning, and Memorandum of Interview with Dr. Colin McLeod, January 27, 1939. RG 88. 69A-2099. Box 1. NDA 90. Vol. 2. W-NRC.

coming. For instances like the treatment of pneumonia, where sulfanilamide had little effect, the merits of sulfapyridine were already established.<sup>59</sup> The press was already convinced. The New York Times found the prospect of saving money as well as lives irresistible: "before long we shall swallow tablets of a complex chemical instead of resorting to expensive injections of serum and thus deal with all 32 types of pneumonia."<sup>60</sup> Opposition to release, several investigators charged, was simply an effort on the part of specialists in serum treatment to protect their investment in knowledge and equipment for pneumococcal typing and serum production.<sup>61</sup> To accumulate more data without excessive delay, the FDA contacted the two principal manufacturers for a list of all investigators to whom they had distributed the drug. On February 1, 1939, they began surveying an additional 45 physicians

---

<sup>59</sup> J.J. Durrett, Memorandum of Interview with E.K. Marshall and Perrin H. Long, December 30, 1938. RG 88. 69A-2099. Box 1. NDA 90. vol. 2. W-NRC. Long urged release despite his speculation that the more serious toxicities experienced with sulfanilamide—hemolytic anemia and agranulocytosis—would eventually show up in sulfapyridine, even though he had not yet seen any cases. P.H. Long to David Bryce, January 13, 1939. RG 88. 69A-2099. Box 2. NDA 160. W-NRC. See also I.S. Ravdin to J.J. Durrett, January 4, 1939 and January 9, 1939; Perrin H. Long to W.G. Campbell, January 9, 1939; John T. Cain and R.W. Weilerstein, Memorandum of Interview with Evan Evans, January 27, 1939. RG 88. 69A-2099. Box 1. NDA 90. vol. 2. W-NRC. For one drug company's representation, see David A. Bryce to W.G. Campbell, January 25, 1939. RG 88. 69A-2099. Box 2. NDA 160. W-NRC.

<sup>60</sup> "The Fight Against Pneumonia," New York Times January 19, 1939, 18. See also J.D. Ratcliff, "Death to the Killer," Colliers (December 24, 1938), 18, 52-53.

<sup>61</sup> R.W. Weilerstein and John T. Cain, Memorandum of Interview with M.H. Dawson and Memorandum of Interview with Harold Thomas Hyman, January 27, 1939. RG 88. 69A-2099. Box 1. NDA 90. vol. 2. W-NRC.

having research experience with the drug.<sup>62</sup> More reason for caution began to emerge: even under current arrangements, which restricted the drug to investigational use, it was hard to confine its uses to pneumococci. Preliminary reports of its use in gonococcal infections were likely to trigger premature and uninformed applications unless the FDA somehow held back the floodgates: surely there was some way, researchers wrote, to release it for pneumonias only, or perhaps limit distribution to "qualified clinicians in many centers ... as an intermediate step".<sup>63</sup> Holding on to the drug until the spring might cost the lives of a few patients, but many more would benefit in the long run by the knowledge obtained of how best to use the drug. The greatest benefit, advocates of serum treatment urged, would come from controlled evaluations on large numbers of patients, by experienced investigators willing to alternate patient assignment between serum and

---

<sup>62</sup> John T. Cain and R.W. Weilerstein, Memorandum of Interview with J.M. Carlisle, Mr. Anderson, Miss Person [Merck], January 25, 1939; and John T. Cain and R.W. Weilerstein, Memorandum of Interview with David A. Byrce [Calco] January 25, 1939. Additional letters went out the following week. RG 88. 69A-2099. Box 1. NDA 90. vol. 2. W-NRC. See this file for a list of investigators contacted: a virtual who's who of clinical investigation and infectious disease. Merck and Calco were the only firms to submit substantial amounts of original clinical and laboratory data: the others relied on published reports. See RG 88. 69A-2099. Box 3. NDAs 422 (E.R. Squibb), and 469 (Sharp & Dohme), and 68A-1292. Box 1. NDA 476 (Abbott Laboratories). W-NRC.

<sup>63</sup> M.A. Blankenhorn to W.G. Campbell, February 2, 1939; Hugh Morgan to W.G. Campbell, February 6, 1939; Memorandum of interview with Harris S. Johnson, February 3, 1939; O.H. Robertson to W.G. Campbell, February 7, 1939; David D. Rutstein to W.G. Campbell, February 7, 1939; W.H. Carroll to W.G. Campbell, February 9, 1939; L.H. Schmidt to W.G. Campbell, February 9, 1939. RG 88. 69A-2099. Box 1. NDA 90. vol. 3. W-NRC.

sulfapyridine.<sup>64</sup> The AMA's Council on Pharmacy and Chemistry and M. Fishbein, editor of JAMA, each counseled patience while additional research was completed under suitable conditions.<sup>65</sup>

Whatever the reservations of skeptics, press reports accentuated the positive. News of favorable investigations reached the public overnight.<sup>66</sup> While continuing to gather data, the FDA was coming under pressure from friend and foe alike to release sulfapyridine. By February, E.K. Marshall, who had initially endorsed the agency's caution about sulfapyridine, wrote:

<sup>64</sup> John T. Cain and R.W. Weilerstein, Memorandum of Interview with Jesse G.M. Bullowa, and Dr. Holle, January 26, 1939. RG 88. 69A-2099. Box 1. NDA 90. vol. 2. W-NRC. To guard against the multiple problems of spontaneous remission, differential response to treatment by age, seasonal variability infectiousness and investigator bias, Bullowa wanted to study a minimum of thirty persons for each of several age groups, over the course of at least one pneumonia season, with automatic treatment assignment to serum, drug, or serum plus drug to guard against investigators biasing treatment selection. See Jesse G.M. Bullowa, Norman Plummer and Maxwell Finland, "Sulfapyridine in the Treatment of Pneumonia," JAMA 112 (February 11, 1939), 570 and Maxwell Finland to W.G. Campbell, February 8, 1939. RG 88. 69A-2099. Box 1. NDA 90. vol. 3. W-NRC.

<sup>65</sup> Council on Pharmacy and Chemistry, "Preliminary Report of the Council," JAMA 112 (February 11, 1939), 538; "Sulfapyridine—The New Sulfanilamide Derivative," JAMA 112 (February 11, 1939), 541. The Council's recommendation was a curious one: Long, the author of the Council's report, was by this time urging release, but the other members felt that the drug should be retained on "experimental status". See Council on Pharmacy and Chemistry, Bulletin, January 18 and 25, 1939, and Paul Nicholas Leech to W.G. Campbell, February 1, 1939, RG 88. 69A-2099. Box 1. NDA 90. v. 3. W-NRC. Fishbein was apparently collecting information on sulfapyridine from his own sources, who reported both "spectacular results" and the absence of any opportunity to conduct the necessary controlled observations as yet. See William S. Middleton to Morris Fishbein, March 11, 1939. William S. Middleton papers, Box 8. MS C 206. National Library of Medicine.

<sup>66</sup> "New Drug is Hailed as Pneumonia Cure," New York Times (January 18, 1939), 21.

I was very much interested and did what little I could to promote the passage of the present Food, Drug and Cosmetic Bill. I do not want to feel that I was mistaken in doing this but cannot help occasionally wondering if the present lack of the drug for seriously ill patients is not worse than operating under the old Food and Drug Act.<sup>67</sup>

As the 6 month deadline for acting on the original application approached, additional calls to release the drug came from former skeptics.<sup>68</sup> Reviewing the evidence from months of interviewing and data collection, Theodore Klumpp, by then Chief of the Drug Division, concluded that the majority of investigators consulted approved of release:

While a few investigators recommended that the drug be withheld from the market, such recommendations upon analysis do not appear to rest upon considerations of the intrinsic safety or danger of the drug. Principally those workers were concerned with the orderly development of medical scientific knowledge concerning the therapeutic efficacy of the drug and the relation of this drug to other available forms of therapy in pneumonia.

While these are important considerations from the point of view of research and medicine, they do not constitute, in our judgment, a substantial basis for withholding this application under the provisions of section 505.<sup>69</sup>

On March 9, 1939, letters went out to each of six firms, indicating that the FDA would not deny their applications for sulfapyridine, provided the manufacturer ensure, through labeling and advertising, that the drug be used "under close, continuous observation of a

<sup>67</sup> E.K. Marshall to W.G. Campbell, February 3, 1939. RG 88. 69A-2099. Box 1. NDA 90. vol. 3. W-NRC.

<sup>68</sup> The New York State and New York City Advisory Committees for Pneumonia Control, which in December had opposed release, endorsed release of the drug in late February. Russell L. Cecil to Theodore G. Klumpp, February 27, 1939. RG 88. 69A-2099. Box 1. NDA 90. vol. 3. W-NRC. The change of heart was noted at the time by the FDA's Klumpp. See his Memorandum for Mr. Campbell. February 23, 1939. RG 88. 69A-2099. Box 1. NDA 90. Vol. 3. W-NRC.

<sup>69</sup> Theodore G. Klumpp, Memorandum for Mr. Campbell. February 23, 1939. RG 88. 69A-2099. Box 1. NDA 90. vol. 3. W-NRC.

qualified practitioner of medicine." It was the manufacturer's responsibility to warn physicians of the drug's toxicity, and instruct them in managing such cases.<sup>70</sup> But neither the FDA nor the industry could take responsibility for the likelihood that sulfapyridine "will undoubtedly be abused by the unwise and ill-informed whenever it is put on the market." This was a "problem of medical practice" and not of drug regulation.<sup>71</sup>

#### A System of Drug Regulation: Conjectures and Refutations

This was the first time in the history of American medicine that I know of where it was possible for the medical profession to have opportunity to be informed about a new drug before the detail man was around at his door importuning him to use a drug about which he knew little.<sup>72</sup>

In their handling of the sulfapyridine case, FDA officials established procedures and enunciated a philosophy of drug regulation which, with minor changes, would govern the agency's behavior and that of the regulated industry for the next eight years. The law mandated the FDA to prevent the sale of "unsafe" drugs. But the idea of an unsafe drug covered a multitude of sins. The toxicity of some drugs could be readily established in the laboratory. The standards imposed by the FDA's toxicologists enabled the agency to screen out demonstrably

---

<sup>70</sup> W.G. Campbell to Dr. Joseph Rosin [Merck & Co], March 9, 1939. RG 88. 69A-2099. Box 1. NDA 90. v. 3. W-NRC.

<sup>71</sup> Theodore G. Klumpp, Memorandum for Mr. Campbell, February 23, 1939. RG 88. 69A-2099. Box 1. NDA 90. vol. 3. W-NRC.

<sup>72</sup> Paul Nicholas Leech, "Relation of the Food, Drug and Cosmetic Act to the Work of the Council on Pharmacy and Chemistry and Manufacturing Pharmacy," American Drug Manufacturers Association Proceedings 28th Annual Meeting (1938), 119.



unsafe drugs, or compounds whose safety for clinical testing had not yet been established.<sup>73</sup>

Drugs such as sulfapyridine presented a different problem: there was no question that it was toxic and little question either that it was of considerable benefit. Only clinical evidence could determine whether it would do more harm to ban the drug or release it. With nearly 1700 applications filed during the first 18 months of the act, the FDA was in no position to do its own research on each drug.<sup>74</sup> Nor, even if it were possible, did they wish to do so:

In our judgement, the question of safety is so important that it is ordinarily desirable to have a number of independent investigators study the question. From this standpoint, then, we go to the literature and study the reports of investigations contained therein, as well as those submitted by the manufacturer. In many cases these are sufficient to establish a prima facie case of safety and then it is unnecessary to go further. In the instances such as, for instance, sulfapyridine, where the problem is a very difficult one, we obtain from the manufacturer a list of all the investigators who have studied the drug and communicate with them, visit them, go over their records, and on the basis of the sum total of experiments with the drug we are in a much better position to arrive at a correct conclusion than if we made tests ourselves.<sup>75</sup>

---

<sup>73</sup> For an example of the FDA's requirements on toxicological data, see the testimony of Robert P. Herwick, Hearings on Quinimid. 1939, 51-59. RG 88. 58A-277. Box 19. F 505.7. W-NRC. In the first year of operation under the new law, more than half of the applications submitted were withdrawn without prejudice, many of these for insufficient data. See Carl M. Anderson, "The 'New Drug' Section," Food Drug Cosmetic Law Quarterly (March, 1946), 84; Theodore G. Klumpp to Arthur DeGraaf, February 17, 1940. RG 88. 59A-2736. Box 24. F 505.1-508.2. W-NRC.

<sup>74</sup> Theodore Klumpp to Arthur DeGraaf, September 27, 1939. RG 88. 58A-277. Box 38. F 511 (General, September-December). W-NRC.

<sup>75</sup> Theodore Klumpp to Arthur DeGraaf, September 27, 1939. RG 88. 58A-277. Box 38. F 511 (General, September-December); testimony of Theodore Klumpp, Transcript of Hearing, Quinimide Corporation. 1939, 228-230. RG 88. 58A-277. Box 19. F505.7; Walton Van Winkle, Jr. to William Deichmann, July 6, 1944. RG 88. 59A-2736. Box 166. F 505.1-505.6. [All references W-NRC.]

The work of one group of scientists, no matter how eminent, was always subject to error. The scientific community was far less likely to err. To evaluate drugs like sulfapyridine, the physicians responsible for administering the 1938 legislation turned to their peers and mentors.

In considering the applications for sulfapyridine, the FDA reviewed over 2,000 cases.<sup>76</sup> For drugs known to produce reactions, like the sulfonamides or the arsenicals used in the treatment of syphilis, "a very large series of cases is necessary" for weighing the severity and frequency of hazards against the therapeutic benefits obtained.<sup>77</sup> But numbers alone were not enough:

Sheer volume of clinical reports or large numbers of cases are not sufficient in themselves to be decisive. Attention must be directed to the character of the investigations and the quality of the investigators.<sup>78</sup>

"Qualified investigators" provided the most valuable data: "We are always delighted to receive the kind of well-controlled, qualified reports that emanate from our better institutions."<sup>79</sup> To those whose

<sup>76</sup> Theodore G. Klumpp, Memorandum for Mr. Campbell, February 23, 1939. RG 88. 69A-2099. Box 1. NDA 90. vol. 3. W-NRC.

<sup>77</sup> J.J. Durrett to Bruce Webster, February 6, 1940; RG 88. 59A-2736. Box 24. F 505.1-508.2, W-NRC. Interestingly, only a handful of the 280 individuals who received "investigational" shipments of sulfapyridine treated more than 100 cases. Of the 98 investigators reporting results to the FDA, only 20 based their arguments on more than 35 patients. As noted above, obtaining multiple "takes" on the data appears to have been more important than the number of cases seen by any one group.

<sup>78</sup> Walton Van Winkle, Jr. to Paul Dunbar and Robert Herwick, January 30, 1946. RG 88. 59A-2736. Box 220. F 505.1074-509. W-NRC.

<sup>79</sup> Theodore G. Klumpp to Arthur DeGraaf, October 28, 1939. RG 88. 58A-277. Box 38. F 511 (General, September-December). W-NRC. FDA physicians were at pains to convince reluctant specialists that the problems of therapeutic evaluation deserved their attention. Theodore G. Klumpp to Fuller Albright, February 4, 1939. RG 88. 58A-277. Box 22. F512.1.10-512.6.31. W-NRC.

work showed little appreciation of experimental rigor, the FDA's scientists emphasized the basic elements of therapeutic research: the importance of controls and the need for a sufficient number of cases.<sup>80</sup> In this, they shared the views of the Council on Pharmacy and Chemistry:

...the word 'control' isn't the only thing. You have to have fair controls and adequate controls, just as much so in clinical work as in pharmacological work.<sup>81</sup>

But good methods alone did not provide good decisions. Where a detailed review of the available evidence did not produce a clear recommendation, then the strategy developed in the sulfapyridine case was called for:

Choosing those investigators with the greatest experience, or those known to be critical in their approach to investigative problems, personal visits should be made to these men and the doubtful points discussed."<sup>82</sup>

Francis Blake's endorsement of sulfapyridine, for example, was "significant" because Blake was "recognized throughout the country as one of the most critical and conservative therapeutists."<sup>83</sup>

<sup>80</sup> On the importance of controls, see the testimony of Sanford M. Rosenthal, B.J. Vos and Ernest King, Transcript of Hearing. Teotisol Labs Corporation. 1940, 56-58, 339, 349-350; testimony of George Bobbs, Transcript of Hearings. The Dominion Laboratories. 'Sunica'. 1940. Both transcripts in RG 88. 59A-2736. Box 24. W-NRC. See also Walton Van Winkle, Jr. to Paul Dunbar and Robert Herwick, January 30, 1946. RG 88. 59A-2736. Box 220. F 505.1074-509. W-NRC.

<sup>81</sup> Paul Nicholas Leech, "Relation of the Food, Drug and Cosmetic Act to the Work of the Council on Pharmacy and Chemistry and Manufacturing Chemistry," American Drug Manufacturers Association. Proceedings. 28th Annual Meeting. (1939), 121.

<sup>82</sup> Walton Van Winkle, Jr. to Paul Dunbar and Robert Herwick, January 30, 1946. RG 88. 59A-2736. Box 220. F 505.1074-509. W-NRC.

<sup>83</sup> Theodore G. Klumpp, Memorandum of interview with Francis G. Blake, February 2, 1939. RG 88, 69A-2099, Box 1, NDA 90, vol. 3. W-NRC. Klumpp came to the FDA from Yale, where Blake had been a senior colleague. E.K. Marshall made a similar remark about the acuity of the clinical observations made by his colleague, Perrin Long. Ernest Q. King, Memorandum of Interview with E.K. Marshall, Jr., April 26, 1941. RG 88. 59A2736. Box 71. F 511.07 (January-July). W-NRC.

Translating the convictions of academic medicine into directives for regulatory policy was not always an easy matter. In the absence of formal criteria, determinations of safety were bound to be complex judgments, dependent upon the beliefs of those consulted as well as the available data. In deciding when to release sulfapyridine, the key question was whether to delay approval until the research community could establish a scientific basis for rational use of the drug. One group of experts felt that thorough therapeutic evaluation of sulfapyridine, like that of any drug, was a matter of years, not months. In the meantime, individual physicians would be better off learning to live with a degree of uncertainty about the drug's toxic effects. Provided they used the drugs intelligently, physicians need not fear harming their patients.<sup>84</sup> Other researchers held that even a brief delay would produce valuable information about the safest and most appropriate use of the drug, benefiting patients and practitioners alike.

Those most reluctant to release the drug immediately were especially skeptical about the ability of general practitioners to use drugs intelligently and safely. Their comments indicate a desire that the FDA not only anticipate but prevent the potential misuse of new drugs by the profession at large. The question was, how far could the FDA

---

<sup>84</sup> See the general remarks of E.K. Marshall, Jr. on the recent tendencies to abandon "time honored slow and laborious method[s] of reaching conclusions..." "Results must be obtained over night and reputations made (or blasted) in a month". E.K. Marshall, Jr., "An Unfortunate Situation in the Field of Bacterial Chemotherapy," JAMA 112 (January 28, 1939), 352-353. On the need for physicians to accustom themselves to managing the toxic effects of the sulfa drugs, see Perrin H. Long and James Haviland, "A Clinical Evaluation of the Use of Sulfanilamide, Sulfapyridine and Sulfathiazole in the Treatment of Bacterial Infections," American Drug Manufacturers Association, Proceedings. 29th Annual Meeting (1940), 91.

go in this direction? Klumpp and his colleagues decided they had gone far enough. In determining drug safety, the FDA would require proof of a drug's clinical value and insist that manufacturers pay careful attention to the therapeutic claims which accompanied their products. Such literature would, where possible, instruct physicians in when and how to use a drug.<sup>85</sup> Sulfapyridine was approved for the treatment of pneumonia, and nothing else. Proposed new uses would require a new review.<sup>86</sup> But whether or not physicians went on to use the drugs as they were intended was not the FDA's problem.

#### The Burden of Regulation

In translating the statutory language of the 1938 drug law into a viable system of drug regulation, the FDA relied heavily on the cooperation of three groups: the regulated industry, the

---

<sup>85</sup> The FDA frequently monitored the nature and extent of therapeutic claims made on behalf of products, though apparently with limited effect in some cases. See J.J. Durrett to Hoffman La Roche, April 20, 1940 and Ernest Q. King, Memorandum for Dr. Durrett, June 4, 1940. RG 88. 68A-1292. Box 1. NDA 776; Ernest Q. King, Memorandum to Dr. Herwick, May 26, 1941. RG 88. 80-22. Box 1. NDA 3726. Vol. 1; R.P. Herwick to Lederle Laboratories, September 29, 1942. RG 88. 59A-2736. Box 107. F 511.07 (August); Walton Van Winkle, Jr. to Lederle Laboratories, Inc., April 13, 1943. RG 88. 59A-2736. Box 142. F 511.07 (January-April). [All references W-NRC.]

<sup>86</sup> When sulfapyridine was subsequently presented for use in treating gonorrhoea, manufacturers were asked to provide "supporting scientific reports" for the new claims, although not as extensive as for the initial review. FDA officials were, if anything, more concerned, however, that the average physician lacked the information to use the drug safely for the new indication. It was up to the manufacturer to provide that information. George W. Merck to J.J. Durrett, March 13, 1940 and J.J. Durrett to Paul N. Leech, March 28, 1940. RG 88. 69A-2099. Box 1. NDA 90. v. 3. W-NRC.

medical-scientific community and practicing physicians. The FDA's success in obtaining this cooperation depended on two conditions:

1) that each group shared the assumptions on which the FDA premised its philosophy of drug regulation and 2) that the actions to which this philosophy led did not conflict with the more fundamental desires and purposes of any group.

On the first condition, there was widespread agreement. All three groups accepted that the safety of new drugs could not be determined in the abstract but depended, as the statute indicated, on the way in which a drug was meant to be used. The viability of the FDA's approach to drug regulation therefore depended ultimately on the costs it imposed on each of the affected groups.

The medical-scientific community, which originated the principles under which the FDA operated, found little to quarrel with in the agency's adoption of their credo.<sup>87</sup> FDA officials expected research physicians to divert energy and attention from their scientific work to provide data and advice on the therapeutic consequences of new drugs.

---

<sup>87</sup> By and large, dissents from this community regarding the FDA's policies are confined to technical disagreements about specific decisions or rules of evidence. See, for example, Hunter F. Kennedy to Perrin H. Long, August 27, 1942. RG 88 59A-2736. Box 107. F 511.07 (August). W-NRC. The most substantial criticism of the FDA's practices I have found is the commentary of several prominent toxicologists that the FDA's approach to toxicity testing may be too standardized and retarding the appearance of new drugs on the market. Chauncey D. Leake, Raymond Gregory, Paul L. Ewing and George A. Emerson, "Appraisal of New Drugs," JAMA 127 (January 27, 1945), 244. To my knowledge, this letter represents the first appearance in print of the induced "drug lag" hypothesis which has achieved ideological popularity in recent years. In practice, the FDA's approach to New Drug Application review appears to have been far less standardized than the critique of Leake, et. al., implies. See the letter of Walton Van Winkle, Jr. to Richard K. Richards, December 19, 1944, which discusses many of the issues in the Leake critique. RG 88. 59A-2736. Box 166. F 505.1-505.6. W-NRC.

Some, if not all, members of the research community complied. Much of the work submitted in support of new drug applications, however, did not come from experienced researchers. But lacking defensible criteria to identify the "line of demarcation between ordinary physicians and those obviously devoting a major portion of their time to investigative work," the agency could only judge the work, not the worker.<sup>88</sup> FDA officials complained frequently about the quality of scientific and clinical work presented to them, but apart from educational efforts, did little about it. For the individual physician, producing adequate therapeutic research was a moral, not a legal, obligation: no unpleasant consequences followed for the individuals who elected not to comply.<sup>89</sup>

The FDA's policies could not avoid "impinging" on the practicing

<sup>88</sup> Walton Van Winkle, Jr. to Paul Dunbar and Robert Herrick, January 30, 1946. RG 88. 59A-2736. Box 220. F 505.1074-509; on the moral "obligations" of medical experts, see J.J. Durrett to Stanley P. Reimann, January 28, 1939. RG 88. 58A-277. Box 38. F 511 (General: January-February); Theodore G. Klumpp to Arthur DeGraaf, October 28, 1939. RG 88. 58A-277. Box 38. F 511 (General: September-December). [All references W-NRC.]

<sup>89</sup> For the FDA's assessment of the quality of work submitted in the initial years of operation under the 1938 Act, see Theodore G. Klumpp, [Address], American Pharmaceutical Manufacturers, Proceedings (1941), 51-55; and J.J. Durrett to Arthur C. Degraaf, March 22, 1940. RG 88. 59A-2736. Box 24. F 505.1-508.2. W-NRC.

On the continued inadequacy of the clinical and laboratory data submitted to the FDA, see Walton Van Winkle, Jr. to Richard K. Richards, December 19, 1944. RG 88. 59A-2736. Box 166 F 505.1-505.6. W-NRC; and Food and Drug Administration, Annual Report, 1944 (Washington: Food and Drug Administration, n.d.), 50. While the FDA did not take action against investigators whose work was not up to par, firms which attempted to market drugs under the guise of distributed them for investigative purposes were considered fair game. Los Angeles District, Division of Regulatory Management, New Drugs. Memorandum. Spicer-Gerhart Company, July 27, 1950. RG 88. 63A-292. Box 55. F 501.2-510. W-NRC.

physician.<sup>90</sup> But the method they chose was quite indirect: by monitoring the information practicing physicians received about new drugs, they hoped to improve the quality of therapeutic practice. The more direction the FDA's reviews could provide about the relative merits of different drugs, the safer the therapeutic practices which might be observed. But for long-term assessments of therapeutic merit, the FDA deferred to the scientific community and the AMA's Council on Pharmacy and Chemistry. By and large, the FDA offered no challenges to physicians who used drugs without regard to agency approved indications or the recommendations of the profession's scientific leadership.<sup>91</sup> Improving therapeutic practice was, in the final analysis, a task for the profession and not the government. The FDA's emphasis on regulating drugs, not medical practice, placed the heaviest burden of regulation on the industry.

---

<sup>90</sup> Walter G. Campbell, "Administration of New Federal Food, Drug and Cosmetic Law," Oil, Paint and Drug Reporter 134 (October 31, 1938), 41.

<sup>91</sup> For evidence of the continuing scepticism of FDA officials towards the scientific awareness and therapeutic capabilities of most physicians, see Ernest Q. King, Memorandum to Dr. Herwick, May 26, 1941. RG 88. 80-22. Box 1. NDA 3726. Vol. 1; P.B. Dunbar, Memorandum of Interview with Dr. Austin Smith, May 26, 1947. RG 88. 59A-2736. Box 234. F. 045.91-046.5. Such concern was behind the FDA's proposal to adopt so-called "prescription labeling" which would enable the FDA to regulate the information manufacturers provided physicians on the appropriate uses of drugs. See J.J. Durrett to the Assistant General Counsel, October 24, 1940. RG 88. 59A-2736 Box 24. F 505.1-508.2. At the same time, the FDA was adamant that the prospect of physicians using drugs "contrary to the recommendations in the labelling is unfortunately a matter which we are not permitted to consider in connection with the new drug applications." Hunter F. Kennedy to Perrin H. Long, August 27, 1942. RG 88. 59A-2736. Box 107. F 511.07 (August). [All references W-NRC.]



Given the costs it imposed on industry, their initial response to the 1938 drug law was surprising. Numerous firms could agree with the representative of Lederle Laboratories who welcomed the FDA's inclination "to tell us things we ourselves didn't fully appreciate about our drug and to improve the directions that were to be put thereon." Although the FDA required firms to provide extensive laboratory and clinical information on the merits (and disadvantages) of new drugs, and restricted the conditions for which those drugs should be marketed, industry generally responded favorably to the burden.<sup>92</sup>

So long as the FDA confined itself to policing the claims which accompanied approved drugs, industry spokesmen acknowledged the FDA's right to consider the benefits, as well as the hazards, of new drugs.

<sup>92</sup> See the remarks of R.S. Childs [Lederle Laboratories] in discussion of J.J. Durrett's "New Drug Application Requisites," Oil, Paint and Drug Reporter (May 22, 1939), 46. For an opposing point of view, see James F. Hoge, "An Appraisal of the New Drug and Cosmetic Legislation From the Viewpoint of Those Industries," Law and Contemporary Problems 6 (Winter, 1939), 121. In general, however, "the trade has accepted [the] decision" to regulate the distribution of drugs by controlling the labeling for the use. "Drug Enforcement Reviewed in Campbell Report," Oil, Paint and Drug Reporter 136 (December 25, 1939), 3. In 1940, both trade associations formed a medical advisory committee to assist the FDA in the determination of "scientific facts" which would aid in the labeling of drugs. The committee, which evidently bit off more than the parent bodies could chew, was asked to curtail its activities substantially the following year. See F.O. Taylor, "Report of the Pharmaceutical Contact Committee," American Drug Manufacturers Association. Proceedings. 30th Annual Meeting (1941), 219-220.

For examples of FDA requests for additional data to accompany new drug applications and/or modifications of therapeutic claims, see J.J. Durrett to Hoffman-LaRoche, Inc., March 26, 1940. 68A-1292. Box 1. NDA 776; Walton Van Winkle to R.J. Strassenbrugh, March 11, 1943. 59-A-2736. Box 142. F. 511.07 (January-April). W-NRC. Some firms, such as Merck & Co., anticipated the FDA's requirements by organizing the lines of investigation to be pursued by research physicians testing new products, and giving them guidance in how to prevent unauthorized use of experimental drugs within their institutions. J.M. Carlisle to L.W. Tucker. February 16, 1939. RG 90. PHS General Classified Records. 1936-1944. Box 57. NA.

But when the FDA refused to approve new products, on the grounds that they were without therapeutic merit, individual firms and industry representatives began challenging not only the FDA's decisions in individual cases, but the premises on which the FDA's entire approach to drug regulation was based. In the post war period, the FDA's right to examine the therapeutic value of new drugs came increasingly under attack.<sup>93</sup>

So long as regulators and regulated shared the same beliefs, the FDA could make case decisions which seemed both valid and just. Once those principles were challenged, FDA officials looked to higher authority to define the rules of the game. Reviewing eight years of operation under the new law, the chief of the FDA's drug division concluded that "it would be of great assistance ... if there were a background of judicial opinion" regarding the weight to be accorded

---

<sup>93</sup> Any confrontation with industry was postponed until 1940 when the FDA's legal authority to move against companies which made exaggerated therapeutic claims on existing products kicked into effect. After 1940, the FDA no longer had to prove fraudulent intent to regulate such claims. "Drug Law Enforcement Reviewed in Campbell Report," Oil, Paint and Drug Reporter 136 (December 25, 1939), 42. One of their initial moves came against the labeling of ovarian substances: "physicians who use such products are entitled to know that there is no scientific evidence that such products when administered orally possess any therapeutic activity." In the view of the trade press, the FDA's dictum was tantamount to telling "the manufacturer that he must undertake to educate the physicians." "Ovarian Substance Abused By F.D.A.," Oil, Paint and Drug Reporter 138 (December 30, 1940), 3. The real confrontation did not come, however, until 1948 when the FDA decided that no labeling for inert glandular substances could satisfy the terms of the law, since it was impossible to write instructions for the safe use of a drug which was without effect. Dunn and Kleinfeld, Federal Food, Drug and Cosmetic Act (n. 28), 756-757. Industry responded by defending the physician's right to ignorant bliss, a prerogative it has since continued to aid and abet. W.W. Wheeler, Jr. "Interference With the Practice of Medicine," Food, Drug and Cosmetic Law Quarterly (September, 1948), 365-375.

efficacy in considering "the 'relative safety' of a new drug" and the role considerations of "relative efficacy" could play in assessing newer products.<sup>94</sup> The industry's challenge did not alter the intellectual commitment of agency officials to the principle of efficacy but they reinforced officials' perception that considerations of efficacy could not be given the deciding weight in reviewing new drug applications.<sup>95</sup> The FDA's circumspection meant that in the postwar period, reformers would turn once again to the profession's scientific elite to take the lead in reforming therapeutic practice.

---

<sup>94</sup> Walton Van Winkle, Jr. to Paul Dunbar and Robert Herrick, January 30, 1946. RG 88. 59A-2736. Box 220. F 505.1074-509. W-NRC.

<sup>95</sup> By 1950, Van Winkle's successor, Robert Stormont, reported that the FDA no longer felt it possible to consider evidence of efficacy in ruling on new drugs. While Stormont's account is probably an exaggeration of actual practice, it suggests a growing consciousness on the FDA's part that they lacked the authority to make give primacy to considerations of efficacy. The FDA's desire to seek new statutory authority for efficacy decisions probably stems in part from their changed perception of the legal situation. See Robert T. Stormont, "Our Mutual Responsibilities in the Regulatory Control of Drugs," American Drug Manufacturers Association Proceedings 38th Annual Meeting (1950), 139-144. [Stormont had left the FDA at the time of this address.]

Chapter Three. Memories of Underdevelopment.  
The Decline and Rise of Cooperative Research

Both professional and government regulation of therapeutic practice presumed a large, active community of well qualified clinical investigators, interested in problems of therapeutic practice. For the first third of this century, only a handful of individuals were willing to commit themselves to an area which seemed on the margins of medical science. In their efforts to improve the quality of therapeutic research, medicine's scientific leadership encouraged the organization of cooperative studies, investigations in which specialists from numerous institutions cooperated in the evaluation of new treatments. The scientific problems these studies were thought to address, and the organizational difficulties researchers encountered in completing such studies successfully, are the subject of the present chapter, which examines the development of the tradition of cooperative studies up to the end of World War II.

The chapter begins by examining the institutional context of medical research in which ideas about cooperative studies developed, and follows with an analysis of a pioneering venture in cooperative research, the Cooperative Clinical Group's study of syphilis treatments. In the second half of this chapter, I examine the contribution wartime research policy made to the theory and practice of cooperative studies. The political and intellectual dilemmas of cooperative studies in the postwar period are analyzed in the following chapter.

## I: Therapeutic Research, 1900-1940

The social development of therapeutic research parallels that of the other medical sciences. The laboratory branches—pharmacology and toxicology—were first to develop. Between 1920 and 1940, the quality of pre-clinical studies on drug action and toxicity improved substantially, as the number of trained researchers and opportunities to conduct pharmacological research increased.<sup>1</sup> In contrast, opportunities for clinical therapeutic research lagged behind.

In recent years, historians have begun to examine sympathetically the situation of researchers interested in clinical problems, who complained repeatedly of the special ideological and institutional obstacles they encountered in pursuing their work.<sup>2</sup> For those physicians

---

<sup>1</sup> On the institutional and intellectual development of research in pharmacology and toxicology, see Ko Kwei Chen, ed. American Society for Pharmacology and Experimental Therapeutics, Inc. 1908-1969 (Washington: Judd and Detweiler, 1969) and John Patrick Swann, The Emergence of Cooperative Research Between American Universities and the Pharmaceutical Industry, 1920-1940. Ph.D. Dissertation. University of Wisconsin. 1985. I am indebted to Dr. Swann for making a copy of his thesis available to me. Certain institutions—the Rockefeller Institute, the University of Wisconsin, the Johns Hopkins University—provided critical loci for the development of research work and the training of others. Nonetheless, as late as 1941, Walter B. Cannon observed that one-third of U.S. medical schools still lacked departments of pharmacology; see his "Problems Confronting Medical Investigators," Science 94 (August 22, 1941), 176.

<sup>2</sup> On the problems of clinical research, see the contemporary perspectives of Samuel J. Meltzer, "The Science of Clinical Medicine: What It Ought To Be and the Men to Uphold It," JAMA 53 (1909), 508-512; and Alfred E. Cohn, "Purposes in Medical Research: An Introduction to the Journal of Clinical Investigation," Journal of Clinical Investigation 1 (October, 1924), 1-11. For a historical perspective on these developments, see Russell C. Maulitz, "'Physician Versus Bacteriologist': The Ideology of Science in Clinical Medicine," in Morris J. Vogel and Charles E. Rosenberg, eds., The Therapeutic Revolution (Philadelphia: University of Pennsylvania Press, 1979), 91-109; see also R.S. Cunningham, "The Organization of Research in Clinical and Preclinical Departments," Southern Medical Journal 26 (July, 1933), 615-620, who makes it clear that the problems of clinical researchers with the social organization of medical schools were not unique.

who sought not merely to transport the accomplishments of the laboratory to the bedside, but to translate the concepts and methods of the laboratory sciences into practices appropriate for investigating human disease, the peculiar organization of medical research seemed ill-suited.

During the first two decades of the present century, medical research flourished in a variety of institutions. Research institutes vied with government laboratories, specialty clinics and universities to provide settings suited to the development of medical knowledge. This organizational diversity permitted researchers to experiment with various strategies for linking the scientific work of the laboratory with the problems seen in the clinic. By 1920, medical schools and their affiliated hospitals had consolidated their position, hitherto weak, by becoming the principal foci for medical research. The subsequent hegemony of medical schools further exacerbated the difficulties of clinical researchers.<sup>3</sup>

---

<sup>3</sup> The public health laboratories of New York City and the Hygienic Laboratory in Washington, subsequently the National Institutes of Health, provide paradigmatic examples of medical research enterprises subsidized by government. The Rockefeller Institute was both the best known and most influential of free standing research institutes organized on the European model, while both the Mayo Clinic and George Crile's clinic in Cleveland provide examples of nationally known centers for surgical treatment and research. The claim that medical schools achieved hegemony does not mean that such prestigious institutions ceased to function, merely that they no longer provided an organizational model for others to emulate.

On the advantages of research institutes over universities as a site for medical research, see Henry H. Donaldson, "Research Foundations in Their Relation to Medicine," Medical Research and Education (New York: The Science Press, 1913), 474-486. The growing importance of universities as a base for research in one medical science—biochemistry—is nicely documented in Robert E. Kohler, "Medical Reform and Biomedical Science: Biochemistry—A Case Study," in Vogel and Rosenberg, The Therapeutic Revolution (n. 2), 27-66. On the development of the

Clinical investigators working in medical schools had to meet the demands of department chairmen to place service obligations before their research. As physicians, they faced competition from their medical colleagues for income, for patients to study and for the allegiance of their students to a research career. As scientists, they faced intellectual competition from laboratory based specialists whose mastery of the relevant intellectual concepts and methodological tools seemed to exceed their own.<sup>4</sup> Whether clinical investigators were as disadvantaged as they thought with respect to their laboratory colleagues is not clear.<sup>5</sup> But outside of a few isolated research

---

Rockefeller Institute, see George W. Corner, A History of the Rockefeller Institute, 1901-1953 (New York: Rockefeller University, 1965). Victoria A. Harden's Inventing the NIH. Federal Biomedical Research Policy, 1887-1937 (Baltimore: The Johns Hopkins University Press, 1986) is now the standard monograph on the early history of the National Institutes of Health.

<sup>4</sup> On the service demands placed on research clinicians, see Cunningham, "The Organization of Research," (n. 2) 615-616. For the difficulties of recruiting students to research, see W.B. Cannon, "The Career of the Investigator," Medical Research and Education (New York: The Science Press, 113), 295-296. Peter English provides an example of the disdain of some physiologists for the work of their physician colleagues in his Shock, Physiological Surgery and George Washington Crile. Medical Innovation in the Progressive Era (Westport: Greenwood Press, 1980), 121-154.

<sup>5</sup> Both pre-clinical and clinical scientists were inclined to emphasize the special problems of each group, and complain that they were unable to keep pace with the economic and intellectual advances of the other. Yandell Henderson, "Teachers in the Preclinical Sciences," JAMA 74 (May 15, 1920), 1416-1417; Joseph Erlanger, C.M. Jackson, Graham Lusk, et. al., "An Investigation of Conditions in the Departments of the Preclinical Sciences. Report of a Committee of the Division of Medical Sciences of the National Research Council," JAMA 74 (April 17, 1920), 1117-1122. The relative disadvantages of each group apparently varied decade by decade and by locale. Historians have not yet begun to collect the kind of quantitative data on careers in various disciplines that would permit us to adjudicate these claims.

centers, few clinical specialists controlled the financial and organizational resources called for by their research.<sup>6</sup>

For those clinicians interested in questions of therapeutic research, social and intellectual difficulties multiplied. The low status of therapeutics restricted the number of clinical specialists willing to spend precious research time evaluating new drugs. The result was readily apparent in the newer journals of medical science: human studies which went beyond the measurement of drug action to evaluate the clinical outcomes of drug treatment were notable by their rarity.<sup>7</sup>

The lack of qualified investigators and institutions posed a special problem for therapeutic reformers. Not only did the demand for carefully conducted therapeutic evaluations exceed the supply but if specialists disagreed among themselves, information to resolve their

<sup>6</sup> For an overview of medical research at specific centers around the country, see A. McGehee Harvey's invaluable Science at Bedside: Clinical Research in America, 1905-1945 (Baltimore: The Johns Hopkins University Press, 1981). Two exemplary but unusual centers were the Mallinckrodt Research Ward at the Massachusetts General Hospital and the "metabolic laboratory" of Graham Lusk and Eugene Dubois at Bellevue Hospital in New York City on which it was modeled. For an informative, if informal, account of the former, see James Howard Means, Ward 4. The Mallinckrodt Research Ward of the Massachusetts General Hospital (Cambridge: Harvard University Press, 1958.)

<sup>7</sup> Any sampling of the pages of the Journal of Clinical Investigation, the Journal of Laboratory and Experimental Medicine or the annual collections of research funded by the AMA's Council on Pharmacy and Chemistry should convince the reader of the paucity of therapeutic studies dealing with clinical outcomes of treatment (as opposed to studies of the physiological or biochemical effects of new drugs.) For a discussion of attitudes toward therapeutic research among the medical elite, see James B. Herrick, "The Relation of the Association to Therapeutics," Transactions of the Association of American Physicians 38 (1923), 1-5. Herrick's presidential address was an attempt to prove the medical research community innocent of any charges that they had neglected therapeutics.



differences was lacking. When reputable investigators employed different criteria to select, treat and measure the results of treatment, it was difficult to compare and judge the outcomes of conflicting studies. In the case of chronic diseases like syphilis and tuberculosis, even larger institutions found it difficult to amass sufficient numbers of patients to form adequate judgments about their therapeutic practices.<sup>8</sup>

To reformers interested in improving the standards of therapeutic research, cooperative studies employing the talents of several institutions seemed a unique device for overcoming the limited vision and opportunities of individual investigators. Earlier generations of scientists held no particular brief for the ideological virtues of cooperation. Rather, cooperative studies seemed to offer an organizational solution to intellectual problems. Their special appeal was in promising to combine several investigative virtues in one: gathering large numbers of patients, to offset the effects of spontaneous recoveries; bringing the combined judgments of experienced investigators to bear on a problem, to offset the effects of individual bias; and, so far as was possible, specifying in advance the means and techniques for selecting patients, delivering treatment and evaluating results.

Despite their perceived scientific advantages, cooperative studies proved especially difficult to carry out. Researchers wishing to engage in cooperative therapeutic experiments not only had to develop

---

<sup>8</sup> For a partial characterization of the problem in the realm of syphilis treatments, see Harry M. Marks, "Notes From The Underground. The Social Organization of Medical Research," in Diana Long and Russell Maulitz, eds. Grand Rounds. One Hundred Years of Internal Medicine (Philadelphia: University of Pennsylvania Press, forthcoming).

the necessary financial resources to support their work, but they had to adopt social norms and organizational controls which would ensure that a plan of study, once agreed on, would be carried out according to agreement.

### Money

In financial terms, therapeutic evaluation in this period resided in a no man's land between laboratory research and public health projects aimed at demonstrating the benefits of community programs in the detection or prevention of disease. The interest of foundations was either in subsidizing new advances in medical knowledge or in convincing communities to devote their energies and resources to proven interventions. Efforts to provide more rigorous evaluation of existing or novel treatments fell somewhere between the two.<sup>9</sup>

Drug companies were a less ample, less respectable source of funds. Industry's reputation for exploiting clinical research in subsequent promotions made them suspect to many physicians engaged in therapeutic research. The desire of many specialists to avoid even the appearance of partiality made them reluctant to seek drug company

---

<sup>9</sup> For examples of each approach, see Robert E. Kohler, "A Policy for the Advancement of Science: The Rockefeller Foundation, 1924-1929" Minerva 16 (Winter, 1978), 480-515 and Peter Buck, "Why Not the Best? Some Reasons and Examples From Child Health and Rural Hospitals," Journal of Social History 18 (1985), 413-429. By the late 1920's, the laboratory side of therapeutic research had begun to cultivate a variety of funding sources, largely commercial in origin. See Swann, The Emergence of Cooperative Research (n. 1) for an analysis of university-corporate ties.

support.<sup>10</sup> And the repeated efforts to establish research consortia financed by a group of firms were largely unsuccessful.<sup>11</sup> The absence of stable sources of funding meant that cooperative investigations proceeded in a climate of uncertainty about their scope and future, if at all.<sup>12</sup> Yet in the final analysis, lack of funding was not the principal barrier to cooperative research.

#### The Division of Labor

Cooperative studies were meant to transcend the limited resources of individual institutions. But cooperative research proved to demand more, not less, of individual researchers: more time, more energy and more willingness to come to agreement about fundamental issues in the

<sup>10</sup> On the general animus towards corporate support of medical research see Richard Harrison Shryock, American Medical Research. Past and Present (New York: Commonwealth Fund, 1947), 141-146, and for the difficulties encountered by one study in accepting such aid, see Marks, "Notes From the Underground" (n. 8). For an analysis of changing attitudes towards accepting drug firm support, see Swann, The Emergence of Cooperative Research. Despite Swann's emphasis on the normalization of relations between universities and the corporate sector, suspicion of research sponsored by drug firms continued at least until after World War II.

<sup>11</sup> For a detailed analysis of one such venture, see John Parascandola, "Charles Holmes Herty and the Effort to Establish an Institute for Drug Research in Post World War I America," in John Parascandola and James C. Whorton, eds. Chemistry and Modern Society (Washington: American Chemical Society, 1983), 94-96. Corporate efforts to establish a consortium for drug evaluation on the model of the British Medical Council were not successful until after World War II; industry interest the British example is documented in the letter and accompanying documents from George W. Merck to Thomas Parran, October 18, 1938. F 19. Thomas Parran, Jr. papers, University of Pittsburgh. [Hereafter cited as "Parran papers."]

<sup>12</sup> For a detailed account of the resulting intellectual and organizational problems which ensued for one study, see Marks, "Notes From the Underground" (n. 8).

biology and management of human disease. The work demanded by cooperative studies overloaded clinicians whose efforts were already divided among laboratory and clinical research, patient care and teaching. The intellectual demands of cooperative research were even greater: rather than resolving differences of opinion about appropriate treatment and classification of patients, early efforts at cooperative studies were undermined by these disagreements. The case of the Cooperative Clinical Group study of syphilis treatments illustrates the problems.

A pioneering effort at cooperative therapeutic investigation, the Cooperative Clinical Group study began in 1928 when six of the most prestigious syphilis clinics in the country agreed to participate in an evaluation of current and future syphilis treatments.<sup>13</sup> Assisted by the Public Health Service, the enterprise promised participants little money but "perhaps a little glory and a chance to do something wonderful for syphilis research in this country."<sup>14</sup> If the experts could agree

---

<sup>13</sup> The participants in the study were John H. Stokes, Professor of Dermatology and Syphilology at the University of Pennsylvania; Udo Wile, Professor of Dermatology and Syphilology at the University of Michigan and Stokes' former mentor; Joseph Earle Moore, Professor at Johns Hopkins who, like Stokes was at work on what were to be the first textbooks summarizing the "modern" treatment of syphilis; Paul O'Leary, who had taken over from Stokes as head of the Dermatology Section at the Mayo Clinic in 1924; and Harold Cole, a respected dermatologist at Western Reserve, and close associate of the therapeutic reformer, Torald Sollman. Thomas Parran, Jr., Commissioner of Public Health for New York State and soon to be Surgeon General, served as the public spokesman for the study and the unofficial arbiter among the feuding clinicians.

On the Cooperative Clinical Group, see Harry Dowling, "The Emergence of the Cooperative Trial," Transactions and Studies of the College of Physicians of Philadelphia 4th Series 43 (July, 1975), 20-29 and Marks, "Notes From the Underground" (n. 8).

<sup>14</sup> John H. Stokes to Paul A. O'Leary. March 26, 1928. F 235. Parran papers.

on uniform standards for selecting, classifying, treating and evaluating patients, they might provide a reliable guide among the welter of opinions which bedeviled practicing physicians in choosing among syphilis treatments.<sup>15</sup>

From the beginning, the Cooperative Group was plagued with the problem they had set out to address: the lack of uniformity among physicians in approaches to treating syphilis. "Astonishing variations between the course pursued by different patients in the same clinic as well as between the practice of individual clinics frequently disclosed themselves...." <sup>16</sup> The analysis of treatment outcomes proved no less problematic than the classification of regimens. Irregularities of patient attendance and the vagaries of clinic record keeping compounded the problem of intentional variations in therapeutic practice. Analyzing the results of treatments on patients whose medical history was unknown and whose medical future was frequently lost from view proved difficult. <sup>17</sup>

Were the difficulties of classifying patients by stage of disease, regimen and treatment outcome merely technical, they might have been readily resolved. But differences of opinion about how to classify

<sup>15</sup> Committee on Research in Syphilis. Scientific Committee. Draft Agenda for Session on Clinical Problems. April 30, 1928. F 220. Parran papers.

<sup>16</sup> John H. Stokes, Harold N. Cole, Joseph Earle Moore, Paul A. O'Leary, Udo J. Wile, Taliaferro Clark, Thomas Parran, Jr. and Lida J. Usilton, "Cooperative Clinical Studies in the Treatment of Syphilis. Early Syphilis," Venereal Disease Information 13 (June, 1932), 207-209 [quote at 208].

<sup>17</sup> For a more detailed discussion, see Marks, "Notes from the Underground," (n. 8).

patients, treatments and outcomes were at the heart of existing therapeutic controversies.<sup>18</sup> Each time a question seemed to be settled, it arose again.<sup>19</sup> As chiefs of prestigious clinics, the members of the Cooperative Group were good at giving orders, and as former interns they were good at taking orders, but neither experience equipped them to share authority. They excelled at originating novel ideas and at criticizing other people's work but not at jointly resolving differences of opinion.

The organization of the work hampered efforts to arrive at a permanent consensus. Committed to a program of rigorous therapeutic investigations, the senior clinicians involved in the study nonetheless found it difficult to devote the necessary time and attention to the cooperative project. The task of classifying patients and outcomes accordingly fell to the statistical clerks of the Public Health Service assisting the study.<sup>20</sup> Questions which arose as a result of ambiguities in the data were resolved by consultation with one or at most, two, of the principals.<sup>21</sup> The problem was organizational, the consequences

<sup>18</sup> John H. Stokes to Thomas Parran, Jr. August 16, 1929. F 236. Parran papers.

<sup>19</sup> Lida J. Usilton, Memorandum for Doctor Joseph Earle Moore, January 15, 1935. F 223. Parran papers.

<sup>20</sup> See Marks, "Notes from the Underground," (n. 8).

<sup>21</sup> It was only when draft papers were circulated that the opportunity arose for others in the group to examine at length the improvised revisions in methods or definitions; accommodating their criticisms meant not merely redrafting the paper, but redoing the analysis. For an explicit discussion of the problem, see Joseph Earle Moore to John H. Stokes, April 1, 1932 (F. 224) and to Thomas Parran, Jr. April 2, 1932; John H. Stokes to Joseph Earle Moore, April 4, 1932 (F 228). See also John H. Stokes to Thomas Parran, Jr. May 23, 1930 (F 230). Experience was no teacher in this regard: the same

intellectual: absent the active participation of experienced clinicians, the resulting analyses were unreliable. The delegation of the abstracting to statisticians "...introduces into the whole study a problematic inaccuracy which ...may invalidate the whole material." <sup>22</sup>

In organizing their work, the Cooperative Group lacked examples to follow, and they lacked what development economists are fond of calling an "infrastructure": a network of resources, personnel and opportunities on which to draw, and the ability to use them productively. The problems to be resolved ranged from issues of intellectual credit—who were the authors in a cooperative study—to more mundane questions of financial accounting—could the funds intended to pay for a nurse in year one be used to buy an automobile for a clinic in year two? Most serious were those circumstances—intellectual and material—directly preventing the principal investigators from accomplishing their scientific aims. The ideological barriers which kept them from leaving non-medical personnel in charge of the data analysis were every bit as real as the material lack of time available for the principals to complete the work on their own. In tandem, they made the already difficult intellectual questions about measuring the outcomes of treatment or the severity of disease practically impossible to solve.

---

problem recurred in developing the Group's later papers on cardiovascular syphilis. Harold N. Cole to John McMullen [Assistant Surgeon General], April 20, 1935. (F 223). All references to Parran papers.

<sup>22</sup> Joseph Earle Moore to John H. Stokes, April 6, 1932. RG 90. PHS. VD Division 1918-1936. Committee on Research in Syphilis. Box 326. NA. See also Joseph Earle Moore to John H. Stokes, July 25, 1934. F 222. Parran papers.

Whatever the study's analytic defects, over the years the results of the Cooperative Clinical Group provided a benchmark for other studies to compare themselves to and emulate.<sup>23</sup> For those outside the specialized field of syphilis, it was not the accomplishments of the Cooperative Group which mattered but their aspirations. Few pre-war projects were as ambitious as the Cooperative Clinical Group and none have left a comparable record of their intellectual and organizational travails.<sup>24</sup> But as subsequent studies were to show, even when opportunity and money were more readily available, effective cooperation remained a rare commodity.

---

<sup>23</sup> Louis Chargin, William Leifer and Theodore Rosenthal, "Marpharsen in the Treatment of Early Syphilis. Comparison of Results in One Hundred and Eighty-Eight Cases With Those of the Cooperative Clinical Group," Archives of Dermatology and Syphilology 40 (August, 1939), 208-217. Clinicians who relied on the Cooperative Group as an authoritative source were sometimes "surprised," on closer examination, "to see what a relatively small number of cases the [critical conclusions about] neoarsphenamine are based on." [Cornell] Conferences on Therapy, "Evaluation of Drugs Used in the Treatment of Syphilis," JAMA 112 (June 10, 1939), 2417.

<sup>24</sup> In the 1940's, the U.S. Public Health Service sponsored a multi-clinic evaluation of the technique of intensive intravenous arsenotherapy reported on by Chargin, et. al in note 81, supra. Lida Usilton, staff statistician for the Cooperative Clinical Group, continued her work for the larger Cooperative project. See Cooperating Clinics of New York and Midwestern Groups, "Massive Arsenotherapy for Syphilis. United States Public Health Service," JAMA 126 (October 28, 1944), 554-558. I have been unable to locate the administrative records for this study. The Cooperative Clinical Group itself continued their activities through the late 1930's: there is nothing in the records of its later studies to suggest that they met with substantially greater success in accomplishing their aims. Indeed, with the principals less involved, and the labor delegated to more and more to junior physicians, they even lost some of the credibility which specialists in the field had given their earlier reports. See R.A. Vondelehr to J.E. Moore, August 22, 1939 and Moore's reply, August 28. Record Group 90. U.S. Public Health Service. General Classified Records, 1936-1944. Box 53. National Archives [hereafter cited as NA].



## II: Science At War

During World War II, the cause of cooperative therapeutic investigation received a substantial boost. The war provided an unprecedented opportunity for the nation's scientific elite to direct the conduct and organization of scientific research. In the case of medicine, responsibility for allocating research funds and problems was inherited by the small group of medical scientists represented on the committees of the National Research Council's Division of Medical Sciences. Officially, medical research policy was formulated by the Committee on Medical Research (CMR) of the Office of Scientific Research and Development (OSRD). As a non-government body, the NRC could not legally allocate government funds. But at its first meeting, CMR decided to rely heavily on the NRC's existing committees: Lewis Weed, Chairman of the NRC's Division of Medical Sciences, was appointed Vice-Chairman of the CMR, and the head of each of the NRC's major medical subject committees was appointed a CMR consultant. The overlap in membership between NRC committees and CMR eroded much of the distinction between the two organizations. 25

The NRC's committees had never suffered from a lack of eminent and capable scientists; having the resources to carry out their plans was

---

25 A.N. Richards, "Foreword," in E.C. Andrus, et. al. Advances in Military Medicine (Boston: Little, Brown and Company, 1948), I, xliii-xliv; Irwin G. Stewart, Organizing Scientific Research for War. The Administrative History of OSRD (Boston: Little Brown and Company, 1948), 45-46, 98-101. This did not mean that significant differences of opinion concerning strategy as well as tactics between the two organizations emerged on occasion, but a good deal of the day-to-day decision making was channeled through the NRC committees and sub-committees with rudimentary oversight by OSRD/CMR.

another matter.<sup>26</sup> The decision to place the management of medical research funding and policy in NRC's hands offered these scientists the opportunity to put pre-war ideals of clinical research into practice. For the first time in their history, medicine's intellectual elite had the opportunity not only to set an example for the rest of the profession but to actually direct the conduct of therapeutic research on a national scale. In each area of medicine, NRC's committees selected the topics to be studied and the best means of attack. Investigators willing to research the problems thereby designated were provided with ample funds.<sup>27</sup>

Prewar convictions regarding cooperative studies dovetailed neatly with military exigencies: putting a group of specialists to work on well-defined problems was deemed both efficient and scientific. In instances where several therapeutic options needed to be evaluated, centrally planned cooperative studies were deemed ideal. Such studies, by virtue of their ability to accumulate large numbers of patients treated according to a common regimen, were thought to yield the most reliable answers in the shortest time. But while wartime circumstances favored the organization and management of cooperative studies, they

---

<sup>26</sup> Stanhope Bayne-Jones, "A History of the National Research Council 1919-1933. VI. Division of Medical Sciences," Science 78 (July 14, 1933), 26-29.

<sup>27</sup> Between 1941 and 1947, OMR spent over \$25m in research funds; by contrast, the National Institutes of Health over this period spent only \$5.6m, \$4m of it in 1947 when OMR had virtually closed up shop. See Stephen Strickland, Politics, Science and Dread Disease. A Short History Of United States Medical Research Policy (Cambridge: Harvard University Press, 1972), 16; NIH Factbook (New York: Marquis Academic Publishing, 1972).

provided no guarantees that medical scientists would execute their well designed studies according to plan.

### A Scientific Ethic

..human experimentation is not only desirable but necessary in the study of many of the problems of war medicine which confront us.<sup>28</sup>

Among the many problems faced by CMR, finding improved ways to treat or ideally, to prevent, venereal disease ranked high on the list. During World War I, syphilis and gonorrhea in the military services caused the loss of nearly seven million days of active duty.<sup>29</sup> Existing methods of "prevention" were of limited effectiveness, and their success depended on the willingness of soldiers to come forth for "chemical prophylaxis" after exposure.<sup>30</sup> Newer methods, such as providing self-administered prophylactic doses of the newer sulfonamide drugs, carried less of a stigma, but the risks they posed to

---

<sup>28</sup> A.N. Richards to J.E. Moore, October 9, 1942. RG 227. OGRD/CMR. Box 36. F Human Experimentation. V.D. NA.

<sup>29</sup> Allan M. Brandt, No Magic Bullet. A Social History of Venereal Diseases in the United States Since 1880. (New York: Oxford University Press, 1985), 115.

<sup>30</sup> Several traditional means of prevention were available: the use of condoms during intercourse; so-called "station" prophylaxis, or administration of soap and/or calomel ointment on the genitals; and individual prophylaxis, which left it up to individuals to anoint themselves. During WWI, condoms and station prophylaxis had proven most effective, but the Army faced substantial moral and political opposition to a policy which was believed to encourage "vice". By WWII, the military found it possible to implement these measure, but their limited effectiveness made the army interested in screening better measures. See Brandt, No Magic Bullet (28), 110-115. Brandt provides an excellent discussion of the political and cultural problem of managing venereal disease in both World Wars.

military operations were unknown. Potential liabilities included adverse drug reactions; sensitization which would affect subsequent use in more serious medical circumstances; and in the case of gonorrhoea, the possibility of producing carriers who might transmit the disease to other soldiers.<sup>31</sup> The military accordingly focused on increasing the efficiency of traditional preventive measures and treatments. Meanwhile, with CMR's support, civilian researchers began the search for more effective interventions.

Of the two diseases, syphilis and gonorrhoea, the latter posed the greater scientific challenge. In the case of syphilis, the immediate problem was to adopt existing methods of treatment to military circumstances. If new remedies arose, viable animal models were available for initial assessments of their benefits and risks. Notwithstanding some difficulties in extrapolating these results to humans, systematic exploratory research could proceed. No such animal models existed for gonorrhoea: developing a vehicle for experimental infections was CMR's first order of business.<sup>32</sup> Until such methods were developed,

---

<sup>31</sup> J.E. Moore to A.N. Richards, February 1, 1943. NAS-NRC Central File. Division of Medical Sciences: Committee on Medicine. Subcommittee on Venereal Diseases: 1943. National Academy of Science [hereafter NAS].

<sup>32</sup> The contrast between the feasibility of using animal models for syphilis, and the absence of such models for gonorrhoea comes out clearly in the deliberations of the NRC Subcommittee on Venereal Diseases. Notwithstanding a history of unsuccessful efforts to produce gonococcal infections in the genital mucosa of experimental animals, NRC recommended that CMR award funds to Dr. Justina Hill of Johns Hopkins Hospital to work on the problem, on the grounds that no such efforts had been tried in recent years. Conference on Chemical Prophylaxis of Venereal Diseases. 23 March 1942. NRC Program Files: DIV NRC: Medical Sciences: CMR, Subcommittee on Venereal Diseases: Conferences: 1941--1943. NAS. On the use of animal models for studying syphilis, see Alan M. Chesney, Immunity in Syphilis (Baltimore: Williams and Wilkins Company, 1927).

however, human studies provided the only means of treatment evaluation.

If studies in humans were necessary, treating deliberately infected volunteers represented the ideal approach from a scientific point of view: the timing and degree of the infections could be controlled and patients carefully monitored while under treatment. Army medical officials were understandably reluctant to undertake such a study.<sup>33</sup> The alternative was to study prophylaxis in "naturally occurring populations": individuals who had already been exposed to gonorrhea. Members of the NRC's committee on venereal disease, long familiar with the difficulties of interpreting such uncontrolled studies, were unwilling to rely on them.<sup>34</sup>

By November, 1942, efforts to induce gonococcal infections in animals were only slightly more advanced than they had been in February, when the NRC's deliberations on the problem began.<sup>35</sup> The

<sup>33</sup> The Army had both scientific and political objections to the use of soldiers as volunteers in such a study, expressing strong doubts that the necessary isolation could be obtained. For a summary of Army objections, see Lewis H. Weed to Ross G. Harrison, March 2, 1943. RG 227. OSRD/CMR. F Human Experimentation. Box 36. NA.

<sup>34</sup> Conference on Chemical Prophylaxis of Venereal Diseases. 23 March 1942. NRC Program Files: DIV NRC: Medical Sciences: CMR, Subcommittee on Venereal Diseases: Conferences: 1941-1943. NAS. The experiences of the committee's chairman, Joseph Earle Moore, in the Cooperative Clinical Study no doubt were a factor in this attitude.

<sup>35</sup> Working under OSRD contracts, Dr. Justina Hill had been unsuccessful in inducing gonococcal infections in the mucous membranes of mice, but had "for the first time" produced gonococcal septicoemia by testicular injections of cultures in mice. NRC, Subcommittee on Venereal Disease, Minutes of a Conference on Chemical Prophylaxis. November 18, 1942. NRC Program Files: DIV NRC: Medical Sciences: Committee on Medicine, Subcommittee on Venereal Diseases: Conferences 1941-1943. NAS. Hill's work continued to make progress but a viable animal model for routine testing continued to be beyond reach. See J.E. Moore's assessment in early February, 1943: J.E. Moore to A.N. Richards, February 1, 1943. NAS-NRC Central File: DIV NRC: Medical Sciences, Committee on Medicine, Subcommittee on Venereal Diseases: 1943. NAS.

value of Sulfathiazole, the newest sulfonamide, in preventing gonorrhea remained relatively untested. The Army meanwhile continued to rely principally on older, less promising unctons.<sup>36</sup> At a conference held to acquaint various OSRD researchers with the progress of work on chemical prophylaxis, two research physicians from New York, Charles Carpenter and Alfred Cohn, proposed to evaluate sulfathiazole on experimentally infected prison volunteers.<sup>37</sup>

The arguments on behalf of a prison study were straightforward and, to the specialists on the NRC's subcommittee, persuasive. There was no question of the need to conduct a human experiment. The difficulties with animal experimentation and observational studies offered ample scientific justification for proceeding with such a study:

---

<sup>36</sup> In 1940, the Army had no standard kits in use against gonorrhea; calomel ointment, used against syphilis, had no specific effect on gonorrhea. A newly developed ointment containing silver picrate was adopted for universal use in the Army in July, 1942. Within weeks of its distribution it became apparent that the local irritation caused by the treatment made it unuseable on any such basis. Sulfonamide ointment was not officially adopted by the Army until March, 1944. Oral sulfathiazole prophylaxis, while widely used in various branches, was opposed by many military physicians: its general use was not approved until the Summer of 1943. Thomas H. Sternberg, Ernest B. Howard, Leonard A. Dewey and Paul Padget, "Venereal Diseases," in United States Army, Medical Department, Preventive Medicine in World War II. Vol. 5. Communicable Diseases Transmitted Through Contact or By Unknown Means (Washington: Office of the Surgeon General, 1960), 198-204. Meanwhile, doubts remained about the precise value of oral prophylaxis. See the remarks of Percy Pelouze [U.S. Public Health Service], in Charles Carpenter, Report of the Conference with the Prison Parole Commission ... Atlanta, Georgia, November 16, 1942. NRC Program Files: DIV NRC: Medical Sciences: CMR, Subcommittee on Venereal Diseases: Correspondence. NAS.

<sup>37</sup> NRC, Subcommittee on Venereal Disease, Minutes of a Conference on Chemical Prophylaxis. November 18, 1942. NRC Program Files: DIV NRC: Medical Sciences: Committee on Medicine, Subcommittee on Venereal Diseases: Conferences 1941-1943. NAS. Carpenter's proposal had been under discussion for over a month, at this point.

It is believed that military necessity and the impossibility of obtaining the desired results immediately or in the predictable future, in any other manner, justify the use of human volunteers for this purpose.<sup>38</sup>

The only remaining question, in the experts' view, was the selection of a suitable population. Any study which deliberately inflicted disease, they reasoned, must have the best chance possible of providing a valid answer. Only a well-controlled study could morally justify the undertaking.<sup>39</sup>

Ideally, the subjects chosen should be closely monitored for untoward responses and drug reactions. Most essential, the study population must be isolated from sexual activity for six months. Years of inconclusive research had convinced the NRC Subcommittee on Venereal Disease that only sexual isolation would guarantee the experiment's scientific success. Without such measures, subjects might naturally acquire infections, hopelessly contaminating any evaluation of treatment outcomes. Neither civilian nor military volunteers, they reasoned,

---

<sup>38</sup> J.E. Moore to A.N. Richards, February 1, 1943. NAS-NRC Central File: DIV NRC: Medical Sciences, Committee on Medicine, Subcommittee on Venereal Diseases: 1943. NAS.

<sup>39</sup> For details of the experimental design, see Proposed Plan of Procedure in the Study of Chemical Prophylaxis Among Human Volunteers Among Prison Inmates. December 19, 1942. RG 227. OSRD/CMR. Box 36. F Human Experimentation. NA. The study called for experimental and control groups with equal numbers of subjects a) with no history of gonorrhoea, b) with one to two previous "attacks", and c) three or more previous infections. The consent form attached to the protocol describing the risks of the study bears favorable comparison with contemporary examples of the genre.

would subject themselves to the required degree of control.<sup>40</sup> After considering several other alternatives, Joseph Earle Moore, chairman of the Venereal Disease Subcommittee, recommended that OSRD/CMR endorse "present" and "possibly future" proposals to study "chemical and chemotherapeutic prophylaxis of gonorrhoea" in a controlled study of prison volunteers.<sup>41</sup>

Customarily, funding recommendations originating with NRC committees were readily acted on by OSRD/CMR, after a brief technical review within NRC. In this instance, however, OSRD officials wanted a more comprehensive assessment, one which took into account the proposal's political ramifications as well its scientific merit.<sup>42</sup> This review

<sup>40</sup> Other populations considered, but rejected included 1) inmates of "institutions for the feeble-minded or insane" who were believed to be unable to offer their consent; 2) military prisoners and 3) conscientious objectors, neither of whom were expected to be fully cooperative. J.E. Moore to A.N. Richards, February 1, 1943. NAS-NRC Central File: DIV NRC: Medical Sciences, Committee on Medicine, Subcommittee on Venereal Diseases: 1943. NAS.

<sup>41</sup> J.E. Moore to A.N. Richards, February 1, 1943. NAS-NRC Central File: DIV NRC: Medical Sciences, Committee on Medicine, Subcommittee on Venereal Diseases: 1943. NAS. The arguments in Moore's letter reflect several months of deliberation with NRC and CMR/OSRD officials about the merits of the proposed study, but although this document responds to certain concerns of Moore's superiors, the logic of the argument is reflected in earlier, less formal communications.

<sup>42</sup> Joseph Earle Moore, whose committee originated the proposal initiated the discussion within OSRD in October, 1942, by consulting A.N. Richards, head of CMR, about their attitudes toward human experimentation. By December, 1942, Moore's proposal had been approved by the NRC's Committee on Medicine, and senior CMR and NRC officials were already involved in discussing the study. Vannevar Bush, head of OSRD, appears to have been officially notified by Richards on January 20; Bush's formal request to Frank Jewett, President of the National Academy of Sciences, originated on February 18, 1943, but Jewett had already known of the problem for at least a week at that point. See Frank Jewett and Ross Harrison to Vannevar Bush, March 5, 1943 and the chronology presented to Bush in mid-January by A.N. Richards. RG 227. OSRD/CMR. Box 36. F Human Experimentation. NA; Vannevar Bush to Frank Jewett,



soon reached the highest levels of NRC and its parent organization, the National Academy of Sciences (NAS). Senior officials in all three organizations recognized the need for the study to receive a careful and thorough peer review. It must be shown to be scientifically necessary for either NRC, NAS or OSRD to consider sponsoring it.<sup>43</sup> But the proposal's political implications were beyond the narrow technical competence of the medical specialists:

In an extreme case their scientific opinion might be the most authoritative in the world and yet their opinion on a matter of public policy have no more value than that of any similar group of intelligent laymen.<sup>44</sup>

Whether or not a human experiment was necessary was a "scientific" question; whether or not prisoners should be used in such a study was a matter of "public policy". To Frank Jewett, President of the National Academy of Sciences, Moore's proposal was "certainly loaded with potential dynamite for those sponsoring it." If objections to the

February 18, 1943 in NAS-NRC Central File: DIV NRC: Medical Sciences: Committee on Medicine, Subcommittee on Venereal Diseases: 1943. NAS. On Jewett's involvement, see Ross G. Harrison to Frank Jewett, February 9, 1943 and Jewett's reply, February 16, 1943. NAS-NRC Central File: DIV NRC: Medical Sciences: Committee on Medicine, Subcommittee on Venereal Diseases: 1943. NAS. The plurality of organizations and the complexity of personal networks among the players make any timetable highly provisional. The dispersal of relevant documents among a half-dozen organizational and personal archives makes it even less likely that the above chronology is more than approximate.

<sup>43</sup> On OSRD's views, see A.N. Richards [CMR] to J.E. Moore, January 29, 1943. RG 227. OSRD. CMR. Box 53. NA; on the views of senior NRC and NAS officials, see Frank Jewett to Ross Harrison, February 16, 1943, [NAS-NRC Central File: DIV NRC: Medical Sciences: Committee on Medicine, Subcommittee on Venereal Diseases: 1943. NAS] and the discussion below.

<sup>44</sup> Frank Jewett [President, National Academy of Sciences] to Ross Harrison [Chairman, NRC], February 23, 1943 [hereafter cited as letter 2]. NAS-NRC Central File: DIV NRC: Medical Sciences: Committee on Medicine, Subcommittee on Venereal Diseases: 1943. NAS.

study should subsequently be raised, it would be difficult to convince anyone that the prisoners had freely volunteered. Before risking the reputation of the Academy and the NRC by endorsing such a project, Moore's committee must demonstrate more conclusively that the alternatives were unworkable.<sup>45</sup> After all,

If the military with some millions of men at their disposal are not prepared to handle the matter on a truly voluntary basis, it would seem to me to raise a very considerable doubt as to the necessity of performing the experiment at all.<sup>46</sup>

Part of Jewett's concern was that of the scientific politician, eager to establish a procedural record behind which he and Vannevar Bush, head of OSRD, could hide in the event of a subsequent outcry.<sup>47</sup>

<sup>45</sup> Frank Jewett to Ross G. Harrison, February 16, 1943; February 23, 1943 [in response to Harrison's letter of February 20: hereafter cited as letter 1] and February 23, 1943 [letter 2]. NAS-NRC Central File: DIV NRC: Medical Sciences: Committee on Medicine, Subcommittee on Venereal Diseases: 1943. NAS. It is apparent that a large part of Jewett's concern rested in the fact that he and Harrison would be in the front lines of any subsequent inquiry into the decision, but also that Jewett, once he had reviewed the scientific rankings of the proposal, did not wish to see it sidetracked by taking it to the Academy's Council or Administrative Committee which would, in his opinion, be deadlocked on such an issue.

<sup>46</sup> Frank Jewett to Ross G. Harrison, February 16, 1943. NAS-NRC Central File: DIV NRC: Medical Sciences: Committee on Medicine, Subcommittee on Venereal Diseases: 1943. NAS.

<sup>47</sup> As Jewett put it to Ross Harrison, in his second letter of February 23, 1943, Bush, director of OSRD and Richards, Chairman of CMR, were "entitled to a record they can use without question in connection with their final decision." See also Jewett's draft of a letter for Ross Harrison to send to Lewis Weed, Chairman of the NRC's Division of Medical Sciences, and the accompanying cover letter to Harrison, February 25, 1943. NAS-NRC Central File: DIV NRC: Medical Sciences: Committee on Medicine, Subcommittee on Venereal Diseases: 1943. NAS. For Bush's concern with the public reaction, see A.N. Richards to J. Earle Moore, February 8, 1943. RG 227. OSRD. CMR. Box 36. F Human Experimentation. NA. This folder also contains the various endorsements of the study solicited from military and public health officials at the requests of A.N. Richards, Jewett and Bush.

But Jewett also required that Moore justify his scientific conclusion that a prison population offered the best means of conducting a valid and informative experiment. Moore's reasons for rejecting studies with military and civilian volunteers not only had to be documented, they had to be persuasive.<sup>48</sup> Once convinced of the medical experts' logic as well as their facts, Jewett and Ross Harrison were prepared to take personal responsibility for endorsing the study:

So far as the risk of adverse public reaction is concerned we realize that opinions differ widely and that the possibility unquestionably exists. It is our mature judgment that in view of the weight of scientific and medical advice and the prospective great and continuing advantage both to the military and civil populations, it is a warranted risk.<sup>49</sup>

Given the extensive political and moral deliberations preceding the study, its eventual outcome was anticlimactic. After several months of research on prisoners in the federal penitentiary at Terre Haute, John F. Mahoney reported that the procedure for inducing gonorrhoea in humans was too unreliable to enable meaningful tests of prophylactic

<sup>48</sup> Jewett's regard for the scientific merits and necessity for the study appears to have progressively increased: see Jewett to Harrison, February 16, 1943, February 23 [letter 2] and March 1, 1943. By mid-February, 1943, Ross Harrison was certainly "convinced" of the study's "scientific soundness" and "desireability". See Harrison to Vannevar Bush, February 19, 1943. His subsequent task seems to have been to resolve Jewett's continuing reservations. See Harrison to Jewett, February 20, 1943. NAS-NRC Central File: DIV NRC: Medical Sciences: Committee on Medicine, Subcommittee on Venereal Diseases: 1943. NAS.

<sup>49</sup> Ross G. Harrison and Frank Jewett to Vannevar Bush, March 5, 1943. NAS-NRC Central File: DIV NRC: Medical Sciences: Committee on Medicine, Subcommittee on Venereal Diseases: 1943. NAS. Neither Harrison nor Jewett was willing, however, to commit their respective organizations to endorsing the proposal, nor were they willing to "advise" on either the study's legal implications or the "probable attitudes of public officials whose sanctions must be obtained" to enable the experiment.

agents.<sup>50</sup> On Mahoney's recommendation, the study was abandoned but the ethical rationale for controlled experiments elaborated while the gonorrhea study was being discussed outlived its meager scientific yield. Research on humans worth doing was worth doing well. A less than adequate study from a methodological point of view was morally unacceptable.

### The War At Home: Penicillin

While plans for investigating gonorrhea treatments were getting underway, CMR officials were beginning a much larger program of clinical investigation, to evaluate the therapeutic potential of penicillin. As early as 1940, British and American researchers had begun to demonstrate penicillin's remarkable abilities in uncontrolled staphylococcal and streptococcal infections.<sup>51</sup> Under normal circum-

---

<sup>50</sup> After initial approval of the project was obtained, the location was shifted to the Federal Penitentiary at Terre Haute, Indiana and conducted under the supervision of the Public Health Service. Mahoney's recommendation to abandon the project was based on his understanding that the research be dropped if it proved "difficult or hazardous" to draw "sound conclusions" from the study. Mahoney's inability to inculcate gonorrhea consistently in the volunteers proved just such an obstacle. See Interim Report, February 9, 1944 and BiMonthly Progress Report No. 4, May 1, 1944. RG 227. OSRD/CMR. Project # M-3169; and John F. Mahoney to A.N. Richards, March 1944. RG 227. OSRD/CMR. Correspondence: "Mahoney". All NA.

<sup>51</sup> The best source on penicillin's history is Gladys L. Hobby's recently published Penicillin, Meeting the Challenge (New Haven: Yale University Press, 1985). A participant in the earliest United States studies of penicillin, Dr. Hobby has complemented her personal knowledge with extensive archival research. Although the present manuscript was substantially advanced by the time Dr. Hobby's book was published, parts of this account have benefited from additional information provided in her account.

According to Hobby, the first clinical studies in this country

stances, these initial reports would be followed by a series of pharmacological and therapeutic studies conducted at various research centers around the country. Results, possibly conflicting, about the effects of penicillin in treating various conditions would slowly accumulate, and over a period of years, authoritative medical opinion about the drug's optimal uses would be formed.

Wartime circumstances provided CMR with the unusual opportunity to determine, more systematically and efficiently, when and how penicillin was best used. Prior to the summer of 1943, supplies of penicillin were severely limited, so much so that it was distributed only to a handful of "experienced investigators" who agreed in return to work "under [CMR's] direction and supervision."<sup>52</sup> CMR's research program

were conducted in October, 1940 at Columbia University and reported at the Society for Clinical Investigation meetings the following May (69-77). The decision to launch a full scale development program was, however, initiated by Howard Florey's visit to the U.S. in the summer of 1940 (Hobby:80-81, 87-92). Although the official OSRD (CMR) program to develop penicillin did not begin until October, 1941, NRC and CMR officials were extensively involved, both officially and unofficially, in assisting the penicillin researchers prior to that date [Hobby: 87, 94; Chester S. Keefer, "Penicillin: A Wartime Achievement" in Andrus, Advances in Military Medicine (n. 25), II, 717-718].

<sup>52</sup> Keefer, "Penicillin: A Wartime Achievement" in Andrus, Advances in Military Medicine (n. 25), II, 719. Through April 1944, Keefer and CMR were in charge of allocating civilian supplies of penicillin, which essentially was distributed for the research program and on an emergency basis only. [U.S. Senate. Committee on Education and Labor, Wartime Health and Education. Part 7. Hearings on Medical Research, 2205-2207. 78th Congress. 2nd Session; J. Solon Mordell, Conference on Requirements and Distribution, December 3, 1943. RG 179. Box 75. Chemicals: Drugs: Penicillin.] As supplies increased, the War Production Board authorized broader civilian distribution, still in accordance with recommendations prepared by Keefer. [Penicillin Producers, Industry Advisory Committee, April 11, 1944. RG 179. Box 1691. 533.8105M; Fred J. Stock (WPB) to All Penicillin Producers and Distributors, April 27, 1944. RG 179. Box 75. Chemicals: Drugs; Penicillin, the Indications, Contradindications, Mode of Administration and Dosage for Penicillin. (War Production Board, May 1, 1944). RG 179. Box 379. 052.524C] Keefer continued as

focused first on the most serious infections, and those for which other drugs performed poorly.<sup>53</sup>

For the military, syphilis was as serious a medical problem as any, and existing therapies quite inadequate. The arsenical treatments favored by civilian experts took months to complete; the need for close medical supervision while these toxic drugs were being used meant that optimal therapy drained operating units and medical facilities alike. Initially, military research focused on accelerated arsenical treatments deemed too experimental for civilian use.<sup>54</sup> The research program to evaluate the use of penicillin in syphilis did not begin until the summer of 1943 when John F. Mahoney demonstrated that, contrary to

---

advisor to the War Production Board on policy for penicillin distribution at least through April 1945. [WPB. Report of the operations, From April 1944 to April 1945...Office of Civilian Penicillin Distribution. RG 179. Box 379. 052.528.] All references NA.

A handful of researchers had their own supplies of the drug prior to the CMR program, and several groups continued to divert official supplies to unauthorized research. See Hobby, Penicillin. Meeting the Challenge (n. 51), 69-77.

<sup>53</sup> Chester S. Keefer. Francis G. Blake, E.K. Marshall, Jr., J.S. Lockwood and W. Barry Wood, Jr., "Penicillin in the Treatment of Infections. A Report of 500 Cases," JAMA 122 (1943), 1217-1224. The initial clinical uses summarized by Hobby, 142-149, are in most instances, as was usually the case with new drugs, case by case reports. Hobby says relatively little about the differences in experimental design and purposes between civilian and military research programs with penicillin.

<sup>54</sup> Paul Padget, "Diagnosis and Treatment of the Venereal Diseases," in U.S. Army. Medical Service. Internal Medicine in World War II. Vol. II. Infectious Disease (Washington: Office of the Surgeon General, 1963), 419-423.

earlier reports, the drug had a pronounced spirocheticidal effect in experimental infections.<sup>55</sup>

Security restrictions on all penicillin research did not prevent news of Mahoney's findings from rapidly circulating in the community of venereal disease experts. Mahoney's report that the initial syphilitic lesions in four sailors had promptly disappeared upon treatment with penicillin, heightened military interest in the drug.<sup>56</sup> But penicillin would not have been the first anti-syphilitic drug which failed to realize its initial promise. Given the Army's pressing need for a rapid and convenient syphilis treatment, they began the use of penicillin in high dosages well in advance of any carefully controlled evaluations. Recognizing the need to examine the merits of penicillin in "the spirit of impartial enquiry" that military users of the drug could not afford,

---

<sup>55</sup> After demonstrating its effects in rabbit syphilis, Mahoney obtained authorization to try the drug in humans. [Chester Keefer], Memorandum on Use of Penicillin in Syphilis [late October, 1943]. NRC Program Files: DIV NRC: Medical Sciences: Committee on Medicine, Subcommittee on Venereal Diseases: Correspondence. NAS. Approval for Mahoney's human trial was granted over the objections of J.E. Moore, who thought that additional animal investigations would be more productive and reliable. J.E. Moore to E.C. Andrus [CMR], July 13, 1943. RG 227. OSRD/CMR. Box 67. F Penicillin VD. NA.

<sup>56</sup> Padget, "Diagnosis and Treatment of the Venereal Diseases," (n. 54), 419-423; William S. Middleton, "European Theater of Operations," in U.S. Army. Medical Service. Internal Medicine in World War II. Vol. I. Activities of Medical Consultants (Washington: Office of the Surgeon General, 1961), 291-299. Both accounts make it clear that extensive informal contacts between NRC committee members and military consultants served to keep the latter abreast of new developments; even when the NRC was not willing to officially endorse a program of venereal disease treatment, its informal advice appears to have encouraged innovation. By the time Mahoney's findings were reported at the annual meeting of the American Public Health Association in October, 1943, planning for the NRC study appears to have been underway.

the Army requested that the NRC organize a more systematic investigation of its potential in treating syphilis. <sup>57</sup>

By the Fall of 1943, increased penicillin production made it possible to begin planning a civilian investigation of syphilis treatment, under the direction of Joseph Earle Moore, chairman of the NRC's Subcommittee on Venereal Disease, and an alumnus of the Cooperative Clinical Group. <sup>58</sup> The NRC's studies had two aims: to determine the optimal ways of using penicillin to treat syphilis and to evaluate its efficacy under more carefully controlled circumstances. A cooperative study, operating according to a fixed plan, promised to meet both

<sup>57</sup> The initial uses of penicillin were on the customary basis: take a few cases in extremis and see what happens. Charles Fletcher, "First Clinical Use of Penicillin," British Medical Journal 289 (22-29 December 1944), 1721-1723. The need for "impartial inquiry" was noted in reference to the testing of penicillin for wound infections: U.S. Army Memo, [Report of Penicillin Conference at No. 48. General Hospital, August 24-25, 1943]. RG 227. OSRD/ CMR. Box 60. F Penicillin, Miscellaneous. NA. The military's evaluation of penicillin for syphilis is reported by Donald M. Pillsbury, "Penicillin Therapy of Early Syphilis in 14,000 Patients: Follow-Up Examination of 792 Patients Six or More Months After Treatment," American Journal of Syphilology and Dermatology 30 (March, 1946), 134-135. The two figures in Pillsbury's title speak volumes about the difficulty of conducting an informative investigation of syphilis treatment within the military. On the NRC's involvement, see N.R.C.-U.S. P.H.S. Meeting of Penicillin Investigators. 7 and 8 February 1946, 1-2; [Chester Keefer], Memorandum on Use of Penicillin in Syphilis [late October, 1943]. NRC Program Files: DIV NRC: Medical Sciences: Committee on Medicine, Subcommittee on Venereal Diseases: Correspondence. NAS.

<sup>58</sup> Pilot studies on humans were being explored as early as August but detailed planning of a cooperative investigation did not begin until October. [Chester Keefer], Memorandum on Use of Penicillin in Syphilis [late October, 1943]. NRC Program Files: DIV NRC: Medical Sciences: Committee on Medicine, Subcommittee on Venereal Diseases: Correspondence. NAS. Moore's committee was under the jurisdiction of the NRC Committee on Chemotherapeutic Agents, chaired by Chester S. Keefer. As Special Advisor and Consultant to CMR and, after 1944, their Chief Medical Administrative Officer, Keefer provided the necessary links to OSRD.



objectives: it would accumulate results more quickly, and more reliably, than a series of less focused individual inquiries.<sup>59</sup>

Participating researchers agreed in advance to "cooperate in a planned investigation, [with] each clinic utilizing a treatment scheme indicated to it by the steering panel." Standardized data collection and laboratory procedures were agreed to at the outset.<sup>60</sup> Final decisions about the conduct of the research were in the hands of CMR, whose virtual monopoly over the civilian distribution of penicillin greatly enhanced its authority.

The clinics involved represented a handful of elite investigators, selected either for their expertise in syphilis or in the study of anti-infectious agents.<sup>61</sup> Thanks to OSRD, they did not lack for funds or manpower in pursuing their researches. But these unusually favorable circumstances found researchers no less reluctant to surrender their intellectual autonomy, even in the pursuit of agreed-upon goals. The

<sup>59</sup> OSRD/CMR, Agenda for Penicillin Conference Nov 9, 1944. RG 227. OSRD/CMR. Box 67. NA.

<sup>60</sup> Minutes of A Conference on Penicillin in the Treatment of Syphilis in Human Beings. October 29, 1943. NRC Program Files: DIV NRC: Medical Sciences: Committee on Medicine. Subcommittee on Venereal Diseases: Minutes: 1940-1943. NAS.

<sup>61</sup> Ultimate decisions concerning the study, while strongly influenced by the specialists on the Subcommittee for Venereal Diseases, were controlled by Chester Keefer of the CMR; it was Keefer who added the clinics of Francis Blake and W. Barry Wood to the study. Twenty-First Meeting, Subcommittee on Venereal Diseases. 11 November 1943. NRC Program Files: DIV NRC: Medical Sciences: Committee on Medicine. Subcommittee on Venereal Diseases: Minutes: 1940-1943. NAS. Blake's and Wood's expertise, like that of Keefer's, was in the more general domain of chemotherapy and infectious disease. An undercurrent of tension between general infectious disease and disease specific specialists runs through the wartime and postwar discussions and deserves a more careful examination than I have been able to provide.

study began by examining the value of penicillin in a range of doses up to 1.2 million units. Participating clinicians, who wished to cure as many patients as possible, resented the protocol's requirement to employ the lower dosages.<sup>62</sup> Even after dosages below 1.2 million units were abandoned, investigators objected to "merely acting as technicians, each dealing with a small phase of a large experiment."<sup>63</sup>

Researchers' requests to use a portion of their penicillin allocations for autonomous investigations were repeatedly rejected by senior CMR officials.<sup>64</sup> Not surprisingly, some individuals followed promising leads anyway, with the pursuit of scientific curiosity resulting in the neglect of patient follow-up in the cooperative study. Yet in a disease like syphilis, only data on long term outcomes could address the question of cure.<sup>65</sup>

---

<sup>62</sup> The initial use of lower dosages was due both to the shortage of penicillin supplies at the outset, and to the theoretical importance of testing the drug's value in the lower ranges. Mahoney's work had already demonstrated that 1.2m units would work, but no one knew if lesser amounts would do as well. See NRC. CMR, Minutes on a Conference on Penicillin in the Treatment of Syphilis in Human Beings, 29 October 1943. NRC Program Files: DIV NRC: Medical Sciences: Committee on Medicine, Subcommittee on Venereal Disease: Minutes: 1940-1943. NAS; and OSRD/CMR, Agenda for Penicillin Conference November 9, 1944. RG 227. OSRD/CMR. Box 67. NA.

<sup>63</sup> J.E. Moore, Memorandum, May 9, 1945. RG 227. OSRD/CMR Correspondence. Box 67. F Penicillin VD. NA.

<sup>64</sup> The NRC Sub-committee on Venereal Diseases passed these requests on to the NRC Committee on Chemotherapeutics, whose chair, Chester S. Keefer, also served as Medical Administrative Officer for OSRD/CMR. Keefer's committee insisted that any proposed studies would have to come up through NRC and CMR for a full review. J.E. Moore to members, Penicillin Panel, Subcommittee on VD, January 24, 1945. RG 227. OSRD/CMR Correspondence. Box 56; J.E. Moore, Memorandum, May 9, 1945. RG 227. OSRD/CMR Correspondence. Box 67. F Penicillin VD [both NA].

<sup>65</sup> See the remarks of Joseph Earle Moore, U.S. P.H.S., F.D.A. and N.R.C., Penicillin Conference 26-27 March 1946 [mimeo text] (Washington: n.p., n.d.), 159.

Despite the fact that they employed a prospective rather than a retrospective design, the methodological difficulties of the penicillin studies evoke those of the prewar Cooperative Clinical Group study. Nearly half the cases accumulated during the war had to be discarded, because of incomplete information or failure to follow the protocol.<sup>66</sup> Loss of patients to follow-up hampered interpretation of the remaining data.<sup>67</sup> Although the study was intended to compare standardized treatments across clinics, individual clinics were rarely assigned more than two or three of the numerous regimens being tested. Variations in race, gender and stage of disease among the clinics further complicated efforts at interpretation, by confounding treatment with clinic

---

<sup>66</sup> As late as 1945, Moore reported that the initial 6,000-8,000 cases would need to be re-abstracted and re-analyzed. J.E. Moore to A.N. Richards, May 2, 1945. RG 227. OSRD/CMR. Box 67. F Penicillin VD. NA. Only 6,000 of 11,000 cases proved useful in the end. See Margaret Merrell in "General Discussion," NRC. U.S. P.H.S., Meeting of Penicillin Investigators. 7 and 8 February 1946, 147. This contrasts with Lowell Reed's estimate in February, 1945, that 25,000 cases would be needed to produce "statistically significant results". See report of J.E. Moore and Lowell Reed to the Committee on Chemotherapeutics. Minutes. 9 February 1945. NRC: Division on Medical Sciences. Committee on Chemotherapeutics and Other Agents. NAS.

<sup>67</sup> Overall loss to follow up averaged 9.5% but was as high as 43.5% in one clinic. See J.E. Moore, "Preliminary Statement," U.S. P.H.S. Conference of Investigators of Penicillin Therapy [February 7-8, 1946], 4.

effects.<sup>68</sup> Before NRC was ready to draw even tentative conclusions from the study, the war itself was drawing to a close.

Despite their problems, cooperative studies had an appeal which the outbreak of peace did little to reduce. If studies conducted according to a standard protocol could guide military policy, there was no reason why they could not also be used to shape the therapeutic practices of civilian physicians. As military pressures abated, the NRC's attention turned to evaluating treatment schedules and modes of administering penicillin which might prove useful in postwar civilian practice.

During the war, the Army's need had been for a rapid and effective means of treating syphilis. The NRC's investigations focused accordingly on treatment schedules which might deliver the greatest protection in the shortest amount of time. With an end to hostilities in sight, the specialists on the NRC committees began giving equal consideration to studies which would anticipate and guide the physician in civilian practice. Some of the questions posed continued to balance the aims of effectiveness and efficiency which had characterized the wartime

---

<sup>68</sup> Assignments were based on estimates of the numbers of patients available in individual clinics and, presumably, on the willingness of investigators to try specific regimens, although this is not explicitly stated in the records. Variability in follow-up rates only increased the "noise" in the data present from the initial experimental design. Other problems evocative of the Cooperating Clinic Study were the inclusion in the study of patients who had been treated prior to entry, and differences of opinion regarding the distinction between relapse and reinfection. See Paul D. Rosahn, "The Treatment of Early Syphilis With Penicillin Alone and Combined With Mapharsen and With Bismuth. Results of a Nation-Wide Study." in U.S. P.H.S. Conference of Investigators of Penicillin Therapy [February 7-8, 1946] and the comments of Joseph Earle Moore and Margaret Merrell in the discussion at this conference: N.R.C.-U.S. P.H.S. Meeting of Penicillin Investigators, 7 and 8 February 1946 [n.p], 146-147.

studies: were the various penicillin emulsions, more convenient to use, as effective as amorphous penicillin? Which dosages produced the best results with the least trouble? Other inquiries, such as the attempt to evaluate the use of bismuth in tandem with penicillin, were pursued in the belief that physicians would try such a combination, however irrational, once penicillin became readily available.<sup>69</sup> When the average practitioner began experimenting with this approach, the specialists' studies would provide them with information about its value. Knowledge, it was believed, translated to influence: better informed physicians might also practice more intelligently.

In January of 1946, OSRD (CMR) turned responsibilities for the penicillin study over to the Public Health Service (PHS). The transfer of authority did little to alter the membership of those directing the

---

<sup>69</sup> The idea of a bismuth-penicillin combination was based on an analogy with the traditional approach to syphilis treatment combining arsenicals and heavy metals. There was no particular theoretical or experimental reason to believe that the action of penicillin was sufficiently similar to that of arsenic to require an initial supplement, but the specialists went ahead anyway, against the advice of the Committee on Chemotherapeutics, in the belief that physicians would try the approach after the war and CMR should be prepared to advise them. J.E. Moore to Chester Keefer, November 21, 1944. RG 227. OSRD (CMR). Box 53. To obtain "proper clinical evaluations" of this and other questions, producers agreed to continue supplying penicillin only to investigators engaged in the authorized program of organized research. Penicillin Producers Advisory Committee, Meeting, April 20, 1945. RG 179. Box 1697. 533.8105M. War Production Board officials feared that if restrictions were lifted, they might soon be handling "letters from people all over the country requesting penicillin for clinical research." Rationing of penicillin supplies continued to be an on-again, off-again affair through the Winter of 1946. See Initial Report on Operations, July 1, 1945 to March 31, 1946. Drug Section, Chemicals Division, Civilian Production Administration. RG 179. Box 73 [all references NA].

study, or their confidence in cooperative investigations.<sup>70</sup> But with penicillin supplies increasing, getting participating researchers to stick to the study protocols became even more difficult, NRC exhortations notwithstanding:

Since money is provided for a particular purpose, it should so be employed. The individual clinics should resist pressure to abandon a given treatment method 'because we have already so treated 100 patients with it,' and to adopt a new one, perhaps because of the latest publication of a new penicillin fraction, of methods of administration or of absorption delaying. The Advisory Committee ...is as anxious as individual investigators to adopt new methods but would prefer, now that military pressure has relaxed, to await definite information from special experimental centers before authorizing general application.<sup>71</sup>

In 1947, projected cutbacks in N.I.H.'s funding for venereal disease research led to a phasing out of the cooperative study in favor of individual investigations, both laboratory and clinical. Those managing the study responded by eliminating clinics with poor track records in following the protocol and finding patients, while maintaining support for the Central Statistical Unit.<sup>72</sup> Well after the war, the

<sup>70</sup> The Public Health Service simply added two or three members to the existing NRC subcommittee overseeing the study, christened it an advisory study section to the National Institutes of Health, and continued operating as before. For a list of committee members, see J.E. Moore, "Preliminary Statement," U.S. P.H.S., Conference of Investigators of Penicillin Therapy [February 7-8, 1946], 1.

<sup>71</sup> J.E. Moore, "Preliminary Statement," U.S. P.H.S. Conference of Investigators of Penicillin Therapy [February 7-8, 1946], 4.

<sup>72</sup> "Financial Support for Medical Research in the Venereal Diseases," American Journal of Syphilis, Gonorrhea and Venereal Diseases 31 (December, 1947), 664-668. There is some evidence that Surgeon General Parran was more concerned with protecting funding for the federal program of venereal disease case finding than with the research program. U.S. Congress. House of Representatives. Committee on Appropriations. Hearings. Department of Labor-Federal Security Appropriation Bill for 1948. February-March 1948. Part 2, 282-285. 80th Congress. 1st Session.

Central Statistical Unit continued to churn out publications on the study's behalf.<sup>73</sup> But despite the impressive numbers of patients they enrolled, the cooperative investigators were forced to draw heavily on speculation and ad hoc interpretations of the data when defending specific findings against the conclusions of other researchers.<sup>74</sup> For the study's statisticians, the most important result had been learned much earlier:

It is less important to get very large numbers of patients on a particular schedule and then not pay much attention to following them, than it is to get a smaller number who are followed through. There is a balance between the two problems of getting large numbers and devoting enough energy to following them up, so that conclusions are not based primarily on pure assumptions.<sup>75</sup>

---

<sup>73</sup> Margaret Merrell, "Estimates of Relapse and Reinfection Rates in Early Syphilis Treated With Penicillin," American Journal of Syphilis, Gonorrhea and Venereal Disease 35 (November, 1951), 532-543.

<sup>74</sup> Frank W. Reynolds, "Penicillin in Early Syphilis: An Analysis of the Discrepancies Between the Results of Arnold et. al. and Those of the Central Statistical Unit," in U.S. Public Health Service, Recent Advances in the Study of Venereal Diseases. A Symposium, April 8-9 1948. (Washington: Venereal Disease Education Institute, 1948), 113-121. In comparing the Cooperative Study with the superior results obtained by John Mahoney and his colleagues, Reynolds notes that the superior "cure" rates obtained by the latter group may have been due to the Cooperative Study's greater diligence in following patients, which lowered their overall success rate.

<sup>75</sup> Margaret Merrell in "General Discussion," N.R.C.-U.S. P.H.S. Meeting of Penicillin Investigators. 7 and 8 February 1946 [n.p], 147-148. Interestingly, the postwar cutbacks in funding for the study apparently concentrated the mind: "The fact that only two issues instead of a battery were under investigation meant that the material was not fractionated so many ways. There were about 3,000 cases, of which over two thirds were concentrated at three large clinics which distributed their cases almost equally over the 4 schedules. Thus the cases under these 4 treatments were almost identical in their distribution as to color, sex and admission diagnosis." Margaret Merrell, "Report From Central Statistical Unit On Comparative Failure Rates for Early Syphilis Treated with Penicillin," A Symposium on Current Progress in the Study of Venereal Diseases. Held Under the Auspices of the Syphilis Study Section, Division of Research Grants and Fellowships, N.I.H. April 7-8 1949. [Washington: Public Health Service, 1949], 40.

The need for more planning and better follow through was an experiential lesson, not easily taught in the textbooks. Many more researchers would have to share the frustrations of the penicillin investigators before the message took hold. Meanwhile, despite its limitations in practice, the systematic approach pioneered in the NRC's studies of penicillin served as a testimony to the virtues of cooperative research:

The first step in the evaluation of a chemotherapeutic agent is the discovery that X drug is "good" in the treatment of Y disease. In the past, once that step has been made, there has been a great tendency for the responsible leaders of the medical profession to lose interest in the subsequent all important but infinitely less dramatic subsequent steps. These include attempts to decide: how "good" is X drug? in what forms is it of little value? does one administer it by the pound or by the ton? daily, weekly or for 18 month periods?; under what circumstances is the treatment definitely worse than the disease from the standpoint of toxicity? from the standpoint of naturally acquired resistance? or from the standpoint of the health of the general public? <sup>76</sup>

The cooperative studies of penicillin demonstrated that it was possible to radically collapse the interval between introducing a new drug and obtaining the knowledge which would enable the practitioner to use it intelligently:

Erhlich made the X-drug-Y-disease step almost forty years ago and yet there is pathetically little information available today on the proper use of the organic arsenicals in the treatment of syphilis.

Mahoney made the X-Y step for penicillin in syphilis less than four years ago and by means of the cooperative approach a vast amount of information on the subsequent steps has already been accumulated. <sup>77</sup>

It comes as no surprise that the specialists involved in the NRC's programs sought to maintain their influence on the conduct of medical

<sup>76</sup> [Walsh McDermott] to Doctors Palmer, Bogen, Barrwell, Hinshaw, Willis and Long [March 27, 1947], Suggestion for the Report From the Panel on Dose Regimens to the Tuberculosis Study Section. Esmond Long Papers, National Library of Medicine, Box 15. [Hereafter cited as Long papers.]

<sup>77</sup> Ibid.



research after the war. During World War II specialists had consolidated their hold on all areas of medical research. But that the tradition of cooperative therapeutic studies should survive with them was by no means a foregone conclusion. Even under the most favorable circumstances—and for getting scientists to cooperate, national emergencies were the most favorable circumstances—obtaining the sustained cooperation of clinical investigators had been difficult to engineer. At the end of the war, medical researchers faced the question of whether to continue working under a common yolk or be free to pursue their intellectual curiosity without restraint. In medicine, as elsewhere in the scientific community, opinions about the merits of organized, purposeful research were mixed.<sup>78</sup> For cooperative studies to win out in the competition for research funding, support for such endeavors was needed outside as well as inside the community of medical scientists. The circumstances favoring the continued development of cooperative studies in the postwar era are the subject of the next chapter.

---

<sup>78</sup> On debates about the control and direction of postwar science policy generally, see J. Merton England, A Patron For Pure Science. The National Science Foundation's Formative Years, 1945-1957 (Washington: NSF, 1982); Daniel J. Kevles, "The National Science Foundation and the Debate Over Postwar Research Policy, 1942-1945," Isis 68 (1977), 5-26. Although perfunctory, Steven Strickland's account of the postwar climate towards medical research in particular suggests that both federal officials at NIH and academic physicians were at best lukewarm about continuing, much less expanding, the wartime involvement of government. See his Politics, Science and Dread Disease (n. 27), 26-27, 38.

Chapter Four. A Tale of Two Studies

The introduction of streptomycin toward the close of the war provided researchers a ready-made opportunity to continue the tradition of cooperative investigation. Like penicillin, streptomycin was initially in short supply, and its distribution was restricted. To CMR officials, in charge of rationing supplies of the drug, using the available quantities of streptomycin to determine its most beneficial uses seemed the appropriate response.<sup>1</sup> Preliminary research had identified tuberculosis as one of the conditions for which streptomycin showed therapeutic promise. For advanced stages of the disease, such as "military tuberculosis," the drug demonstrated dramatic and unquestioned effects. For the initial stages, its advantages over conventional therapies were far from clearcut. Additional studies were needed to

---

<sup>1</sup> Although CMR no longer controlled research funds as of January 1946, it continued to be in charge of allocating supplies of scarce drugs. The publicity given streptomycin prior to establishing production facilities on a industrial scale led producers to request that distribution of the drug be regulated. Apart from researchers conducting pilot studies who were supplied directly from drug firms, civilian allocations of streptomycin through October 1946 were to be made by Chester Keefer. See Initial Report on Operations. July 1, 1945 to March 31, 1946. Drug Section, Chemicals Division, Civilian Production Administration. Record Group 179, Box 73; Streptomycin Industry Advisory Committee, Civilian Production Administration. Meeting. January 15, 1946, RG 179, 533.8405; Civilian Production Administration, Memo, February 13, 1946 RG 179, 533.845; and Ernest M. Allen to R.E. Dyer, February 20, 1946. RG 443, NIH, Office of the Director, Box 142 [all National Archives, hereafter cited as NA.] Keefer continued to wear two hats, and as an NRC official was involved in planning the cooperative study. See Chester Keefer to Esmond Long, March 9, 1946. Esmond Long papers. Box 15. National Library of Medicine [hereafter cited as "Long papers"].

specify the precise benefits (and hazards) of treating tuberculosis with streptomycin.<sup>2</sup>

In the spring of 1946, NRC officials began meeting with representatives of the Veteran's Administration, Army, Navy, and the Public Health Service to discuss plans for a collaborative investigation of streptomycin.<sup>3</sup> Of the individuals involved, only John Barnwell, newly appointed head of the VA's Tuberculosis Division and Carroll E. Palmer, head of the PHS' Tuberculosis Field Studies Division, lacked personal experience with the NRC's wartime penicillin studies.<sup>4</sup> The continuity of personnel assured a continuity of methods and purpose

<sup>2</sup> The earliest work on TB was done under the auspices of the American Trudeau Society. By January, 1946, Streptomycin appeared "as a potentially useful adjunct to approved and time-tested therapeutic procedures in tuberculosis, but by no means a substitute for them." H. Corwin Hinshaw, et. al. "Report of the Committee on Therapy [American Trudeau Society, January 11-12, 1946]," American Review of Tuberculosis 54 (October-November, 1946), 442; H. Corwin Hinshaw and William H. Feldman, "Streptomycin in Treatment of Clinical and Experimental Tuberculosis," Annals of the New York Academy of Science 48 (September 27, 1946), 177-182; H. McLeod Riggins and H. Corwin Hinshaw, "American Trudeau Society. The Streptomycin-Tuberculosis Research Project," American Review of Tuberculosis 56 (August, 1947), 168-173.

<sup>3</sup> As the OMR official in charge of allocating streptomycin supplies, Chester Keefer was in a position to indicate the best research opportunities for the drug; as an NRC official, Keefer was involved in planning the cooperative study. See Chester Keefer to Esmond Long, March 9, 1946. Long papers. Box 15.

<sup>4</sup> The VA's Arthur Walker took charge of the streptomycin investigation after completing a history of the penicillin studies for OSRD. Esmond Long, who spent the war as a consultant on tuberculosis to the U.S. Army, joined the study as the NRC's representative, at the request of Lewis Weed, head of NRC's Division of Medical Sciences. Both Walsh McDermott and Chester Keefer brought their experience with the NRC's penicillin studies to the task of analyzing streptomycin. William B. Tucker, "The Evolution of the Cooperative Studies in the Chemotherapy of tuberculosis of the Veterans Administration and Armed Forces of the U.S.A.," Advances in Tuberculosis Research 10 (1960), 3-4; Veterans Administration, Minutes of the Third Streptomycin Conference, May 1, 2 & 3, 1947. (St. Louis: Veterans Administration, 1949), 7.

between the two investigations: a cooperative study, by accumulating more patients and handling them in a uniform manner, promised to provide results more quickly and more reliably than any independent, albeit coordinated, series of researches. <sup>5</sup>

As originally planned, the proposed streptomycin study was intended as a joint venture between the Veteran's Administration, the Army, Navy and the Public Health Service. Lack of funding prevented the PHS from immediately joining a major research initiative. <sup>6</sup> With 9,000 tuberculosis patients in its hospitals, and more on the way, the VA could not afford to wait: in June of 1946, the first of its studies began.<sup>7</sup>

---

<sup>5</sup> Projected Plan for a Joint Study of TB. April 25, 1946. Long papers. Box 15.

<sup>6</sup> In June 1946, the Public Health Service refrained from adding streptomycin to its research program due to "conditions prevailing in Congress at that time." Discussions of a study, within the PHS and by the Bureau of the Budget continued, however. National Health Advisory Council, Minutes. December 6-7, 1946. Vol 1: 51. RG 443, Box 2. NA.

<sup>7</sup> Tucker, "Evolution of Cooperative Studies," (n. 4) 6. While the VA study was getting started, joint discussions with the PHS continued. Even after beginning its own study, PHS representatives, along with various NRC figures, continued to advise the VA on their own investigation. The overlap in membership between the VA's Streptomycin Committee, the Tuberculosis Study Section of N.I.H., and the Committee on Research of the American Trudeau Society makes it difficult to determine at times which study is being discussed, especially in the spring and summer of 1946 when the VA study was just getting underway. The continuing involvement of the NRC only further complicates the problem. The key players with multiple hats were John Barrwell, who served as head of the VA's TB Division and on the N.I.H.'s TB Study Section; Esmond Long, a key figure in the American Trudeau Society, who served as the NRC's representative to the VA study, and later helped design the PHS investigation; and Carroll E. Palmer, who headed the PHS' TB Division and consulted frequently to the VA study. Other ubiquitous players include H. Corwin Hinshaw, J. Burns Amberson, Chester Keefer and Walsh McDermott.

### Research in the Bureaucracy

The VA may have seemed like the ideal organization to conduct a controlled investigation of streptomycin treatment—a centralized bureaucracy, newly invigorated by an infusion of medical and scientific talent.<sup>8</sup> The reality was somewhat different. Decisions made in the Washington office had to anticipate the objections of the veterans' lobbies, whose strength in many Congressional districts gave local concerns substantial weight with the central bureaucracy. The very idea that the VA was conducting experiments had to be approached gingerly:

...we don't like to use the word 'experiments' in the Veterans Administration; 'investigation' or 'observations,' I believe is the approved term for such a study in the VA hospitals....<sup>9</sup>

The VA's delicacy in these matters posed difficulties for what was initially planned as a controlled experiment.<sup>10</sup> In principal, the arguments for including a control group were well understood. The course of tuberculosis was highly erratic. In the absence of an untreated control group, crediting improvements to streptomycin, or any

<sup>8</sup> Paul R. Hawley, "Medical Problems of the Veterans Administration," JAMA 129 (October 13, 1945), 521-522; Paul R. Hawley, "New Opportunities for Physicians in the Veterans Administration," JAMA 130 (February 16, 1946), 403-405.

<sup>9</sup> [John Barnwell] in Veterans Administration, Minutes of the First Streptomycin Conference, December 12, 13 & 14, 1946. (Chicago: Veterans Administration, 1949), 29. See also U.S. Army-U.S. Navy-Veterans Administration-U.S. Public Health Service-National Research Council-National Tuberculosis Association, Streptomycin Conference, May 10, 1946. Long papers. Box 15.

<sup>10</sup> Streptomycin Committee, "The Effect of Streptomycin Upon Pulmonary Tuberculosis in Man—Preliminary Report of a Cooperative Study of 223 Cases by the Army, Navy and Veterans Administration," Veterans Administration Technical Bulletin TB 10-37 (September 24, 1947), 4.

novel treatment, was problematic.<sup>11</sup> But for the VA to withhold treatment from one group of patients while providing it to others required additional justification:

In general, and in particular with a disease as various and unpredictable as pulmonary tuberculosis, there can be no doubt as to the theoretical desirability of untreated controls, selected by alternation or randomization. In the laboratory, this is axiomatic. In the clinic, however, such a series seems justifiable to us on only one of two grounds: (1) a genuine ignorance or doubt that the drug in question has any therapeutic value; or (2) a shortage of supply which, by making it impossible to treat all cases, makes it fair to treat alternate cases.<sup>12</sup>

The use of a control group was better science, but the VA investigators proved unwilling to abandon existing treatments purely in the name of science. Initially, their study included a control group, but as the drug became more readily available, withholding it became more difficult. Given the potential political repercussions, the VA's commitment to the necessity for untreated controls flagged.<sup>13</sup> Despite

<sup>11</sup> ibid. The lack of a control group had hampered the efforts of the penicillin investigators to distinguish between the effects of therapy and the numerous other factors which might affect outcomes: patient selection, adjunct therapies and the natural history of the disease. See the remarks of Margaret Merrell, N.R.C.-U.S. P.H.S. Meeting of Penicillin Investigators, 7 and 8 February 1946 [n.p.], 147-148.

<sup>12</sup> Arthur M. Walker and John Barnwell, "Clinical Evaluation of Chemotherapeutic Drugs in Tuberculosis," Annals of the New York Academy of Science 52 (Dec. 14, 1949), 746.

<sup>13</sup> The initial control group was dropped within a few months, after investigators realized that there was a shortage of eligible patients in the designated VA study hospitals, too few to sacrifice half their number for a control group if the study was to produce quick answers about the merits of streptomycin. See Streptomycin Committee, "The Effect of Streptomycin Upon Pulmonary Tuberculosis," (n. 10), 4. What began as a pragmatic decision, motivated by a desire to get answers about streptomycin as quickly as possible, soon became a matter of policy. The VA's subsequent decision not to participate in the Public Health Service study of streptomycin was based in part on the conviction that the use of control groups would produce "undesirable repercussions" from "certain groups in this country." Veterans Administration, Minutes of the Fourth Streptomycin Conference, October 9, 10, 11, & 12, 1947. (St. Louis: Veterans Administration, 1947), 61.

the arguments of some consultants that a control group was necessary, the study proceeded without such a safeguard.<sup>14</sup>

The lack of an untreated control group forced the VA researchers to rely on ad hoc comparisons of study patients with the results of conventional therapy obtained in the recent past on comparable patients.<sup>15</sup> The difficulties of interpreting such comparisons soon became evident.<sup>16</sup> The study was further compromised by the decision, in October, 1947, to no longer require individual investigators to collect a two month baseline period of observations on patients before beginning treatment.<sup>17</sup> But despite its methodological shortcomings,

<sup>14</sup> The most insistent of these critics was Carroll Palmer, PHS representative to the planning group. See VA, Minutes of the First Streptomycin Conference. December 12, 13 & 14, 1946 (n. 9), 5-6; VA, Minutes of the Third Streptomycin Conference. May 1, 2 & 3, 1947. (n. 4), 155-156. Palmer was supported in his views by both VA and NAS statisticians, who were increasingly sceptical of the VA's ability to produce reliable conclusions. See Gilbert W. Beebe, Memo for Lewis Weed, Organization of Research in Use of Streptomycin in TB June 11, 1947. NRC. Division of Medical Sciences. Committee on Medicine. Subcommittee on TB: General. Archives, National Academy of Sciences [hereafter cited as NAS]. As new regimens came under consideration, the issue of controls recurred, with the same themes being sounded. See the discussion on a control group for thoracoplasty: Minutes of the Third Streptomycin Conference, (n. 4), 33-34.

<sup>15</sup> VA, Minutes of the First Streptomycin Conference. December 12, 13 & 14, 1947 (n. 9), 5.

<sup>16</sup> In an effort to prevent knowledge of treatment from influencing assessments of outcome, "blind" evaluations of x-ray results were obtained from observers who did not know whether or not they were scoring cases which had received streptomycin. Statisticians reviewing these results reported that inter-observer agreement about the degree of improvement was no better than might have been expected by chance. See VA, Minutes of the Third Streptomycin Conference. May 1, 2 & 3, 1947. (n. 4), 147-165, especially 150-152; Streptomycin Committee, "Effect of Streptomycin Upon Pulmonary Tuberculosis," (n. 10), 6.

<sup>17</sup> VA, Minutes of the Fourth Streptomycin Conference October 9, 10, 11, & 12, 1947 (n. 13), 67.

the VA study remained the largest, if not the only, program investigating streptomycin treatment:

"...absolutely the whole profession is going to have to depend on the Veterans Administration to tell us what we are going to be able to learn about streptomycin. There is no other organization which is likely to be able to learn about streptomycin on such a wide scale....<sup>18</sup>

The value of their study to the community of tuberculosis researchers depended upon the VA's capacity to treat large numbers of patients according to a standard protocol. To deliver on this promise, the VA investigators had to secure effective compliance with the aims and conditions of therapeutic research:

...the integrity of this whole thing depends upon regarding it as an investigative job. It is not indifference to the welfare of the patient in my mind but there is something much more important than the welfare of any particular patient--we are trying to find out something which will be of use to a great many thousand patients, in the future, or not of use.<sup>19</sup>

In maintaining the scientific "integrity" of their study, organization and ideology were the VA's principal assets. In the long run, neither proved adequate. Initially, the "Streptomycin Committee" placed in charge of the research made all decisions about which VA

<sup>18</sup> The importance of the VA program only increased as other projects encountered funding difficulties. [H. Corwin Hinshaw] in Veterans Administration, Minutes of the Second Streptomycin Conference. January 23 & 24, 1947. (Chicago: Veterans Administration, 1949), 87, 50-51; see also [Esmond Long], Veterans Administration, Minutes of the Fifth Streptomycin Conference. April 15, 16, 17 & 18, 1948 (Chicago: Veterans Administration, 1948), 149. Apart from having access to larger numbers of patients, the VA could also expect a greater proportion of their patients to return for follow-up, as such exams were a condition of receiving disability checks. Walker and Barrwell, "Clinical Evaluation of Chemotherapeutic Drugs in Tuberculosis," (n. 12), 742-743.

<sup>19</sup> [Arthur M. Walker] in VA, Minutes of the Second Streptomycin Conference. January 23 & 24, 1947. (n. 18), 59.



institutions and patients were eligible to participate.<sup>20</sup> Backed by the authority of the VA's medical director, the Streptomycin Committee was able to restrict use of streptomycin to the elite group of VA hospitals and physicians "most competent" to assess the drug. So long as little streptomycin was available, centralized allocation according to the guidelines of the research protocol seemed effective. But as supplies of the drug improved, regulating the use of streptomycin within the VA became increasingly difficult.<sup>21</sup>

As with penicillin, publicity concerning streptomycin created a demand for the drug among both patients and physicians. Of the two groups, patients and their families seemed easier to manage.<sup>22</sup> Handling the demands of VA physicians and hospitals prevented from using streptomycin proved more difficult. To the study's leadership, including more hospitals in the study seemed the only alternative to

---

<sup>20</sup> On the organization of the study, see VA, Minutes of the Second Streptomycin Conference. January 23 & 24, 1947. (n. 18), 66-67, 72-74 and, more generally, Tucker, "Evolution of Cooperative Studies," (n. 4), 8-15.

<sup>21</sup> VA, Minutes of the Second Streptomycin Conference. January 23 & 24, 1947. (n. 18), 5-6, 67, 69, 84-85. Discussions of the implications of increasing streptomycin supply began almost as soon as the study started. See A.M. Walker to John B. Barnwell, Esmond Long and George Owen, June 17, 1946. In November 1946, less than six months into the study, the VA decided to make streptomycin available generally available for all non-TB uses, while keeping the Streptomycin Committee in charge of allocations for TB cases. Paul Hawley to Esmond Long, November 25, 1946. [Both Long papers, Box 15.]

<sup>22</sup> "What we tell the relatives is this--that this is an experimental drug, that its efficacy has not been proven and that we feel sure they would not want their own husband, father or brother being experimented upon because we have heard so many times, complaints about people being experimented upon." Patients who persevered were successfully discouraged by a detailed enumeration of hazards on the consent form they were asked to sign. [Delmar Goodel] in VA, Minutes of the Second Streptomycin Conference. January 23 & 24, 1947. (n. 18), 88.

abandoning their research entirely. If streptomycin had to be made more available, then enlarging the study seemed the most prudent course.<sup>23</sup> But as the study grew larger and more complex, continuing centralized selection and monitoring of patients seemed unfeasible. The alternative was to delegate control of the study to the regions, but with the regions in charge, "each and every individual in our offices who has political power will attempt to break down the investigative program into one of purely therapy."<sup>24</sup>

Fears about the consequences of losing central control proved well-founded. With an increasing number of VA institutions enrolled in the study, ability to ensure compliance with the protocol diminished.<sup>25</sup> As reports from other, smaller, studies became available, the impulse to explore new directions suggested by these findings grew stronger.<sup>26</sup>

---

<sup>23</sup> By December, 1946, the principal investigators believed that restricting the study (and use of the drug) to the original units was no longer a viable strategy. Arthur M. Walker to John B. Barnwell, Esmond Long, H. Corwin Hinshaw, et. al., December 3, 1946. Long papers, Box 15. For the arguments, pro and con, on expanding the study, see VA, Minutes of the Second Streptomycin Conference. January 23 & 24, 1947. (n. 18), especially 65-74.

<sup>24</sup> [F.G. Bell] in VA, Minutes of the Second Streptomycin Conference. January 23 & 24, 1947. (n. 18), 72. Bell's fears proved justified, to judge from the various remarks at the next meeting.

<sup>25</sup> While some regions experienced no problems, others alluded to difficulties keeping VA physicians in line. In general, there were two distinct kinds of protocol violation: one involved continuing the study at sites without the correct laboratory facilities to monitor reactions to the drug; the other involved selecting patients who did not fit the protocol's rules of eligibility, or varying the treatment regimen in some fashion. It was the latter type of deviation which caused difficulties in interpreting results. VA, Minutes of the Third Streptomycin Conference. May 1, 2 & 3, 1947 (n. 4), 63, 67, 69, 152-153, 157.

<sup>26</sup> See the discussion in VA, Minutes of the Fifth Streptomycin Conference. April 15, 16, 17 and 18, 1948. (n. 18), 147-150.

University based VA affiliates posed a particular problem:

Streptomycin is a new toy with a lot of our attending men and in one of our hospitals I think our Dean's Committee has been wanting to use it. We are not particularly sure about the type of cases in which they are using it. <sup>27</sup>

Once decisions about eligibility were decentralized, moral suasion provided the principal means for ensuring that participating investigators were following the research protocol:

It is strongly urged that each unit remember that this is an experiment. Whenever any question of interpretation arises in the selection of cases or in the post-treatment management, the investigator should adopt that course which will supply the most valid evidence. <sup>28</sup>

Apart from appeals to the scientific conscience of researchers, threats that the study would be defunded "if it [streptomycin] becomes too widely disbursed" were the leadership's only remaining recourse for persuading errant investigators. <sup>29</sup>

<sup>27</sup> [H.L. Mantz], VA, Minutes of the Third Streptomycin Conference, May 1, 2 & 3, 1947 (n. 4), 66 and discussion, 57-58. The study directors had only partial success in persuading local VA investigators that they were "not telling a medical school how to treat a patient" but only "telling them what the requirements are ... to admit a case to this investigation." Ibid., 58.

<sup>28</sup> Investigators were granted permission to use the drug on ineligible patients at their discretion, so long as they did not include those patients in their reports on the study. Streptomycin Committee to Study Units, December 20, 1946. Long papers. Box 15.

<sup>29</sup> The threat of a cutoff came from the Bureau of the Budget, which expressed concern that the project be conducted "wholly on an investigative basis". [John Barnwell], VA, Minutes of the Third Streptomycin Conference, May 1, 2 & 3, 1947 (n. 4), 51. In 1949, cutbacks in the size of the study enabled the directors to request that units which did not feel comfortable following the protocol elect to leave "active" status in the study; despite these initiatives, protocol deviations still continued. Veterans Administration, Minutes of the Seventh Streptomycin Conference, April 21, 22, 23 & 24, 1949. (Washington: Veterans Administration, 1949), 335-336.

To non-medical observers, the VA study looked like their idea of research: focused, purposeful investigation in a large bureaucracy.<sup>30</sup> The watchword of the VA was organization, but the organizational means available to the VA study ultimately proved inadequate to the task. To the medical community, the VA's investigation of streptomycin demonstrated the limits of organization alone in producing convincing findings. Another avenue to producing reliable cooperative studies had to be found.

#### Politics in the Service of Science

While the VA study was getting underway, the Public Health Service was planning its own researches into streptomycin. Mindful of the penicillin experiences, PHS advisors anticipated some of the difficulties physicians would face in following the research protocols:

Innumerable physicians throughout the country will be treating small series of patients with this or that regimen and will be publishing their results. This will constitute a pressure in the form of competition which is most difficult to resist. We must be prepared, however, to accept the risk that some one of these unsponsored programs may discover something which we have not yet had an opportunity to study. We can do this only if we are secure in our minds that what we are studying, limited in scope as it will be, should provide us with some tangible answers at the end of a net period of time.<sup>31</sup>

---

<sup>30</sup> "If we only had the haphazard grant in aid system, the chances are that we would not yet have electron tubes and television would only be a dream. No industrial laboratory would think of conducting research as it is now conducted in medicine." New York Times editorial, 1944, quoted in Richard Harrison Shryock, American Medical Research (Commonwealth Fund, 1947), 104.

<sup>31</sup> [Walsh McDermott] to Doctors Palmer, Bogen, Barnwell, Hinshaw, Willis and Long, Suggestions for the report from the Panel on Dose Regimens to the Tuberculosis Study Section [March 27, 1947]. Long papers, Box 15.

To keep participating physicians in line, the planners called for a strictly defined protocol, with explicit rules about eligibility and treatment schedules. Physicians would be expected to continue treating patients on a given regimen, until authorized by a steering committee to discontinue treatment.<sup>32</sup>

PHS officials contemplated a organized program of cooperative research, complementary to the ongoing VA investigation. Individual tuberculosis researchers, however, wanted the PHS to sponsor a free ranging program of research, not purely confined to evaluating streptomycin treatment in humans.<sup>33</sup> The announcement of cutbacks in Congressional appropriations for studying tuberculosis forced researchers to accept a change in

...philosophy from free research to a target [sic] study directed at the specific question of the merit of streptomycin in tuberculosis therapy. [....] The essence of this portion of the program, as distinguished from the various proposals of the Study Section, is

---

<sup>32</sup> [Walsh McDermott] to Doctors Palmer, Bogen, Barnwell, Hinshaw, Willis and Long, Suggestions for the Report from the Panel on Dose Regimens to the Tuberculosis Study Section [March 27, 1947]. Long papers, Box 15. Even at this date, incorporation with the VA study was not yet ruled out. Two of the "advisors", Bogen and Barnwell, were deeply engaged in the VA study, while the others (especially McDermott and Long) were actively involved in discussions of the VA's interim results and future plans.

<sup>33</sup> The initial request to Congress was for an appropriation of \$3m, of which \$1.25m was intended for research on streptomycin independent of the cooperative project. See testimony of R.E. Dyer, U.S. Congress, House of Representatives. Committee on Appropriations. Hearings. Department of Labor-Federal Security Appropriation Bill for 1948. February-March 1946. Part II, 491. 80th Congress. 1st Session. I have been unable to trace the records of the TB Study Section's deliberations; however, it appears as if a Steering Committee consisting of Esmond Long, Walsh McDermott, H. Corwin Hinshaw, Carroll Palmer, H. McLeod Riggins and H. Stuart Willis found itself in the position of having to trim the research plans without help from the study section. See H. Stuart Willis to Esmond Long, May 16, 1947 and Long's reply, May 19. Long papers, Box 15.

that a group of special experts in the field of clinical tuberculosis, in different institutions, in different parts of the country, agree to cooperate in a large scale, rigidly controlled project, which is operated in such a way as to insure the collection of uniform observations that may be combined or pooled to furnish statistically significant evidence in the treatment of certain well defined types of pulmonary tuberculosis.<sup>34</sup>

The Bureau of the Budget approved funding for the PHS study on condition that the research would "be carefully coordinated with similar work by other government agencies ... and be closely controlled in extent and direction by the Study Section, the [National Health Advisory] Council and appropriate specialists in the Public Health Service." The Bureau of the Budget's principal concern was that appropriations for medical treatment not be slipped in under the guise of research. There is no evidence that BoB had any particular interest in the details of the experimental design, but BoB's intervention provided PHS officials with an opportunity to engage the contentious issue of experimental controls.<sup>35</sup> They insisted that the PHS study, unlike the ongoing VA investigation, contain a pre-selected control group of patients who did not receive streptomycin:

---

<sup>34</sup> National Health Advisory Council, Minutes. June 6-7, 1947, V. 1: 168. RG 443. N.I.H. Office of the Director. Minutes. National Advisory Health Council, 1945-1960. Box 2. NA.

<sup>35</sup> National Health Advisory Council, Minutes. June 6-7, 1947 V. I: 167-168. RG 443, N.I.H. Office of the Director. Minutes. National Advisory Health Council, 1945-1960. Box 2. NA. BoB's own records shed little light on the matter, but are consistent with this interpretation. See Mr. C. Martin to Director, January 5, 1948. RG 51. BoB Estimates Division. Series 39.14a. Federal Security Agency. 1939-1952. Box 12. NA.

The cases chosen by the Panel shall, by proper random device, to avoid all possibility of bias, be divided by the Central Unit into cases for treatment and cases for control.<sup>36</sup>

Like the VA investigators, PHS officials anticipated difficulties from physicians asked to withhold streptomycin from one group of patients while treating others with the drug:

...it seems very likely that the men responsible for various phases of this project may encounter criticism from people who are already convinced of the value of streptomycin, or who for some other reason do not consider necessary a program providing for withholding the drug from one group of patients. Since we have agreed to go ahead with such a program, it is important to protect the individual investigators from possible serious consequences of this criticism.<sup>37</sup>

Advocates of a control group wanted backing from the medical authorities on the study's Steering Committee, in the form of a statement justifying the withholding of streptomycin. The limited amounts of streptomycin available, coupled with uncertainty about the drug's precise value, could serve as an initial justification. Sceptics were sure that any such statement would serve its purpose: to stiffen

---

<sup>36</sup> Minutes. Meeting of the Tuberculosis Study Section Steering Committee and Special Consultants May 24-25, 1947. Long papers. Box 16. There appears to have been some disappointment with the VA's decision to proceed without controls; see the call for "more adequately controlled" studies from the American Trudeau Society: "Annual Report of the Committee on Therapy and the Subcommittee on Streptomycin," JAMA 135 (November 8, 1947), 642. One of the strongest advocates of controls was Carroll E. Palmer, Director of Field Studies for the PHS' Tuberculosis Division. See Esmond Long, "The Award of the Trudeau Medal for 1972," American Review of Respiratory Disease 106 (1972), 627.

<sup>37</sup> Carroll E. Palmer to Esmond R. Long, October 21, 1947. Long papers, Box 15.

the backbone of investigators faced with a patient whose conditions was deteriorating.<sup>38</sup> The proposed compromise was

...that physicians do not communicate to patients the fact that they are being considered for inclusion in this series. Hence patients who are in the control group are not to realize that they have been denied streptomycin.<sup>39</sup>

The majority of investigators participating in the PHS study proved willing to go along with the idea of a control group.<sup>40</sup> What remained unresolved was the handling of control patients whose disease worsened substantially during the study. Should they receive the drug, and under what circumstances? PHS representatives proposed that investigators submit such cases to an Appeals Board which would decide if an exemption was warranted. Provided the exemption criteria were sufficiently narrow, and specified in advance, only a few patients would be lost and the research design need not be compromised.<sup>41</sup> This

<sup>38</sup> Carroll E. Palmer to Esmond R. Long, October 21, 1947; H.C. Hinshaw to Carroll E. Palmer, October 29, 1947; Walsh McDermott to Carroll E. Palmer, November 3, 1947. Long papers, Box 15. Hinshaw's and McDermott's letters make it clear that both were uncomfortable with signing a statement denying that streptomycin had any value, but whereas Hinshaw proposed a modified statement justifying a control group, McDermott thought that the best approach was saying nothing at all on the subject.

<sup>39</sup> H.C. Hinshaw to Carroll E. Palmer, Proposed Statement of Investigations.... October 29, 1947. Long papers, Box 15.

<sup>40</sup> It appears, however, that both control patients and those receiving streptomycin were permitted to have other traditional treatments such as "collapse therapy"; when the control group is referred to as "untreated" it is the use of streptomycin which is meant. In the event, 79.2% of the streptomycin patients and only 73% of the controls received no surgical intervention. Esmond R. Long and Shirley H. Ferebee, "A Controlled Investigation of Streptomycin Treatment in Tuberculosis," Public Health Reports 65 (November 3, 1950), 1424.

<sup>41</sup> TB Study Section. Steering Committee. Minutes, November 22, 1947, 6. Long papers, Box 16; Shirley H. Ferebee to Esmond Long, December 2, 1947. Long papers, Box 15.



proposal only altered the terms of the debate. According to one dissenting study section member, it all boiled down to a question of clinical integrity:

As a matter of fact I do not believe it is possible to give a definition [of life-threatening conditions] which would cover all the possibilities. Fundamentally, it rests on the judgement of the physician who is treating the case and who knows the patient best. He is in a far better position than anyone else to make the decision. If he is capable of undertaking a clinical investigation of therapy, he is certainly capable of assuming the responsibility for such judgement. <sup>42</sup>

To advocates of experimental controls, this approach, if allowed free rein, "would completely invalidate the control study" and "jeopardize the entire program of the Study Section." <sup>43</sup>

The PHS study of streptomycin was much smaller and less complex than the VA investigations, enabling PHS officials to exercise an unusual degree of influence over the conduct of the research. Decisions

<sup>42</sup> J. Burns Amberson to Esmond R. Long, December 18, 1947. Long papers, Box 15. The continued obstructions placed by Amberson in the face of maintaining effective controls are all the more striking as in 1931 he had conducted what appears to be the first U.S. clinical study employing an untreated control group where the assignment of treatment/no treatment was left to chance. J. Burns Amberson, B.T. McMahon and Max Pinner, "A Clinical Trial of Sanocrysin in Pulmonary Tuberculosis," American Review of Tuberculosis 24 (1931), 401-435.

<sup>43</sup> H. McLeod Riggins to Esmond Long, November 14, 1947; H. McLeod Riggins to H. Stuart Willis, November 10, 1947. Long papers, Box 15. The subsequent decision to create an Appeals Board did not, however, permanently resolve the underlying issue: questions about the scope and operating procedures of the appeals process continued to recur. See Tuberculosis Study Section, Report of Informal Meeting, June 17, 1948. Long papers, Box 16. This discussion implies that some physicians were referring cases which could not possibly meet the appeals criteria, perhaps in the hopes of having the Appeals Board take responsibility for withholding the drug. The document also raises questions about the investigators' understanding of the concept of blind allocation to treatment and controls: at several points it is suggested that the panel reviewing patient eligibility would be better off knowing whether the patients are intended for control or treatment.

about admitting patients to the study were centrally reviewed, and subsequent assignment to treatment or control groups was handled by the study's statistical unit. But organization alone could not forestall the desire of investigators to raise questions which were not contemplated in the original research plans.

Nearly 18 months into the study, the Evaluation Policy Committee proposed that "an adequate evaluation [of outcome] must take into account everything that can be known about a patient," including data that only the treating physician could provide.<sup>44</sup> The VA study had begun to demonstrate problems with the traditional reliance on roentgenographical measures of outcome.<sup>45</sup> Holding an improvised case conference on each patient, clinicians argued, would "lend greater accuracy to interpretations of questionable features and in the long run give greater significance to the interpretation of results."<sup>46</sup>

---

<sup>44</sup> Shirley H. Ferebee to J. Burns Amberson, October 20, 1948. Long papers. Box 15.

<sup>45</sup> William B. Tucker, "Evaluation of Streptomycin Regimens in the Treatment of Tuberculosis. An Account of the Study of the Veterans Administration, Army and Navy, July 1946 to April 1946," American Review of Tuberculosis 60 (1949), 745-746. Lawrence B. Hobson and Walsh McDermott, "Criteria for the Clinical Evaluation of Antituberculosis Agents," Annals of the New York Academy of Sciences 52 (December 14, 1949), 782-787. The initial VA discussions (prior to 1949) emphasize the problem of interobserver agreement in interpreting x-rays, and means to improve it. Hobson and McDermott's discussion implies a more fundamental problem: x-rays are simply not good prognostic indicators in early pulmonary tuberculosis; observer variation simply aggravates the lack of sensitivity inherent in the method, which cannot detect at an early stage of disease the tissue changes which are predictive of subsequent course.

<sup>46</sup> J. Burns Amberson to Esmond R. Long, November 3, 1948. Long papers. Box 15.

To the statisticians in charge of the PHS study, there were enough difficulties producing trustworthy scores for data they had agreed to collect, without trying to introduce clinical material which was neither standardized nor uniformly available:

It would be a tragic mistake to distort the original pattern of the study now to try to make it yield information it was not designed to produce, because in so doing, the kind of answers it can give will lose their validity. <sup>47</sup>

The statisticians' inclination was to distrust measures which could not be reliably reproduced: in this, as in other respects, the future of clinical research was theirs. <sup>48</sup>

Both the PHS and VA cooperative investigations remained in operation for well over a decade, evaluating newer drugs in the treatment and prevention of tuberculosis, and serving as models (and training grounds) for physicians interested in therapeutic research.<sup>49</sup> The primary interest of the VA and PHS studies, however, lies neither in their scientific accomplishments nor in their subsequent historical

<sup>47</sup> Shirley H. Ferebee to J. Burns Amberson, October 20, 1948. Long papers. Box 15.

<sup>48</sup> The soliloquy of Emil Bogen on the subject of clinician's reliability is quite interesting in this regard. To the statisticians' plaint that experienced clinicians cannot agree in even in judging x-rays, Bogen replied, in effect: they can if you throw out the bad films. Veterans Administration, Minutes of the Eighth Streptomycin Conference. November 10, 11, 12, & 13, 1949 (Washington: Veterans Administration, 1949), 279.

<sup>49</sup> On the subsequent activities of the VA and PHS enterprises, see the discussions by Harry Dowling, "The Emergence of the Cooperative Trial," Transactions and Studies of the College of Physicians of Philadelphia 4th series 43 (July 1975), 20-29; Shirley Ferebee Woolpert, "Acceptance of the Trudeau Medal for 1972," American Review of Respiratory Disease 106 (1972), 629-630; and that of William Tucker, "Evolution of Cooperative Studies," (n. 4).

influence, but in what they can tell us about the changing purposes and means of therapeutic investigation.

### Good Science and Organized Research

The initial impetus for cooperative studies came from a desire to improve the quality of therapeutic research. It was only natural that during World War II, the specialists at the NRC should turn to cooperative studies as the quickest way to obtain reliable answers about the merits of novel therapies. Whether or not the penicillin investigations deserved this confidence remains an open question: the NRC's findings concerning the drug's use in treating syphilis can hardly be described as timely. Nonetheless, during the war cooperative research came to represent the ideal blending of science with efficiency: to the talents of individual researchers, cooperative studies added the leaven of organization. Both within and beyond the medical community, the development and clinical investigation of penicillin served as an eloquent testimony to the virtues of cooperative research.<sup>50</sup>

The streptomycin studies conducted by the VA and PHS represent the principal attempt to carry on the tradition of cooperative studies immediately after the war. Among those planning the streptomycin studies, there was no question that adequate therapeutic research demanded the participation of experienced investigators, knowledgeable about the vagaries of tuberculosis and the mechanisms of drug action

---

<sup>50</sup> The wartime work with penicillin, along with that on radar, were the two examples with which Vannevar Bush, head of OSRD, introduced his report to President Roosevelt on postwar science policy, Science. The Endless Frontier (Washington: U.S. Government Printing Office, 1945), 5.

and resistance. If the effects of streptomycin on tuberculosis were going to be studied, the general practitioners of Paducah would not be doing the research. What distinguished the VA and PHS studies from other work on streptomycin was not the involvement of specialists, but their apparent willingness in this instance to subordinate individual judgement to a common purpose.

If decisions about the future of these studies had been left solely up to the scientific community, it is an open question whether they would have been supported, or if they would have taken the precise form they did. But to those footing the bill, the streptomycin studies represented organized, purposeful research: a means of quickly finding answers to practical questions about the therapeutic use of streptomycin. What gave cooperative studies like these a competitive advantage, in the quasi public debates over the direction and funding of postwar medical research, was not their association with better science but their reputation for efficiency.

The irony, then, of the VA and PHS studies of streptomycin is that, in tandem, they demonstrated the inadequacy of good organization alone to produce good science. To the generation of clinical investigators trained before the war, the participation of specialists in joint projects of therapeutic evaluation was in itself a partial guarantee of a study's scientific merit. The VA investigations of streptomycin demonstrated that specialists, no less than anyone else, were capable of self-deception, selecting the most or least promising

cases for treatment, depending on their particular prejudices.<sup>51</sup>

Both the VA and PHS testified favorably on behalf of streptomycin, but it was the PHS studies, properly randomized, which received credit for demonstrating the new drug's benefits in treating tuberculosis.<sup>52</sup> To contemporaries, the lessons were clear: cooperation and expertise, planning and standardization, were necessary but not sufficient to ensure a successful investigation.

Long after the technical details of the procedures they employed were obsolete, the PHS' studies of streptomycin served as an example of

<sup>51</sup> [Morris C. Thomas], VA Minutes of the Fifth Streptomycin Conference. April 15, 16, 17 and 18, 1948. (n. 18), 15; William Stead, "A Suggested Change in the Method of Randomization of Patients in Therapeutic Trials," Veterans Administration, Transactions of the 16th Conference on the Chemotherapy of Tuberculosis. February, 1957. (Washington: Veterans Administration, 1957), 117-119. Manipulation of treatment assignment by physicians who "randomized" on the basis of patients' chart numbers was widely suspected in another VA study, of anti-coagulants, conducted in the 1950's. See Louis Lasagna, "The Controlled Clinical Trial: Theory and Practice," Journal of Chronic Disease 1 (April, 1955), 357-358.

<sup>52</sup> The PHS shared credit for demonstrating the value of streptomycin with the British Medical Research Council, whose trials also used randomized controls. The VA subsequent decision to "finally" adopt randomization was influenced by disagreements between the three cooperative groups (VA, PHS and MRC) concerning the relative value of isoniazid versus streptomycin plus PAS. William B. Tucker, "A Controlled Study of the Variables in the Chemotherapy of Pulmonary Tuberculosis. An Account and Critique of the Investigation by the Cooperative Group of the Veterans Administration, Army and Navy, 1946-1953," Veterans Administration, Transactions of the 12th Conference on the Chemotherapy of Tuberculosis. February 1953. (Washington: Veterans Administration, 1953), 31-32; Tucker, "Evolution of Cooperative Studies," (n. 4), 28.

It is difficult to reconcile Tucker's claim in 1960 that "repeated checks" found only "minor deviations" from randomization (30) with Stead's observation as late as 1957 that statistically significant differences existed among the treatment arms in severity of disease. What is clear is that even "internal" critics of the VA study such as Stead and Tucker continued to believe in the VA findings: the adoption of centralized randomization was to convince others. Stead, "A Suggested Change in the Method of Randomization," (n. 51) 119.

scientific progress in therapeutics. Along with centrally controlled randomization, their use of objectively measured indicators of progress and blinded assessments of therapeutic outcomes constituted adherence to a program of methodological reform in the postwar era. The rationale contemporaries offered for such innovations was that they served to limit the exercise of subjective judgment: rather than pitting the clinical acumen of individual physicians against each other, evaluations conducted according to the canons of methodology would provide an objective measure of therapeutic progress. What went unmentioned was that these procedures also reduced the clinician's ability to deviate spontaneously from an agreed upon plan of research, whatever the reason.

The scientific accomplishments of the PHS study were credited to its superior methodology. But the methodological advances adopted by the PHS served organizational purposes as well. By taking decisions about treatment assignment and outcome evaluation out of clinicians' hands, they provided researchers with a mechanism which reduced the investigator's opportunity to change his mind in mid-stream about the methods and purposes of a study. None of these innovations could eliminate the need for someone to enforce the details of the experimental protocol. But putting the central statistical office in charge of this task relieved participating clinicians of the duties of policing themselves. How and why statisticians came to play the policeman's role is the subject of the next chapter.

Chapter Five. Managing Chance. Statistics and Therapeutic  
Experimentation After World War II

In the decades following World War II, unprecedented numbers of physicians began to be aware and make use of the methods and concepts of statistics. For many of these individuals, the randomized controlled trial (RCT) came to represent both the symbol and substance of the statistical method in medicine: "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."<sup>1</sup> This chapter will examine the intellectual revolution brought about as statistics and statisticians, formerly relegated to the sphere of public health, began to make substantial inroads on the domain of clinical medicine.

The occasional use of experimental controls in therapeutic research has a venerable history. The hagiographical tradition traces the practice at least as far back as James Lind's 18th century experiments in the British Navy using citrus fruit to prevent scurvy.<sup>2</sup>

---

<sup>1</sup> John Bailar, "Introduction," in Stuart Shapiro and Thomas Louis, Clinical Trials: Issues and Approaches (New York: Marcel Dekker, 1983), 1.

<sup>2</sup> Lind's example is one of the most commonly cited by authors tracing the genealogy of controlled therapeutic experiments. J.P. Bull, "The Historical Development of Clinical Therapeutic Trials," Journal of Chronic Diseases 10 (1959), 218-248; Abraham Lillienfeld, "Ceteris Paribus: The Evolution of the Clinical Trial," Bulletin of the History of Medicine 56 (Spring, 1982), 1-18. What these reviews generally overlook is that Lind's insight was lost to subsequent generations of the British Navy, due to Lind's use of the contemporary terminology, "lime juice," to designate a ration largely consisting of juice from lemons; 19th century polar explorers, issued rations of juice from limes, discovered to their detriment that Lind's remedy no longer worked. See Alice Henderson Smith, "The Relative Content of Anti-



But like any search for precursors, the quest for forgotten ancestors in the history of statistics and medicine risks leading us astray in understanding the more recent past. Once we have noted that Lind employed a control group in his study of scurvy prevention in 1747, that Hewlett used "blind" testing in his 1913 comparison of synthetic and 'natural' salicylates, or that Evans and Hoyle alternated the use of various active remedies with placebo in their 1933 study of angina treatments, what do we conclude?<sup>3</sup> That physicians took a long time to understand and accept the intellectual framework and techniques of experimental inference? But what exactly are we to take as evidence such a "lag" existed?

Whether or not a cultural lag exists in medicine's acceptance of "modern" statistical concepts in therapeutic research depends on which "event" one takes as evidence that a statistical framework was adopted. Is it denoted by the initial awareness that spontaneous recoveries (chance) as well as treatment affect the recovery of patients? the insistence on studying large numbers of patients to offset the effects of chance? the first use of statistical tests of significance to

---

worked. See Alice Henderson Smith, "The Relative Content of Anti-ascorbic Principle in Limes and Lemons. B. Historical Inquiry," Lancet ii (November 30, 1918), 737-738. Lind's lack of knowledge about the underlying active agent which worked to prevent scurvy suggests a different moral than that suggested by the hagiographers of experimental controls: a purely empirical knowledge of therapeutics, however methodologically sound, has its limitations and even its hazards.

<sup>3</sup> For the less celebrated examples cited, see A.W. Hewlett, "Clinical Effects of 'Natural' and 'Synthetic' Sodium Salicylate," Annual Report of the Investigations Carried Out Under the Supervision of the Therapeutic Research Committee of the Council on Pharmacy and Chemistry 2 (1913), 30-36; William Evans and Clifford Hoyle, "The Comparative Value of Drugs Used In the Continuous Treatment of Angina Pectoris," Quarterly Journal of Medicine 7 (1933), 311-338.

evaluate experimental effects? the first use of randomly assigned control groups? the first instance in which professionally trained statisticians play a major role in experimental design? Any one of these marker events might be said to denote an awareness of statistics, and for each one, we might find one or more competing claims to be "first", spread out not merely over several decades but several centuries.

In understanding the revolution wrought by statistics in the latter half of the 20th century, it is desirable to abandon the long view, and consider the following questions: when clinical researchers began to employ randomized controls in therapeutic research, what did they intend to accomplish? What problems, intellectual and practical, did they hope to solve? How was the conduct of therapeutic research affected by the new ideas about experimental design and practice introduced after World War II? And what role, if any, did statisticians play in bringing about these changes?

It is tempting to regard the adoption of statistical methods and concepts in biomedical research as a classical instance in the diffusion of new scientific ideas. Evidence of the diffusionists' s-shaped curve is not wanting: a few isolated publications introducing statistical concepts in the early 1930's; followed by a steadier stream of articles at the end of the decade, indicating the use of such practices in specific laboratory fields; and culminating in a crescendo of new applications and developments after World War II. But to regard the evidence in this way is to accept the inevitability of progress and to

overlook the specific circumstances, intellectual and social, in which these innovations began to appear attractive.

Therapeutic reformers and investigators had long been interested in improving the quality of clinical drug research. As manufacturers in the postwar period produced an increasing number of drugs meriting clinical consideration, reformers sought a dependable means for distinguishing products with therapeutic merit from those which merely had good copy writers. Increasingly, they turned to statisticians as colleagues and as mentors in this task. The statisticians' contribution to this alliance was not only to assist researchers in the design and analysis of experiments, but to present their ideas and innovations about experimentation to the broader medical community as if they represented little more than the conventional wisdom of conscientious and intelligent investigators. The result was an incomplete revolution, one in which the majority of physicians were acquainted with neither the intellectual power which lay behind the procedures advocated by statisticians nor with the limitations of statistical methods.

The geneticist and statistician R.A. Fisher is generally credited with orienting the theory of experimental design within the conceptual framework of statistical inference. The initial section of this chapter accordingly examines Fisher's ideas about the uses and purposes of randomization in experiments. Fisher's views and arguments about experimental controls are contrasted with pre-World War II ideas about controls in therapeutic research, as well as with the arguments and concepts presented by post-war advocates of randomized experiments in medicine. Subsequent sections examine the allegiance of statisticians

to the cause of therapeutic reform, and the changing relation between statistician and clinician implicit in postwar ideals of therapeutic experimentation. The concluding section explores the tension between the ideologies of statisticians and physicians in their views of medical treatment and therapeutic research.

### **The Management of Chance: Three Perspectives on Experimental Controls**

Most recent students of biostatistics will recall their first introduction to the fallacy of hasty conclusions. Consider, their mentors advised, the comforting illusion engendered when the initial patients treated with a new drug respond dramatically to the innovation. Do not be deceived, statisticians warn: chance alone could easily produce a run of spontaneous recoveries. Only a proper attention to experimental design and the procedures of statistical analysis can protect the unwary. However salutary the lesson, statisticians were hardly the first to call attention to such chance effects, nor the first to advocate the use of experimental controls to guard against them. To understand their precise contribution to the theory and practice of experimental design, it is necessary to consider the views and practices of their predecessors.

In the initial decades of the 20th century, experienced researchers repeatedly counseled the inexperienced physician about the role of chance in creating the illusion of effective treatment. The use of experimental controls was one means to protect oneself against this illusion, as were care in the selection of cases and efforts to study the outcomes of treatment in large numbers of patients. A purely

statistical knowledge, such as might be gained by the comparative review of large series of cases was, however, regarded as inferior.<sup>4</sup> Ideally, clinical researchers aspired to the determinate conditions of the laboratory investigation, where the factors that affected outcomes were both known and manipulable. Yet even the best clinical study could only approximate this ideal:

Clinical observations can be made just as accurate as laboratory observations; but in the human subject, observation cannot be as readily controlled, the conditions cannot be so easily kept uniform or varied—in one word, the problems cannot be analyzed, as they can be in the animal.<sup>5</sup>

Unable to stabilize the conditions of therapeutic research, clinical investigators in the first half of the 20th century sought to master uncertainty by accumulating experience. Experience alone brought detailed knowledge of the vagaries of specific disease, knowledge which might then be applied in devising proper experimental controls. Within this context, a 'well-controlled' experiment might refer to one in which a carefully selected comparison group was employed to "control for" the effects of disease severity and spontaneous

<sup>4</sup> See, for example, the arguments of W.D. Sutcliffe in "Adequate Tests of Curative Therapy in Man," Annals of Internal Medicine 10 (July, 1936), 89-96. Sutcliffe, an advocate and practitioner of controls, nonetheless accepted most of Claude Bernard's criticisms of statistical approaches to therapeutics and experimentation.

Even statisticians of this era might espouse reservations about statistics rooted in an underlying determinism: "...let the experimenter who is driven to use statistical methods not forget this, that the very fact that he is compelled to use statistical methods is a reflection on his experimental work. It shows that he has failed to attain the very object of experiment and exclude disturbing causes." G. Udny Yule, The Function of Statistical Method in Scientific Investigation (London: His Majesty's Stationary Office, 1924), 5. Medical Research Council. Industrial Fatigue Research Board. Report No. 28.

<sup>5</sup> Torald Sollman, "Experimental Therapeutics," JAMA 58 (January 27, 1912), 243.

recovery, but it might equally refer to a study in which the perturbing factors of diet, co-morbidity and the like had been minimized to the greatest extent possible.<sup>6</sup> In both instances, the use of 'controls' depended on the experimenter's ability to recognize those circumstances which might affect the results of treatment. The value of a study depended on the prior state of knowledge concerning disease and treatment: the more that was known, the better the experiment which could be designed. It also depended, as we have seen in previous chapters, on one's confidence in the experimenter: the more experienced and perceptive the individual, the more reliable the findings.

Prior to World War II only a handful of investigators recognized the need to rely on statistical tests to assess the effects of random variations on therapeutic outcomes. For the most part, chance was regarded as an enemy of knowledge rather than an ally, working its most powerful effects when the researcher was ignorant of "true" causes: no researcher could afford to overlook these effects but each researcher hoped to reduce them to a minimum. Experimental controls, in this context, gave researchers a technique for managing chance. In contrast, R.A. Fisher's ideas about randomized experiments called for researchers

---

<sup>6</sup> The 1931 study by J. Burns Amberson and his colleagues, previously cited by myself and others as an early example of controls chosen at random, is typical of contemporary thinking on the subject. The experimenters took twenty-four carefully selected tuberculous patients and divided them into twelve pairs, with the cases "individually matched, one with another" on the basis of the experimenters' clinical assessments. Group "A" was then assigned to receive a drug treatment; Group "B" to serve as controls "by a [single] flip of the coin". Such a procedure has little in common with the statistical theory underlying randomization, however much an advance it may have seemed on the standard experimental procedures of the day. See J. Burns Amberson, B.T. McMahon, and Max Pinner, "A Clinical Trial of Sanocrysin in Pulmonary Tuberculosis," American Review of Tuberculosis 24 (1931), 403-404.

not merely to acknowledge but to embrace chance: in the face of a sometimes perverse nature, the prudent investigator would give up trying to approximate certainty and concentrate on finding a means to measure the inevitable uncertainty which remained.

Fisher's initial interest was in plants, not patients. He was involved in agricultural experiments comparing the yields of different varieties of grain and was looking for a way to simultaneously solve two problems: first, to maximize the information gained from a single experiment, and second, to maximize the likelihood that a given experimental result could be relied on. Suppose, Fisher asked, an experimenter finds a 10% difference in yields between two grain varieties planted in different fields. How can one tell if the difference in yields is due to a real difference between the grains, or to differences in the soil, temperature, moisture and light in the two fields? One way is to rely on the experimenter's past experience that a difference of such magnitude is never due solely to variations in plot conditions; another is to replicate the experiment several times. Fisher rejected the first solution as unsound and the second as uneconomical. It could take as long as 500 hundred years, he calculated, to demonstrate that this particular finding was due to chance only once in twenty times.<sup>7</sup>

Instead, Fisher proposed that the experimental plots be divided into narrow strips, and that the grains be assigned to their place in the field by use of a chance mechanism. By subdividing a single field,

---

<sup>7</sup> R.A. Fisher, "The Arrangement of Field Experiments," [1926] in R.A. Fisher, Collected Papers (Adelaide: University of Adelaide Press, 1972), Vol 2, 86.

the number of observations gained from a single experiment is vastly increased, and the effects of variations in soil and atmospheric conditions on experimental error are greatly reduced.<sup>8</sup> For Fisher these were laudable but not crucial consequences of the proposed experimental procedure. The essential advantage for Fisher was only gained by the use of a chance mechanism (randomization) to assign treatments, which ensures the validity of the inference that the experimental difference in yields reflects a true difference in grain productivity. Where no chance mechanism is employed, what Fisher termed the experimental "estimate of error" is invalid.<sup>9</sup> The "physical act of randomization," Fisher wrote, "is necessary for the validity of [using] any test of significance."<sup>10</sup>

---

<sup>8</sup> Fisher, "The Arrangement of Field Experiments," (n. 7), 86-87. The practice of subdividing a field into narrow strips was already common in agricultural experimentation; Fisher's contribution was to provide an approach to the practice within the framework of statistical inference. See Joan Fisher Box, "R.A. Fisher and the Design of Experiments, 1922-1926," The American Statistician 34 (1980), 1-7; W.G. Cochran, "Early Development of Techniques in the Comparative Experimentation," in D.B. Owen, ed. On the History of Statistics and Probability. (New York: Marcel Dekker, 1976), 3-25.

<sup>9</sup> Fisher, "The Arrangement of Field Experiments," (n. 7), 87-88; R.A. Fisher, The Design of Experiments (London: Oliver and Boyd, 1935), 46-49.

<sup>10</sup> Fisher, The Design of Experiments (n. 9), 51. Fisher's argument is that any experiment produces only an estimate of the real error (or difference), and that for this estimate to reflect the real error, it is necessary that causes of variation which do not influence the real error (difference) should not be allowed to influence the estimate, while making it equally certain that any causes which do affect the real error must equally affect the estimate. [46-47] The use of a chance mechanism to assign treatment determines "...whether this particular ingredient of error [location in the field] shall appear in our average with a positive or negative sign. Since each particular error has thus an equal and independent chance of being positive or negative, the error of our average will necessarily be distributed in a sampling distribution, centered at zero, which will be symmetrical in the sense that to each possible positive error there corresponds an equal negative error, which, as our procedure guarantees, will in fact occur with equal probability." [48].



Couched in terms of evaluating new crops and fertilizers, Fisher's arguments nonetheless provided a conceptual basis for the use of statistical methods in all forms of experimentation. Ultimately, Fisher's work would prove relevant to an expansive range of practical and theoretical problems in the experimental sciences.<sup>11</sup> But in the years immediately following publication of his views regarding experimental design, Fisher and his followers had sufficient difficulty in convincing agricultural experimenters to apply the methods and theories developed with regard to grains to studies in animal breeding and care.<sup>12</sup> For much of the 1930's and 1940's, only a small group of

---

<sup>11</sup> For accounts of the initial lag among statisticians in accepting Fisher's ideas, see Harold Hotelling, "The Impact of R.A. Fisher on Statistics," Journal of the American Statistical Association (March, 1951), 45-46 and F. Yates, "Sir Ronald Fisher and the Design of Experiments," Biometrics 20 (June, 1964), 316. Not the least of the controversies over Fisher's methods was the dispute with his friend and colleague, W.S. Gosset, of "Student's t-test" fame, over the virtues of randomized versus systematic designs. Gossett believed that one could reduce the "real error" (or difference) between yields by use of a sandwich design ABBA in which unlike grains were placed as close as possible to each other. Two issues emerged as key in the controversy: 1. Whether "it was better, scientifically, to have smaller real errors, when comparing different treatment means," even if one thereby lost some precision in the estimate of error; 2) what was one to do when the result of randomization was an extremely biased (unbalanced) allocation? For accounts of the debate, see Cochran, "Early Development of Techniques," (n. 8), 21-23; and E.S. Pearson, "'Student' as Statistician," in E.S. Pearson and M.G. Kendall, eds. Studies in the History of Statistics and Probability (London: Griffin, 1970), 360-364.

<sup>12</sup> As summarized by John Wishart, the technical objections to randomized experimentation with animals sound remarkably similar to the subsequent reservations of medical and biological researchers: "Thus, it was stated that animals were more variable than field plots, too variable, in fact, for small differences in growth to be detected. Further, if this variability was to be reduced, all animals in one experiment ought to be offspring of the same parents, and ought to be of the same age and weight at the start of the experiment." And so on: the instinct of agricultural researchers was to look for, and seek to control, the biological factors which accounted for variable outcomes. See John Wishart, "Statistical Treatment of Animal Experimentation," Journal of the Royal Statistical Society, 6 (1939) Supplement no 1., 1-12. See also Yates, "Sir Ronald Fisher and the Design of Experiments," (n. 11), 314.

agricultural and biological researchers were familiar with Fisher's work. Despite a steady stream of pilgrims to Fisher's experimental farm at Rothamstead, the visibility and import of Fisher's ideas for medical research remained minuscule.<sup>13</sup> Regardless of their intellectual power and ultimate influence, Fisher's ideas about randomization would deserve little more than a footnote here, were it not for the illuminating contrast they provide with the way in which the virtues of randomization were initially articulated to medical audiences.<sup>14</sup>

For Fisher, the paramount virtue of randomization was that it enabled the statistics to work: with randomization, you knew how to interpret an experimental finding, without it you were lost. An

---

<sup>13</sup> On the pilgrimages to study with Fisher, see Joan Fisher Box, R.A. Fisher, The Life of A Scientist (New York: John Wiley & Sons, 1978), passim; and W.J. Youden, "The Fisherian Revolution in Methods of Experimentation," Journal of the American Statistical Association 41 (March, 1951), 49-50. Unfortunately, neither Box nor Youden offers anything near a complete list of the individuals who came to study and work with Fisher, although Youden mentions the existence of such a list for the period 1934-1944. Beginning in 1931, and intermittently after that, Fisher reversed the flow by coming to the United States to lecture at various statistical centers, most notably the University of Iowa. Fisher's sojourns here were interrupted during wartime, a fact to which his biographer attributes some of the intellectual distance between Fisher and the younger generation of mathematical statisticians who came into their own here during World War II.

One of the most touching accounts of Fisher's openness towards strangers is Box's report (245) of her future husband's initial arrival at Rothamsted, cap in hand and sargent's stripes on sleeve, to ask for Fisher's help in solving a problem the U.S. Army had saddled him with, which involved the analysis of some esoteric and fractious distributions. Fisher spent an afternoon of precious time working out an approach to the problem with the young sargent.

<sup>14</sup> Donald Mainland suggests that some of the problem is due to Fisher himself, Fisher's influence coming initially through his book on Statistical Methods rather than the more sophisticated Design of Experiments. See Mainland, "The Use and Misuse of Statistics in Medical Publications," Clinical Pharmacology and Therapeutics 1 (July-August, 1960), 412-413.

important, but secondary advantage of randomized experiments, was that they limited the sources of objective bias, by which Fisher meant the unknown factors which might favor one outcome over another, without the investigator's realizing it.<sup>15</sup> Both features aided the experimenter to orient himself in what Fisher apparently regarded as forays against a capricious, if not malevolent nature.

According to his biographer, Fisher viewed experimentation as a form of gambling with a somewhat perverse devil:

To play this game with the greatest chance of success, the experimenter cannot afford to exclude the possibility of any possible arrangement of soil fertilities, and his best strategy is to equalize the chance that any treatment shall fall on any plot by determining it by chance himself. Then if all the plots with a particular treatment have higher yields, it may still be due to the devil's arrangement, but then and only then will the experimenter know how often his chance arrangement will coincide with the devil's.<sup>16</sup>

In introducing randomization into medicine, the emphasis was not so much on the perversities of nature as on the credulity of man. To medical researchers, randomization's greatest asset was to neutralize the investigator's beliefs about the value of novel therapies: "This principal of the elimination of personal bias is fundamental in all experiments but it is of particular importance in clinical research."<sup>17</sup> To physicians engaged in therapeutic experimentation, randomization

---

<sup>15</sup> The limitation of objective bias, on Fisher's argument, derives not so much from the use of randomization per se, as from the use of randomization in tandem with the multiple replications of the experimental comparison which occur in randomized designs. It is this internal replication, not randomization per se, which effects the control of unknown factors. See Fisher, Design of Experiments (n. 9), 68-71, 78-79.

<sup>16</sup> Box, "R.A. Fisher and the Design of Experiments," (n. 8), 3.

<sup>17</sup> D.D. Reid, "Statistics in Clinical Research," Annals of the New York Academy of Sciences 52 (March 10, 1950), 932.

appeared to offer a mechanism which, by limiting the investigator's role in the selection and assignment of patients, would augment the medical community's confidence in the therapeutic claims being made.

For many physicians, the writings of A. Bradford Hill provided their initial introduction to the methods and purposes of the clinical trial. As the architect of the Medical Research Council's (1948) study of streptomycin, Hill's success in getting British physicians to adopt the principles of controlled experimentation was admired and envied in the United States. Hill's work and writings served as an example through which researchers hoped to persuade domestic audiences of the proper approach to clinical evaluation: the randomized controlled trial (RCT).<sup>18</sup> And in certain crucial respects, the principal claims of Hill and his interpreters regarding randomization had little to do with Fisher's theory of statistical inference.

To some extent, the arguments of Hill and others resembled the traditional skepticism of therapeutic reformers: human disease is highly variable, making individual case reports inadequate for assessing the merits of new treatments; clinicians have a limited

---

<sup>18</sup> For biographical background on Hill, see the various essays gathered in Statistics in Medicine 1 (1982), especially that by Sir Harold Himsworth, "Bradford Hill and Statistics in Medicine," 301-303. For citations of Hill's work with the MRC, see E.K. Marshall and Margaret Merrell, "Clinical Therapeutic Trial of A New Drug," Bulletin of the Johns Hopkins Hospital 85 (1949), 228-229; D.D. Reid, "Statistics in Clinical Research," (n. 17) 932; Donald Mainland, "Statistics in Medical Research," Methods in Medical Research 6 (1954), 152. Hill's importance as a propagandist on these shores should not deceive historians into thinking that American ideas about the need for randomized clinical trials originated in 1948 with the MRC studies of streptomycin; rather Hill's example was a convenient one to point to when trying to convince domestic audiences of the fundamental importance of controlled clinical trials in modern medicine.

ability to determine the factors (other than treatment) which might affect the outcome of therapy, making control groups a necessary adjunct to therapeutic studies; and so on.<sup>19</sup> To medicine's scientific elite, these were familiar and plausible arguments. But the reservations of earlier generations regarding the individual clinician's capacity to critically assess new therapies now applied with equal force to the researchers themselves. To Fisher's concept of objective bias, advocates of randomized therapeutic experiments added the notion of subjective bias, the hopes of the experimenter that a new treatment might work. The most compelling reason for randomization in medicine, according to its advocates, was to regulate the effects of the investigator's therapeutic preferences on the conduct of experimentation. The use of randomization ensured:

...that neither our personal idiosyncrasies, consciously or consciously applied, nor our lack of judgement have entered into the construction of the two (or more) treatment groups and thus biased them in any way....<sup>20</sup>

The difficulty with the traditional selection of controls by experts, Hill argued, is that one could never be sure that the groups were truly comparable or that factors other than treatment had not affected the results.<sup>21</sup> Even trained investigators have prejudices which may operate "unconsciously," if not consciously, to undermine the comparabi-

---

<sup>19</sup> A. Bradford Hill, "The Clinical Trial," British Medical Bulletin 7 (1951), 279; idem., "Assessment of Therapeutic Trials," Transactions of the Medical Society of London 68 (1953), 129-131, 136; Marshall and Merrell, "Clinical Therapeutic Trial of A New Drug," (n. 18), 224.

<sup>20</sup> Hill, "Assessment of Therapeutic Trials," (n. 19), 132.

<sup>21</sup> Hill, "The Clinical Trial," (n. 19), 278-279.

lity of treatment groups. Efforts to correct for such preferences simply result in "reverse" biases which undermine the integrity of the study.<sup>22</sup> "The best way to avoid such bias is to assign cases by some technique which eliminates the possibility of prejudice."<sup>23</sup>

As statisticians acknowledged, randomization could not, in fact, guarantee that experimental groups were comparable.<sup>24</sup> However, when used in tandem with other recommended methodological reforms, such as "blinded" assessment of outcomes, randomization could free a researcher from the accusation that his beliefs had affected a study's execution. The researcher who followed correct procedure could reassure not only himself but his colleagues that the results could be trusted. By freeing the investigator from the charge of bias as well as the act, randomization ensured that even

...the sternest critic is unable to say eventually when we dash into print that quite likely the groups were biased through our

<sup>22</sup> A. Bradford Hill, Principles of Medical Statistics. 3rd Edition. London: The Lancet, Ltd., 1945), 8; idem., "The Clinical Trial," 280; idem., "Assessment of Therapeutic Trials," (n. 19), 132. See also Mainland, "Statistics in Medical Research," (n. 18), 155.

<sup>23</sup> Louis Lasagna, "The Controlled Clinical Trial: Theory and Practice," Journal of Chronic Diseases 1 (April, 1955), 357.

<sup>24</sup> While Hill, and other statistically trained authors were at pains to insist on the need to check the comparability of the experimental groups after randomization, it was not difficult for the unmethodical and less informed reader to overlook such caveats, arriving at the conclusion that randomization took care of the problem. The occasional remarks by Hill and others that randomization led to equalization of the groups "in the long run" no doubt contributed to the problem. See A.B. Hill, "Principles of Medical Statistics: I. The Aims of the Statistical Method," Lancet i (January 2, 1937), 42; Brian MacMahon, "Statistical Methods in Medicine," New England Journal of Medicine 253 (October 13, 1955), 648. For caveats, see Hill, "The Clinical Trial," (n. 19), 281-2; idem., "The Assessment of Therapeutic Trials," (n. 19), 132; Marshall and Merrell, "Clinical Therapeutic Trial of A New Drug," (n. 18), 225; Reid, "Statistics in Clinical Research," (n. 17), 932.

predilections or through our stupidity. The random method removes all responsibility from the observer.<sup>25</sup>

The naivete of such statements would discomfort subsequent generations of statisticians.<sup>26</sup> Yet behind such imprecise exaggeration lay a desire to impress researchers with the critical contribution which statistical methods and reasoning might make to therapeutic experimentation.<sup>27</sup> Where R.A. Fisher had sought to give the experimenter a measure of the uncertainty which characterized his results, randomization was offered to medical researchers as a technique for bolstering confidence in their experimental findings.<sup>28</sup> An earlier

<sup>25</sup> Hill, "Assessment of Therapeutic Trials," (n. 19), 132.

<sup>26</sup> Jerome Cornfield, "Recent Methodological Contributions to Clinical Trials," American Journal of Epidemiology 104 (1976), 408-421; Frederick Mosteller, personal communication.

<sup>27</sup> It is interesting to note that Fisher's argument on behalf of randomization does not appear in any of A. Bradford Hill's most-oft cited papers. Perhaps Hill's subsequent remark to a colleague and friend explains his reluctance to enter into the complex Fisherian rationale for randomization: "My skill, if any was, I believe, in offering the clinician something simple that he could understand....If one had started with something abstruse, the answer would have been 'Go to Hell'—and we would still be there." Cited in Harold M. Schoolman, "The Clinician and Statistician," Statistics in Medicine 1 (1982), 315.

<sup>28</sup> Individuals who had studied with Fisher, or who followed the logic of his arguments in the Design of Experiments acknowledged the importance of randomization in providing "a basis for valid inference," [Mainland]. Despite the care which individuals like Mainland took, "Many people believe they [statistical methods] eliminate chance when in fact they merely give us an idea as to the probability of the results being due to chance." Cornell Conference, "How to Evaluate A New Drug," American Journal of Medicine 17 (November, 1954), 727. That physicians ended up thinking this way is not surprising, since even sophisticated authors at times put the case for randomization in a curious way: "The differences observed .. between treated and control groups may be due to chance, and it is the function of the statistical test of significance to test just that hypothesis. [....] Chance is always considered to be guilty or responsible for the differences until its innocence has been proved by the results of technical tests of significance." [my emphasis] Reid, "Statistics in Clinical Research," (n. 17), 933.

generation's trust in the judgement of experienced researchers was to be replaced by a reliance on experimental method: "The use of properly designed clinical trials permits us to move from an authoritative frame of reference to a scientific one."<sup>29</sup>

### 'The Gift Relationship': Statisticians and Clinicians

Once the clinician has grasped the simple techniques that have been brought to his aid, the statistician has no further part to play. Along with the old soldier he can fade away, contentedly if, sometimes, wistfully.<sup>30</sup>

It is currently fashionable in some circles to consider the clinician member of the team as some sort of minor excrescence, 'a fifth cousin about to be removed.'<sup>31</sup>

Prior to the late 1940's, the statisticians involved in clinical research played an ancillary and subordinate role to their physician colleagues. Apart from a few individuals who took the lead in conceptualizing and interpreting bioassays, statisticians generally served as collectors and computers of data, or as occasional critics of the research enterprise. The statistician, like the pathologist, was consulted only after the damage was done, or not at all. Yet by the mid-1960's, no clinical researcher embarking on a major therapeutic experiment would think of planning such a study without the active colla-

---

<sup>29</sup> J.J. Nickson and A.S. Glicksman, "Clinical Trials in Radiation Therapy: An Application of Statistical Methods to Clinical Research," American Journal of Roentgenology 96 (July, 1966), 236.

<sup>30</sup> A. Bradford Hill, "Reflections on the Controlled Trial," Annals of Rheumatic Diseases 25 (1966), 113.

<sup>31</sup> Lasagna, "The Controlled Clinical Trial," (n. 23), 354.



boration of one or more professional statisticians, a curious outcome for a group whose stated purpose was "professional suicide."<sup>32</sup> The circumstances which led to such results bear further examination.<sup>33</sup>

To the modern sociologist of science, on the lookout for Hobbesian forays in discipline building and professional self-aggrandizement, the behavior of medical statisticians would seem self-defeating. Rather than emphasize the esoteric aspects of their technical craft, they disparaged the importance of "statistical arithmetic"; rather than lay claim to unique theoretical insights, they insisted that "modern statistics does not claim to be something intrinsically different from the principles and methods of experimenters in general"; rather than recording each minor assist to members of the research community, they accepted credit in the literature only for contributions "above and beyond the ordinary call of duty."<sup>34</sup>

<sup>32</sup> Hill, "Reflections on the Controlled Trial," (n. 30), 113.

<sup>33</sup> Of the potential means for examining the role of the statistician, the most promising avenue—following the path from the logic of the experiment to the practice of contemporary therapeutic experimentation—is at present severely obstructed. Few of the individuals engaged in the building of this research enterprise have yet retired, and of those who have, the correspondence and other documentary records of their activities are markedly lacking either in information about the planning and conduct of specific studies or the discussions which statisticians engaged in, among themselves and with their physician collaborators, about the application of statistical theory to clinical experiments. In some situations, where face to face contact eliminated much need for correspondence, little such documentation is likely to surface. This may be the case for the small group of statisticians at NIH (Cornfield, Mantel, Greenhouse, Moore).

<sup>34</sup> See Mainland, "Statistics in Medical Research," (n. 18), 122; *idem.* "The Use and Misuse of Statistics," (n. 14); and Reid, "Statistics in Clinical Research," (n. 17) 934. In assisting experimenters, statisticians initially saw themselves as carrying out the logical implications of a developed theory; the idea that some of their contributions might in fact be of scientific interest, and hence publishable only gradually emerged. See Nathan Mantel, "A Personal Perspective on Statistical Techniques for Quasi-Experiments," in Owen, On the History of Statistics and Probability (n. 8), 124-125.

Statisticians' insistence that the planning of clinical trials was more "a matter of hard work and attention to detail" than an application of "esoteric intellectual principles" was hardly insincere. Their initial counsels to clinical investigators were little more than "common sense talk" with few pretensions to theoretical originality.<sup>35</sup> Although not the product of a self-conscious strategy to accumulate influence, the conduct of statisticians could not have been better calculated to do so. In emphasizing the need to raise the standards of therapeutic experimentation, statisticians created a natural alliance with reformers who sought to improve the practice of medicine, and thereby created an opportunity for themselves to contribute to the cause of reform.

There was nothing novel about the desire of therapeutic reformers to adjudicate the profession's adoption of new treatments. But in the post-war era, reformers faced a situation different in quantity, if not quality, from previous decades. The flourishing production of new remedies required that the doctor abandon the time-consuming process of

---

<sup>35</sup> W.G. Cochran, "Designing Clinical Trials," in Francis M. Forster, The Evaluation of Drug Therapy (Madison: The University of Wisconsin Press, 1961), 71; See also Cochran's letter to William G. Madow, December 6, 1961, characterizing the existing literature as "common sense talk with precautions about the [?] sources of bias." W.G. Cochran papers. Box 5. Harvard University Archives. The realization that clinical trials provided an opportunity for statisticians to make original contributions to statistical theory and techniques came only gradually, as some of the more senior statisticians began to realize the inadequacies of traditional theories of experimental design and inference for certain clinical problems, and as some of the younger statisticians became aware of the importance of publication.

"weighing imponderables" and slowly "coming to a decision."<sup>36</sup> Traditional means for discriminating among the relative merits of "undoubtedly effective drugs" were seen as inefficient as well as unreliable: "For instance, I do not think the relative values of aureomycin, chloramphenicol and penicillin in clinical pneumonias could have been determined without a fairly large-scale and statistically designed trial."<sup>37</sup> Moreover, the enthusiasm physicians formerly expressed for remedies of questionable value was of even greater concern in the case of newer, more powerful drugs coming on the market.<sup>38</sup> The potential of "properly designed" trials to curb the enthusiasm of physicians for novel treatments did not go unnoticed.<sup>39</sup> "Modern methods of experimental design" offered a means to avoid the "needless suffering of animals and human

<sup>36</sup> L.J. Witts, "Introduction," in L.J. Witts, ed. Medical Surveys and Clinical Trials. Some Methods and Applications of Group Research in Medicine. (London: Oxford University Press, 1959), 3. See also Lasagna, "The Controlled Clinical Trial: Theory and Practice," (n. 23), 353; Cornell Conference, "How To Evaluate A New Drug," (n. 28), 722; J. Solon Mordell and C.K. Himmelsbach, "An Objective Approach to New Drug Therapy," Public Health Reports 68 (January, 1953), 47.

<sup>37</sup> Hill, "Assessment of Therapeutic Trials," (n. 19) 136.

<sup>38</sup> See James Whorton, "'Antibiotic Abandon': The Resurgence of Therapeutic Rationalism," in John Parascandola, ed. The History of Antibiotics: A Symposium (Madison: American Institute of the History of Pharmacy, 1980), 125-136. My only quarrel with Whorton's excellent discussion of the elite's critical reaction to antibiotics is the tacit implication that between Osler and the antibiotic era, therapeutic rationalism was in suspended animation.

<sup>39</sup> Hill, Principles of Medical Statistics, (n.22) 171; E.M. Glaser, "Volunteers, Controls, Placebos and Questionnaires in Clinical Trials," in Witts, Medical Surveys and Clinical Trials, (n. 36) 111.

beings [which] occurs when an investigation is so conducted that any conclusion that can be drawn from it must inevitably be equivocal."<sup>40</sup>

From the reformers' perspective, unnecessary equivocation about the merits of new treatments was as undesirable as unwarranted enthusiasm. Their aim was not to prevent progress but to provide a reliable basis for distinguishing it from puffery:

The doctor of today is under constant bombardment with claims as to the efficacy of drugs, new and old. It is difficult, if not impossible, to read a journal, attend a medical meeting, or open the morning mail without encountering a new report on the success or failure of some medication. The clinician who would avoid nihilistic rejection or trusting acceptance of all such claims, or capricious decisions as to their merits, is well advised to adopt a yardstick, a set of criteria, that will improve his chances of making sound evaluations.<sup>41</sup>

The techniques of modern statistical experimentation offered just such a yardstick: a well designed trial should offer clinicians "as decisive an answer" to questions about of therapeutic merit as "can be foreseen or as the statistical approach can ever give."<sup>42</sup>

<sup>40</sup> Donald Mainland, "The Planning of Investigations," in Methods of Research in Medicine 6 (1954), 138.

<sup>41</sup> Lasagna, "The Controlled Clinical Trial: Theory and Practice," (n. 23) 353. For a similar argument, see Mindel C. Sheps, "The Clinical Value of Drugs: Sources of Evidence," American Journal of Public Health 51 (May, 1961), 650-652. Sheps' account is unusual for its emphasis on the need to have hospitals and "organized medical groups" support such evaluations, by participating in such studies and by creating institutional policies based on such evaluations (653). Sheps is also unusually frank about the propensity of drug firms to put clinical trials to "serve promotional [rather than scientific] ends".

<sup>42</sup> A. Bradford Hill, The Philosophy of the Clinical Trial, National Institute of Health Annual Lectures--1953. The British Medical Research Council's "Therapeutic Trials Committee" with which Bradford Hill worked was formed in 1931 in response to industry requests for "an authoritative body to arrange clinical tests of [promising] new remedies." F.H.K Green, "The Clinical Evaluation of Remedies," Lancet ii (November 27, 1954), 1087-1090. Despite changes in organization and name, it continued to function as an adjudicator of

In allying themselves with reformers, statisticians eschewed therapeutic nihilism.<sup>43</sup> The RCT's function, accordingly, was to be "as informative and as convincing as possible."<sup>44</sup> To convince, a clinical study not only needed to offer guarantees of impartiality, but to be as simple and as transparent as possible. A trustworthy answer to a simply put question, statisticians argued, was preferred to a contestable reply to a more complex inquiry.<sup>45</sup> The key to a simple investigation was planning, necessary not only to determine which questions were to be asked but to decide if they were worth asking

Proper planning does not guarantee a successful experiment, but it makes success more likely; and attempts to draw up a plan are equally valuable if they reveal that a proposed investigation would be futile.<sup>46</sup>

As much as any methodological innovations, what distinguished the "modern" clinical trial was the participation of statisticians in its

therapeutic claims. The American Medical Association's Council on Pharmacy and Chemistry, in emulation of the MRC, formed its Therapeutic Trials Committee in 1946. While the activities of the American TTC were less extensive than those of the MRC, it aspired to a similar role. See Council on Pharmacy and Chemistry, "The Therapeutic Trials Committee," JAMA 131 (June 15, 1946), 596-597; Austin Smith, "Some Research Interests of the American Medical Association," Food, Drug, Cosmetic Law Quarterly 3 (June, 1948), 220-223.

<sup>43</sup> Hill, "The Clinical Trial," (n. 19) 282.

<sup>44</sup> A. Bradford Hill, "Aims and Ethics," Conference on Controlled Clinical Trials-Vienna, 1959 (Oxford: Blackwell Scientific Publishers, 1960), 4.

<sup>45</sup> Hill, "The Clinical Trial," (n. 19) 279; Marvin Schneiderman, "Controlled Clinical Trials: Monday's Countdown for Tuesday's Launching," Journal of New Drugs 1 (November-December, 1961), 251.

<sup>46</sup> A group of investigators who could not agree on the kinds of patients to be enrolled, the appropriate schedule of treatment and the means for measuring improvements, probably were "not ready" to undertake a trial. See Cochran, "Designing Clinical Trials," (n. 35), 71. On planning see Mainland, "The Planning of Investigations," (n. 40), 138-145.

planning. The investigator was advised to "consider your statistical colleague rather as an architect, to be consulted before the work is started, so that" the experiment can produce "the maximum amount of accurate information."<sup>47</sup> In improving the quality of therapeutic experimentation, it was not the statistician's technical contributions—his repertoire of experimental designs or his ability to calculate the necessary sample size and choose the appropriate statistical tests— which mattered so much as his intellectual gifts for "ferreting out weaknesses in experiment design, risks of bias, and undesirable variability."<sup>48</sup> As statisticians were quick to acknowledge, they had no particular monopoly on such "statistical tact," which was as much a matter of temperament as of training. But if professional education provided "no guarantees" of such aptitudes, the individuals drawn to statistics were more likely to possess that affinity for obsessive doubt which in their view made for good experiments. And if neither nature or nurture engendered a taste for the "policeman's duties" in statisticians, circumstances did: detecting (and repairing) problems in the design and conduct of experiments required time, and the statistician

---

<sup>47</sup> Reid, "Statistics in Clinical Research," (n. 17) 931. On the importance of having the statistical contribution made while the study is being planned, see also Donald Mainland, "Statistics in Clinical Research," Annals of the New York Academy of Sciences 52 (March 10, 1950), 923; Lasagna, "The Controlled Clinical Trial: Theory and Practice," (n. 23) 356.

<sup>48</sup> Donald Mainland, "The Clinical Trial: Difficulties and Suggestions," Journal of Chronic Disease 11 (May, 1960), 493.

had more such time (if never enough) than the busy clinicians with whom they worked.<sup>49</sup>

The role of house skeptic was an uncomfortable one, which statisticians sought, unsuccessfully, to abdicate.<sup>50</sup> In the short run, their place in medical schools depended on their ability to help, not hinder, the experimenter's progress.<sup>51</sup> If it often fell to the statistician to remind the investigator "too readily persuaded by his data...just how significant or insignificant" the results were, statisticians also elected to "concentrate on what has been accomplished positively" thereby allaying the reservations of young investigators about the merits of their contributions and assisting them in getting products out the

<sup>49</sup> The notion of statistical tact, "which is rather more than simple good sense" is Major Greenwood's, cited by Bradford Hill: "some are born with it; the rest of us have to acquire it." "Statistics in Medicine," Manchester Statistical Society, Transactions (1946-1947), 4. The talent to which Hill refers is a temperamental ability "to ask oneself such questions as is this 'control' an efficient control, is that difference explicable on any other grounds than the obvious grounds, is this comparison a just comparison for the question at issue, what, here, were the 'exposed to risk?'" (4-5). On the affinity of statisticians for their duties see Mainland, "The Clinical Trial: Difficulties and Suggestions," (n. 48) 492-495. As Mainland observed, two "experienced" statisticians could be had for the price of one "suitable" clinician, another circumstance which favored the division of labor.

<sup>50</sup> On the reluctance of statisticians to assume the pre-eminent role in managing clinical trials, see J. Yerushalamy, "The Planning of a Clinical Trial--Introduction," in Forster, Evaluation of Drug Therapy (n. 35), 60.

<sup>51</sup> Donald Mainland, "We Wish to Hire A Medical Statistician. Have You Any Advice to Offer?," JAMA 193 (July 26, 1965), 290-291. Mainland's article, largely a reprint of a paper composed in 1958 but rejected for publication then, is largely about the difficulties which medical school statisticians will face in refusing help to colleagues in the face of demands that they serve as methodological fireman, offering consultations by the bushload. Mainland's point is that the quality of advice, and hence of the resulting research, depends on the statistician's opportunity to immerse himself in the experimenter's problem to a greater extent than circumstances generally allowed.

door.<sup>52</sup> In the long run, however, the success of their program for experimental reform depended on the "investigator himself" mastering "the principles of statistical reasoning." Statisticians expected that clinically trained investigators would ultimately take responsibility for routine therapeutic investigation, as understanding of the relevant statistical principles became more commonplace.<sup>53</sup>

Effective statistical education was propaganda by the deed, requiring the active collaboration of statisticians and clinicians in identifying and solving the problems of therapeutic experimentation.<sup>54</sup>

The most successful statisticians were those with gifts for persuasion as well as polemic, those who learned to speak the clinician's language well and to present complicated statistical ideas as if they were little more than common sense.<sup>55</sup> After a few "intimate collaboration[s]", the qualified medical investigator might obtain enough "insight into the methods of applying the general principles" of statistics to enable him to work on his own, limiting future consults to particular technical questions or instances where more abstruse methodological issues came into play.<sup>56</sup>

<sup>52</sup> Mantel, "A Personal Perspective," (n. 34) 120.

<sup>53</sup> Mainland, "Statistics in Clinical Research," (n. 47) 929; Hill, "Statistics in Medicine," (n. 49) 3-5

<sup>54</sup> Mainland, "Statistics in Clinical Research," (n. 47), 929; and "Statistics in Medical Research," (n. 18), 124-5.

<sup>55</sup> See the remarks of Fred Ederer on Jerome Cornfield and of Sir Harold Himsworth on Bradford Hill: Fred Ederer, "Jerome Cornfield's Contributions to the Conduct of Clinical Trials," *Biometrics* (March 1982), Supplement, 25-26; Himsworth, "Bradford Hill and Statistics in Medicine," (n. 18), 302-303.

<sup>56</sup> Mainland, "Statistics in Medical Research," (n. 47), 125.



The supply of "suitably trained" statisticians, however, could provide only a fraction "of the guidance that is needed." To investigators unable to undertake a suitable apprenticeship, statisticians could only offer the admittedly second best prescription of "a limited number of simple techniques of design and analysis, which should be rigidly adhered to."<sup>57</sup> The articles and then books on "modern" methods of therapeutic investigation which began to appear in the 1950's were part of this missionary effort. But the desire to inculcate the philosophy, and not merely the techniques, of statistics among physicians could not be fulfilled by the writing of books and the organizing of symposia. To the clinician not personally exposed to statistical reasoning through contact with the small group of biostatisticians interested in medical research, the statistician's dictates could seem unnecessarily ritualistic and even inimical to the progress of clinical investigation.<sup>58</sup>

#### Ideological Skirmishes: Ethics and Epistemology

To judge from the published writings of medical statisticians in the 1950's, the road to responsible therapeutic investigation was paved over the objections of clinicians, which had to be dismantled before RCTs could become an acceptable and routine practice. Whatever reser-

---

<sup>57</sup> Mainland, "Statistics in Medical Research," (n. 47), 125.

<sup>58</sup> The impression of ritualism was no doubt abetted by the fact that after the mid 1950's, few individuals thought it worthwhile to present the various arguments for RCTs ab novo, and the importance of randomization was as frequently presumed as explained. That even before this date a very small group of authorities mostly quoted each other might also have given a few independent thinkers cause for reflection.

vations physicians held privately about the enterprise, however, few gave voice to them publicly. Their case must be presented largely through the writings of those who sought to rebut it. For Bradford Hill, the problem began with the physician's personality:

The person who is attracted to medicine may be one who is interested in individuals, individuals who are sick and in need of expert knowledge and attention. Interest in the group, the statistical concept, is, therefore, not present or is weak.<sup>59</sup>

Training and professional experience only reinforce the physician's habit of thinking about individuals.<sup>60</sup>

To critics of the statistical enterprise, by emphasizing the study of groups, the statistician at best discarded (valuable) information on "individual variability" and at worst, discouraged efforts to investigate why one patient responded to treatment and another did not:

When the statistician speaks of the design of an experiment, he or she, means the design of a trial to ensure that it provides the requisite data for a test which (if valid) will justify or discredit the assertion that a treatment guarantees an average level of benefit for a representative group of patients; but ... it is not our ultimate goal to state that treatment A is or is not better than treatment B in this sense, though such a statement may well be a useful and unavoidable preliminary. We are in search of a one-to-one correspondence [between patient and treatment], and any assessment of remedies stated within the linguistic framework of averages prescribed by the requirements of the statistician is at best a provisional answer to the question which medical ethic and biological curiosity alike prompt us to explore in the context of the clinical trial.<sup>61</sup>

To statisticians, such reasoning was fallacious: physicians do, unavoidably, deal with groups. They compare individual patients

<sup>59</sup> Hill, "Statistics in Medicine," (n. 49), 2-3.

<sup>60</sup> Hill, "Statistics in Medicine," (n. 49), 3.

<sup>61</sup> Lancelot Hogben, "The Assessment of Remedies," The Medical Press (October 13, 1954), 353.

and the results of treatment in groups of patients. Consider, Bradford Hill suggested, the case of the "careful" clinician:

Faced with rapid recovery of a single case of rheumatic fever under a new drug what, I suggest, the careful clinician would do is not to generalize but to test the treatment on a second case. If it again worked well he would test it on a third. And without being accused of undue caution, or even of mathematical leanings, he might go so far as to seek a fourth. And so with, perhaps, somewhat halting steps he unwittingly directs himself up the statistical garden path. I believe he might sometimes fare better if he straightaway walked boldly up the path and without any ado opened the gate to a designed and clinical trial.<sup>62</sup>

If the clinician was reasonably confident that certain groups of patients responded to treatment differently than others, well and good: the statistician could accommodate that supposition in the experimental design.<sup>63</sup> But in most cases, statisticians averred, the clinician's supposition would not withstand the critical scrutiny of his colleagues, much less that of an experimental test. The statistician's insistence on randomization accordingly represented a not too subtle reminder of the limitations of medical knowledge:

It is impossible to identify all the elements that affect the outcome and therefore we require some process that will bring about similarity in our groups without attempting to categorize all the important factors and match the groups on these.<sup>64</sup>

The physician who persisted in claiming to effectively discriminate between those who would, and would not, benefit from therapy, was only deluding himself (and his patients). But the researcher who refused to subject those beliefs to a controlled experimental test was

<sup>62</sup> Hill, "Assessment of Therapeutic Trials," (n. 19) 131.

<sup>63</sup> Hill, "Assessment of Therapeutic Trials," (n. 19) 132-133.

<sup>64</sup> Marshall and Merrell, "Clinical Therapeutic Trial of a New Drug," (n. 18), 224.

jeopardizing the community's chance to ever determine the truth. Once it was claimed that a new therapy worked, it became increasingly difficult to deny that treatment to other patients, even for the sake of testing the claim. Trials should be begun, statisticians argued, before less conclusive evidence establishes a prejudice on behalf of a novel treatment. The logic was plain: to apply the practice of randomized experimentation at all, it ought to be universal.<sup>65</sup>

To members of the research community, claims for the application of controlled experimentation to new treatments were most compelling when contrasted with uncontrolled testing by individual clinicians: "Deliberate experimentation on a group of cases with adequate controls is merely an efficient and convenient means of collecting and interpreting data that would otherwise be dispersed and inaccessible."<sup>66</sup> When contrasted instead with other strategies of scientific investigation, controlled trials, with their emphasis on the clinical outcomes of treatment, faced an uphill fight for intellectual respectability and access to resources.

---

<sup>65</sup> For the statistician's insistence on the need to conduct well-controlled trials before the credibility of new therapies was established on the basis of inferior evidence, see Marshall and Merrell, "Clinical Therapeutic Trial of A New Drug," (n. 18) 230; and Hill, "The Clinical Trial," (n. 19) 279. Concern about other studies foreclosing the opportunity to do a "strictly controlled clinical trial" of cortisone and ACTH in treating rheumatic fever prompted the organization of an international study of these drugs. Rheumatic Fever Working Party [MRC] and American Council on Rheumatic Fever and Congenital Heart Disease, "The Treatment of Acute Rheumatic Fever in Children. A Cooperative Clinical Trial of ACTH, Cortisone and Aspirin," Circulation 11 (March, 1955), 343-371.

<sup>66</sup> Michael B. Shimkin, "Problem of Experimentation of Human Beings. I. The Research Worker's Point of View," Science 117 (February 27, 1953), 205.

Few individuals publicly opposed large scale cooperative therapeutic studies in principle. The problem was presented instead as one of competing claims on a limited pool of research support, to be adjudicated "on the merits" in each instance.<sup>67</sup> For those who favored clinical trials, the path of least resistance was to augment their scientific yield by collecting data on various physiological, biochemical and clinical parameters of disease: "There seems little Nobel about a clinical trial. If you complicate it with some biochemistry, a little epidemiology, and so forth, perhaps it will seem more worth doing."<sup>68</sup> Such data, while often incidental to the study's principal aim of determining therapeutic benefit, satisfied researchers interested in adding to the store of knowledge about the biology and natural history of disease. But to methodological reformers intent on promoting studies

---

<sup>67</sup> The cautious remarks of James Shannon, Director of the National Institutes of Health, illustrate the stance taken toward "applied" research of all sorts, including clinical trials: "I do not believe it is reasonable to question whether we should do this type of work in general. Rather the critical question is a highly specified one: in each area, is the knowledge adequate at the moment to undertake such a study, or would it be more profitable to delay? The answer to this question is most important since such studies are always costly, always profligate with scientific talent, and—if embarked upon with insufficient knowledge—may well be predisposed to failure." See James A. Shannon, "Factors Influencing the Substance and Dimension of Medical Research in the United States," in G.E.W. Wolstenholme, Cecilia O'Connor and Maeve O'Connor, eds. Symposium on Significant Trends in Medical Research (Boston: Little Brown & Co, 1959) 313-314. Private research organizations faced similar problems, pace the evident relief of the American Heart Association that the Public Health Service (NIH) had agreed to finance Irving Wright's proposed clinical test of dicumerol. See American Heart Association, Board of Directors, Minutes, June 28, 1946. Howard B. Sprague papers, Francis Countway Library of Medicine. Harvard Medical School. [The Sprague papers are currently uncatalogued.]

<sup>68</sup> To the researcher so tempted, the statistician's advice was "Don't! Resist the temptation! Keep it simple!" See Schneiderman, "Controlled Clinical Trials, (n. 45), 251.

with fewer objectives and simpler protocols, even the most noteworthy clinical trials of the 1950's represented compromises with their experimental ideals.<sup>69</sup>

Like their predecessors, who thought to dictate therapeutic practice from the confines of the laboratory, postwar reformers were not particularly troubled by the slowness with which medical researchers adopted recommended improvements in therapeutic experimentation. Fields which had not yet recognized the contribution of statistical methods to therapeutic research simply represented new territory to conquer. As the practices advocated by reformers became more widespread, it became easier to admonish the outlying provinces with their failure to keep step with scientific developments elsewhere

---

<sup>69</sup> The multi-year study of anticoagulants in the prevention of coronary thrombosis, sponsored by the American Heart Association and the National Institutes of Health, provides an extreme example of the genre, with its 656 pages of text and dozens of tables providing data about the natural history of coronary thrombosis. See American Heart Association. Committee on Anticoagulants. Myocardial Infarction. Its Clinical Manifestations and Treatments with Anticoagulants (New York: American Heart Association, 1954).

The 1956 trials of the Salk polio vaccine, perhaps the largest of postwar "therapeutic" experiments, reflect the interests of virologists in collecting massive amounts of data on antibody formation within the vaccinated and control population. See Thomas Francis, Jr. "An Evaluation of the 1954 Poliomyelitis Vaccine Trials--Summary Report" American Journal of Public Health 45 (1955), 26-32, The polio investigators were quite fortunate that there were no obvious points of contradiction between the immunological and the clinical evidence, and that the vaccine seemed effective regardless of whether clinical or laboratory confirmed diagnoses were taken as a measure of outcome.

in medicine.<sup>70</sup> But underlying the reformers' portrait of uneven but inevitable progress lay a more complex intellectual and social process.

However much research physicians might agree in theory about the ability of randomized controls to improve the reliability of therapeutic knowledge, in practice the experienced clinician often claimed to know more about the factors which influenced therapeutic outcomes than those who had planned the latest authoritative clinical trial. In the case of many acute diseases, where small amounts of money or time would suffice to conduct a study, the physician with different assumptions about the patients who would best benefit from treatment or the dosages which would produce the best results could conceivably follow the statistician's advice to "conduct another controlled experiment." In instances where the outcomes of treatment were unambiguous, and academic physicians had well established interests in determining the relative merits of specific drugs, the counsels of methodological reformers increasingly took hold. Simplifying experimental objectives, adopting the use of randomized controls, and incorporating statistical tests were acceptable, if hardly universal, measures of methodological improvement. The incorporation of statistical methods in these

---

<sup>70</sup> See, for example, Erik Juhl, Erik Christensen and Niels Tygstrup, "The Epidemiology of the Gastrointestinal Randomized Clinical Trial," New England Journal of Medicine 296 (January 6, 1977), 20-23; J.J. Nickson and A.S. Glicksman, "Clinical Trials in Radiation Therapy-An Application of Statistical Methods to Clinical Research," American Journal of Roentgenology 96 (1966), 236-240.

domains accordingly enhanced, rather than subtracted, from the authority of academic physicians to direct therapeutic practice.<sup>71</sup>

In the treatment of chronic diseases, where controversies over the merits of particular therapies ran deepest, the promise of improvements in experimental method to adjudicate differences of opinion about clinical and scientific questions was harder to realize. Here the strategy of collecting more and more data in the course of a study ran the risk of producing more, not less, controversy, as physicians attached different interpretations to the therapeutic findings on the basis of the reported information.<sup>72</sup> And the obvious methodological solution to scientific disputes--to conduct a better study--was hardly a routine option in circumstances where hundreds of patients and years of follow-up might be needed to complete an experiment.

While acknowledging that the study of chronic disease might pose greater logistical and organizational difficulties, statisticians had no reason to think that the biological complexity of chronic disease posed more fundamental challenges to the program of methodological reforms. Yet the case of chronic diseases brought to the fore a

---

<sup>71</sup> The support of drug companies for improvements in the methods of clinical evaluation assisted in these developments. See Hart E. van Riper and Donald Boyer, "The Planning and Reporting of Clinical Trials," New York State Journal of Medicine (October 1, 1961), 3337-3342. Progress was nonetheless slow, when viewed from the perspective of therapeutic reformers.

<sup>72</sup> The opposition of George Burch, a participant in Irving Wright's comprehensive evaluation of anti-coagulants, to granting Wright's study the official endorsement of the American Heart Association speaks directly to the continuing limitations of controlled trials in setting authoritative practice standards. See Executive Committee of the Scientific Council, American Heart Association, Minutes, March 13, 1954. Sprague papers.



critical ambiguity in the framework of statistical inference: what was the role of experience in interpreting experimental data? For many statisticians, prior experience had no legitimate place in the formal analysis of experimental data. The scientist's principal contribution to an experiment came in the planning of a study, not its interpretation. Few statisticians were naive enough to believe that the interpretation of therapeutic experiments was purely an exercise in mathematics. In practice, an experiment whose results conflicted with experience or biochemical theory might deserve closer scrutiny. But statistical theory offered no clues to the researcher interested in incorporating his own beliefs about the pathogenesis of disease or the mechanisms by which a drug acted in interpreting a given experimental result.<sup>73</sup>

The inability of statisticians to offer a formal methodology for incorporating "outside" knowledge into the interpretation of an experiment might have remained a purely academic issue, of interest primarily to those concerned with the logic of statistical inference. Instead, it took on wider significance, as statisticians sought to extend their accomplishments to the sphere of chronic disease. To a extent greater here than elsewhere in medicine, the scientific authority of specialists in chronic disease rested on their claim to formulate complex, unquantifiable judgments about the myriad factors which

---

<sup>73</sup> At best, statisticians offered an informal division of labor between the scientist, concerned with the problem of scientific truth, and the statistician, concerned with the problem of experimental precision. See George W. Snedecor, "The Statistical Part of the Scientific Method," Annals of the New York Academy of Medicine 52 (March 10, 1950), 792-799.

determined why one patient responded to treatment and another did not. The inability of statisticians to specify a formal place for the judgments of experts in the evaluation of clinical trials directly challenged the authority of such clinical experts. The insistence of methodological reformers that statistical methods were superior to individual judgment in evaluating the effects of treatment added insult to injury. And the underlying uncertainty and ignorance about the nature of these diseases made it possible for clinical trials of their treatment to become quite controversial. In the next chapter, I examine the controversy over one such study, the University Group Diabetes Program evaluation of treatments for chronic diabetes.

Chapter Six. Anatomy of a Controversy:  
The University Group Diabetes Program Study

I. Introduction

On May 21, 1970, some 800,000 diabetics were greeted with the news that tolbutamide (Orinase), a drug widely used in the treatment of milder diabetes, was causing "early death" in a substantial number of individuals. Investigators in the "biggest, most sophisticated and probably the longest study of diabetes ever made" had found that patients taking tolbutamide experienced significantly higher mortality than patients receiving a placebo.<sup>1</sup> On the basis of these findings, the U.S. Food and Drug Administration (FDA) was reviewing indications for the use of tolbutamide.<sup>2</sup> For all but a handful of physicians, the Washington Post report provided their first introduction to the University Group Diabetes Program (UGDP) study, whose findings were to be the subject of a decade long controversy concerning the safety and utility of oral hypoglycemic drugs.

Few recent controversies in medicine are comparable in length and rancor to that over the UGDP. More extended debates--over the treatment of breast cancer or the safety of oral contraceptives--never focused as much attention, for so long, on a single study. And however much suspicions about the integrity of researchers may have figured in privately held judgments in other disputes, the UGDP is unusual in the degree to which such accusations were publicly traded. As an example

---

<sup>1</sup> Morton Mintz, "Antidiabetes Pill Held Causing Early Death," Washington Post May 21, 1970, 1, 7.

<sup>2</sup> Harold M. Schmeck, Jr. "Scientists Wary on Diabetes Pill," New York Times May 22, 1970, 56.

of a study which tried-and failed-to resolve a controversy about the merits of medical treatment, the UGDP certainly represents an extreme case. But if the intensity of the debate over the UGDP was atypical, the conceptual issues and political dilemmas raised by the study are not.

At the level of statistical theory, the UGDP posed questions about the proper relation between statistical inference and scientific decisions: what criteria apply in deciding when a clinical experiment should be stopped? how are unanticipated results from a unique study to be interpreted?<sup>3</sup> At the level of policy, the controversy posed questions about the appropriate relation between scientific claims and regulatory decisions: which studies count as proper evidence? when competing experts disagree about the merits of a study, how are the disagreements adjudicated? how are the results of clinical trials to be translated to medical practice? None of these questions are unique to the UGDP. The persistence of the debate over the UGDP, however, calls attention to a prior methodological problem which must be addressed: how is the controversy to be explained?

To proponents of the UGDP, the persistence of the controversy requires little explanation. They regard the study as a pioneering effort whose conclusions have been repeatedly affirmed as valid in the

---

<sup>3</sup> Jerome Cornfield, "Recent Methodological Contributions to Clinical Trials," American Journal of Epidemiology 104 (1976), 408-421; Paul Meier, "Statistics and Medical Experimentation," Biometrics 31 (June, 1975), 511-529.

face of intensive scrutiny.<sup>4</sup> The critics' failure to accept the study can only be explained as due to non-rational factors: ignorance or cupidity. Not surprisingly, the UGDP's critics regard it as a flawed study from which no conclusions can be drawn, and suggest that support for the UGDP is the result of the clan loyalty of statisticians and their camp followers.<sup>5</sup> In each case, there are purportedly two sides, truth and error; and only error calls for a social explanation, truth providing its own justification.

Historians and sociologists of science have found that asymmetric arguments of this sort: "I have reasons but you have prejudices," are not uncommon in scientific debate. Such debates are frequently found to entail disagreements not merely about the interpretation of individual experiments, but about the rules by which different experimental evidence is evaluated and weighed. The analyst's first duty is to examine the way in which both parties to a debate have constructed their arguments, and to identify the underlying methodological issues which each party believes to be at stake.<sup>6</sup> In the case of the UGDP, what began as a debate over the merits of a particular finding broadened into a dispute about the relative merits of statistical and clinical

---

<sup>4</sup> Thaddeus E. Prout, Genell L. Knatterud and Curtis Meinert, "Diabetes Drugs: Clinical Trial," [Letters] Science 204 (April 27, 1979), 362-363; Martin Goldner, ibid., 363-365.

<sup>5</sup> Holbrooke S. Seltzer, "Efficacy and Safety of Oral Hypoglycemic Agents," Annual Review of Medicine 31 (1980), 261-272; Alvan R. Feinstein, "How Good is the Statistical Evidence Against Oral Hypoglycemic Agents?," Advances in Internal Medicine 24 (1979), 73-76.

<sup>6</sup> G. Nigel Gilbert and Michael Mulkay, "Warranting Scientific Belief," Social Studies of Science 12 (1982), 383-408; G.D.L. Travis, "Replicating Replication? Aspects of the Social Construction of Learning in Planarian Worms," Social Studies of Science 11 (1981), 11-32.

expertise in drawing inferences from the study. This scientific debate was exacerbated by the FDA's decision to draw on the UGDP's report in modifying the package labeling for the oral hypoglycemic drugs. The agency's action led critics of the study to publicly challenge both the validity of the UGDP's conclusions and the basis of the FDA's authority for prescribing the labeling of drugs.

To understand the controversy, it is necessary to understand both the clinical context in which it arose and the political context in which it was addressed. The remaining sections in this introduction describe the state of knowledge about the management of diabetes in the period just prior to the design of the UGDP. The study, and the development of the controversy are described in part II. The efforts of the FDA and other groups to adjudicate the dispute are presented in part III, followed by an account of the controversy's implications.

### The Disease

"Diabetes" refers to a variety of metabolic disorders characterized by a common inability to efficiently process and use foods containing sugar. Normally, the body processes sugar with the aid of a hormone, insulin. For a variety of reasons, only partially understood, diabetics either fail to produce sufficient quantities of insulin or, in some instances, are unable to effectively employ their insulin in processing sugars.

Specialists currently distinguish between two broad classes of diabetes, Type I and Type II, each with its own distinctive etiology,

symptoms, and prognosis.<sup>7</sup> Type I, or insulin dependent diabetes (once termed "juvenile diabetes"), usually appears by age 40 and entails a total or near total inability to produce insulin. Prior to the introduction of insulin in 1922, the majority of such patients died within five years of diagnosis.<sup>8</sup> The vast majority of diabetics, however, are known as Type II diabetics, formerly termed "adult-onset diabetes".<sup>9</sup> "Non insulin dependent diabetes," as it is now known,

---

<sup>7</sup> The current classification and nomenclature was introduced in 1978 by the National Diabetes Data Group, National Institutes of Health. In addition to distinguishing between Type I and II diabetes, they recognized three additional syndromes: gestational diabetes, or diabetes associated with pregnancy; diabetes secondary to other conditions; and impaired glucose tolerance, formerly termed "chemical" or "sub-clinical" diabetes. While the specific labeling of Types I and II is relatively recent, it represents an official recognition of a distinction which has long been clinically recognized. The principal significance of the new terminology was to provide a standardized terminology for clinical and epidemiological research, and to negate the customary association between age of onset and type of diabetes which was embodied in the traditional nomenclature. See National Diabetes Data Group, "Classification and Diagnosis of Diabetes Mellitus and Other Categories of Glucose Intolerance," Diabetes 28 (1979), 1039-1057.

<sup>8</sup> Mortality rates from diabetic coma had improved from 64% to 42% with the introduction of Allen's "semi-starvation" therapy in 1915. Mayer B. Davidson, "The Continually Changing 'Natural History' of Diabetes Mellitus," Journal of Chronic Disease 34 (1981), 5. For descriptions of the dramatic impact of insulin on the lives of young diabetics, see Michael Bliss, The Discovery of Insulin (Chicago: University of Chicago Press, 1984).

<sup>9</sup> Precise up-to-date numerical estimates of the two types of diabetes are not available; it is known, however, that 81% of existing diabetics are 45 years or older, and that 74% of newly diagnosed diabetics are 45 years or older. These statistics do not refer, however, to the date at which diabetes first manifested itself. Report of the National Commission on Diabetes to the Congress of the United States. vol. 3. part 1. Scope and Impact of Diabetes (1) (Washington: Department of Health, Education and Welfare (pub) no. (NIH) 77-1021, 1976), 71. 73. The problem is further exacerbated by the fact that the severity and precise nature and development of the impairment in metabolism are more significant than exact age of onset in categorizing types of diabetes. Moreover, estimates of the incidence of Type II diabetes, in particular, are particularly sensitive to changes in

represents a less severe condition, in which the body often retains some capacity to produce insulin, but cannot employ it effectively to prevent elevated blood sugar levels.<sup>10</sup> Such individuals frequently do use insulin but unlike insulin dependent diabetics, they do not require insulin injections to stay alive.

Despite the discovery of insulin, diabetics of both types have substantially lower life expectancy than the non-diabetic population, and remain at elevated risk for a series of life-threatening and disabling complications. Type II diabetics in particular are at substantially higher risk for death from cardiovascular disease and stroke, as well as being subject to diabetes related blindness and kidney disease.<sup>11</sup> While physicians can usually distinguish, on the basis of laboratory and clinical findings, between the diagnosis of Type I and Type II diabetes, the severity of the disease and its long run complications vary considerably from individual to individual within each group.<sup>12</sup> These complications represent the major threat

---

official diagnostic criteria. See L. Joseph Melton III, Pasquale J. Palumbo and Chu-Pin Chu, "Incidence of Diabetes Mellitus by Clinical Type," Diabetes Care 6 (January-February, 1983), 82-85.

<sup>10</sup> For a recent review on the etiology of Type II diabetes, see Edward S. Horton, "Role of Environmental Factors in the Development of Noninsulin-Dependent Diabetes Mellitus," American Journal of Medicine 74 (November 30, 1983), Supplement: 32-40.

<sup>11</sup> Type II diabetics are not subject to diabetic retinopathy and nephropathy (kidney disease) to the same degree as Type I diabetics; they are however, at greater risk than the general population. Davidson, "The Changing 'Natural History' of Diabetes Mellitus," (n. 8) 6-7.

<sup>12</sup> Even the group which generated the classification schema acknowledges that discrimination between Type I and Type II diabetes may be difficult in certain individuals. National Diabetes Data Group, "Classification and Diagnosis of Diabetes Mellitus," (n. 7) 1044. Of far greater consequence, however, is the heterogeneity of



to the well being and future life expectancy of adult diabetics; whether or not they are preventable has been the subject of a long standing controversy among specialists.

### The Treatment

Effective and appropriate management of diabetes is a complicated and controversial task. What all therapeutic regimens have in common is an attempt to keep the blood sugar levels of the patient somewhere within "normal" range--the means by which this is accomplished and the degree to which blood sugar levels are strictly controlled vary from physician to physician and patient to patient. One school has held that the development of complications is directly and causally linked to the failure to maintain strict control of blood sugar. Others have maintained that the complications are independent manifestations of the disease process, whose progression is largely unaffected by blood sugar levels. The controversy has proven difficult to resolve: when analysis are made on the basis of retrospective chart review, it has been difficult to sort out the confounding effects of duration of disease, severity of disease and mechanisms of patient selection. The apparent heterogeneity of the underlying disorder aggravates the problem: some patients develop more complications, and sooner after diagnosis than

---

symptoms and prognosis within each class. See Stefan S. Fajans, Michael C. Cloutier, and Robert L. Crowther, "Clinical and Etiological Heterogeneity of Idiopathic Diabetes Mellitus," Diabetes 27 (1978), 1112-1125. These underlying differences are compounded by the fact that physicians differ substantially in the proportion of patients they would classify as "diabetic," given a specific laboratory result. See Kelly M. West, "Substantial Differences in the Diagnostic Criteria Used By Diabetes Experts," Diabetes 24 (July, 1975), 641-644.

others. Is the rapid progression of complications in these patients due to uncontrolled glucose levels, or are the complications and the difficulties in controlling their blood sugar both evidence that these patients have a more severe form of the disease?<sup>13</sup>

This theoretical dispute has substantial implications for the lives of patients: the more one believes in the long run medical benefits of control, the more likely one is to advocate aggressive management of diet and life-style and/or the use of insulin, despite its inconveniences and hazards. Patients using insulin must do more than simply inject the drug at regular intervals. Use of insulin, activity and diet must be kept in an even balance. An error in insulin dosage or a failure to snack between meals can lead to an excess amount of insulin and a condition known as insulin shock.<sup>14</sup>

<sup>13</sup> Philip K. Bondy, "Therapeutic Considerations in Diabetes Mellitus," in F.J. Ingelfinger, A. Relman and M. Finland, eds. Controversy in Internal Medicine (Philadelphia: W.B. Saunders Company, 1966), 499-500; Alexander Marble, "Control of Diabetes Lessens or Postpones Vascular Complications," ibid., 493-496. While agreeing on the difficulty of studying the problem, Bondy and Marble take opposing views on the use of "strict" control. For an assessment of the state of knowledge regarding the pathogenesis of vascular complications around the time the UGDP study was designed, see Alexander Marble and George F. Cahill, The Chemistry and Chemotherapy of Diabetes Mellitus (Springfield: Charles C. Thomas, 1962), 75-76.

<sup>14</sup> The rate of errors in insulin dosage or failure to adhere to diet ranges as high as 50% in some observational studies. See Julia Watkins, T. Franklin Williams, Don A. Martin, et. al., "A Study of Diabetic Patients at Home," American Journal of Public Health 57 (March 1967), 453-455; Kelly M. West, "Diet Therapy of Diabetes: Analysis of a Failure," Annals of Internal Medicine 79 (1973), 426. The possibility that immunologic responses to insulin bring about further autoimmune damage to the pancreas gave physicians another reason to avoid the use of insulin. George F. Cahill, Jr., "Some Thoughts Concerning the Treatment of Diabetes Mellitus," in Ingelfinger, et. al., Controversy in Internal Medicine (n. 13) 509.

Concerned about these side effects of insulin, physicians welcomed the introduction to the United States in 1956, of tolbutamide, an orally administered drug for diabetics. One of several compounds initially intended as substitutes for insulin in treating juvenile diabetics, tolbutamide was soon found to be of little use in these patients. It was, however, reported effective in a substantial number of adult-onset diabetics, especially in individuals whose diabetes appeared after age 40 and those whose daily insulin requirements were under 20 units.<sup>15</sup>

The oral hypoglycemic agents soon became a popular substitute for insulin injections in Type II diabetes.<sup>16</sup> Their primary advantage was convenience. Not only was the need for injections eliminated, but an end to insulin injections reduced the need for between meal feedings, formerly necessary to ensure that blood sugar levels did not drop to

---

<sup>15</sup> Tolbutamide, which belongs to a class of compounds known as sulfonylureas, was developed in the mid-1950's, after German investigators observed the hypoglycemic effects of chemically related compounds used in the treatment of infections. One of these related compounds, carbutamide, was introduced into the use late in 1955 but withdrawn due to toxic side effects by the fall of 1957. See Helmut Mehnert, Rafael Camerini-Davalos and Alexander Marble, "Results of Long-Term Use of Tolbutamide (Orinase) in Diabetes Mellitus," J.A.M.A. 167 (June 14, 1958), 818. For a review of the history of research for insulin substitutes, and an overview of the initial American clinical findings, see C.J. O'Donovan, "New Orally Effective Adjuvants in the Management of Diabetes Mellitus," Journal of Chronic Diseases 4 (December, 1956), 635-643. O'Donovan reports that the German work was preceded by earlier French investigations reporting the hypoglycemic effects of these compounds, but that wartime publication may have inhibited dissemination of the French results.

<sup>16</sup> One early study of the use of tolbutamide found that 61.7% of over 9,000 patients studied had formerly been managed by a combination of dietary restrictions and insulin. See C.J. O'Donovan, "Analysis of Long Term Experience with Tolbutamide (Orinase) in the Management of Diabetes," Current Therapeutic Research 1 (November, 1959), 74.

excessively low levels following the injection.<sup>17</sup> Experts soon recommended tolbutamide as the drug of choice in cases of mild diabetes (i.e. those with no history of ketoacidosis and with insulin use under 40 units daily); tolbutamide's special attractiveness lay in the low rate of toxic reactions and hypoglycemic episodes it provoked.<sup>18</sup> Although both experts and the drug's manufacturer, Upjohn Company warned that tolbutamide was no substitute for maintaining proper dietary restrictions, in fact, its use was associated with a further de-emphasis on dietary control, in which tolbutamide and other oral hypoglycemic agents "became readily used as substitutes for dietary restrictions."<sup>19</sup>

---

<sup>17</sup> Ibid., 85. See also Mehnert, et. al. "Results of Long-Term Use of Tolbutamide," (n. 15) 826-827. On the preferences of patients for oral drugs over insulin, even against medical advice, see Marble and Cahill, The Chemistry and Chemotherapy of Diabetes (n. 13), 141, 158.

<sup>18</sup> Craig M. Arnold and Ronald W. Lauener, "Oral Hypoglycemic Drugs," Canadian Medical Association Journal 91 (August 22, 1964), 395. Tolbutamide is less likely than other oral hypoglycemic agents to lower blood sugar levels excessively because it has a short "effective span of action." Thus regular doses are less likely to build up to produce hypoglycemia. The low toxicity and rapid excretion were also reported by the American Medical Association's Council on Drugs in their initial assessment of tolbutamide, although they stressed the possibility of acute hypoglycemia when trying to transfer patients from insulin to tolbutamide. See Council on Drugs, "New and Nonofficial Drugs. Tolbutamide," J.A.M.A. 164 (July 20, 1957), 1333-1335. Researchers from the Joslin Clinic, which had helped to introduce the drug in the United, did not feel that such acute hypoglycemia was a significant problem. See Mehnert, et. al. "Results of Long-Term Use of Tolbutamide," (n. 15), 826-827.

<sup>19</sup> R. Camerini-Davalos, O. Lozano-Casteneda and A. Marble, "Five Years Experience with Tolbutamide," Diabetes 11 (1962), Supplement 1, p. 78; Council on Drugs, "New and Nonofficial Drugs. Tolbutamide," (n. 19), 1333-1335; "Orinase [Upjohn Company Flyer]," in W.J.H Butterfield and W. Van Westering, eds. Tolbutamide ...After Ten Years (New York: Excerpta Medica Foundation, 1967), 326.

## II. The Study and the Controversy

The increasing use of tolbutamide during the late 1950's generated widespread interest among physicians specializing in diabetes. Particularly intriguing to researchers and clinicians was the suggestion that patients using tolbutamide might develop fewer cardiovascular complications, or develop them later than patients on other treatments.<sup>20</sup> The National Institute of Arthritis and Metabolic Diseases (NIAMD) accordingly began planning a multi-clinic evaluation of the oral agents. At the suggestion of Christian Kliment, a young epidemiologist recruited to the study, the investigators decided on a randomized, placebo controlled design to evaluate the standard treatments then used in treating adult-onset diabetes: insulin, tolbutamide, and diet.<sup>21</sup>

The UGDP was organized as a multi-center trial with three aims:

1. To evaluate the effects of controlling blood glucose levels on the development of vascular (and other) complications;
2. To study the natural history of these complications;

---

<sup>20</sup> Rafael A. Camerini-Davalos and Harold S. Cole, eds., Early Diabetes, 1st International Symposium on Early Diabetes, 1968 (New York: Academic Press, 1970), 443-470; Rafael A. Camerini-Davalos, "Treatment of Chemical Diabetes," in Butterfield, Tolbutamide ... After Ten Years (n. 19), 231-232; Robert Felman and Dwight Fitterer, "The Prophylactic use of Oral Hypoglycemic Drugs in Asymptomatic Diabetes," in Tolbutamide...After Ten Years (n. 19), 243-254; Martin Goldner, Gennell L. Knatterud, and Thadeus E. Prout, "Effects of Hypoglycemic Agents on Vascular Complications in Patients with Adult Onset Diabetes," J.A.M.A. 218 (November 29, 1971), 1400.

<sup>21</sup> Testimony of Floyd Daft (Director, NIAMD) in U.S. Congress. House of Representatives. Committee on Appropriations. Hearings. Departments of Labor, and Health, Education and Welfare Appropriation for FY 1960, 751. 86th Congress, 1st Session; Harry M. Marks, Interview with Christian R. Kliment, May 16, 1984.

3. To improve methods in clinical trials.<sup>22</sup>

The study's primary purpose was to resolve the long-standing controversy among physicians as to whether strict and effective control of blood sugar levels would delay or prevent the onset of vascular complications. Much of the evidence in support of this claim was based on the accumulated clinical experiences of individual practitioners, or on retrospective analyses of patients' records from diabetes clinics. A prospective, randomized and double blinded study, it was hoped, might avoid the problems of selection bias and partisan assessment which had heretofore plagued historically controlled studies.<sup>23</sup>

The UGDP was intended as a model of clinical investigation. Newly diagnosed diabetics from twelve university clinics were randomly assigned to one of four treatment groups: a group receiving insulin at dosages which were adjusted at regular intervals to keep blood glucose

<sup>22</sup> University Group Diabetes Program, "Study of the Effects of Hypoglycemic Agents on Vascular Complications in Patients with Adult-Onset Diabetes. I. Design, Methods and Baseline Results," Diabetes (1970) Supplement 2, 747. The UGDP's role as an "almost unique example of a prospective study" and the "promise of secondary gains in the methodology of clinical trials" was strongly emphasized by the committee which considered their initial application. Special Review Committee, General Comments on Application 06876-01, May 15, 1960. [on file at Division of Research Grants, National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases; hereafter cited as DRG, NIADDKD.]

<sup>23</sup> Harvey C. Knowles, "The Problem of the Relation of the Control of Diabetes to the Development of Vascular Disease," Transactions of the American Clinical and Climatological Association 76 (1964), 142-147. Not all specialists agreed that a prospective trial, however desirable was feasible. The Joslin Clinic's Alexander Marble raised questions about the feasibility of such a trial in Type II diabetics, in whom the effects of treatment on the appearance of vascular complications was potentially confounded by the duration of undiagnosed disease, and by the prevalence of vascular disease unrelated to diabetes in such a middle-aged population. See Marble, "Control of Diabetes," (n. 13) 495.

levels at targeted levels (IVAR); a group receiving a fixed insulin dosage (ISTD); a group receiving the oral hypoglycemic agent, tolbutamide (TOLB); and a group receiving a placebo. Neither patients nor treating physicians knew who was receiving tolbutamide and who placebo. At the onset of the study, all 1027 patients received instruction in the use of a standard diet for diabetics, regardless of which treatment group they belonged to.<sup>24</sup>

The program of data collection was to be both comprehensive and exemplary: examinations were scheduled upon admission to the trial and at yearly intervals thereafter. The investigators' focus was on the incidence of complications, and they gave little thought to the possibility of finding differences in mortality among the treatment

---

<sup>24</sup> Insulin dosages were calculated by a formula relating insulin to body surface: in the IVAR group, dosages would be adjusted at intervals according to how well controlled the patient was; in the ISTD group, initial dosages tailored to individuals were selected in the same way, but unless the patient's health was threatened, the dosages were not to be adjusted. Naturally, it was impossible to blind clinicians or patients to the fact that they were injecting insulin; all evaluations, however, were done by laboratories blind to the patient's treatment status. A fifth treatment group, receiving phenformin, another oral hypoglycemic agent, was added in the second year of the study in 6 of the UGDP's 12 participating clinics. In order to obtain sufficient numbers of patients in the phenformin group, 6 patients of every 14 were allocated to phenformin in these clinics, and two each were assigned to IVAR, ISTD, placebo or tolbutamide. See UGDP, "I. Design, Method and Baseline Results," (n. 22) 748-750. The effect of this procedure was to drastically reduce the number of patients on tolbutamide and placebo in half the UGDP clinics, a decision which subsequently became the source of some controversy in interpreting the results; the number of tolbutamide and placebo patients in the study overall remained equal, however.

groups.<sup>25</sup> As the study progressed, however, routine monitoring of mortality began to indicate an unfavorable trend for patients on tolbutamide, particularly for those dying of cardiovascular causes. The statistical director, Christian Klimt, first called the attention of this trend to the other investigators in 1967, and began to search for baseline differences which might explain results which, if anything, were the opposite of what might have been expected when the trial was conceived.<sup>26</sup> Klimt, who monitored the data on a weekly basis, became increasingly concerned about the trend. At the onset, few of the clinicians agreed:

"They couldn't conceive, nobody could conceive, a drug which had been by that time, I don't know, eighteen years on the U.S. market ...could potentially do harm of such a serious nature that you would get a difference in total death. And by derivation, you know, a much greater difference in cardiovascular mortality."<sup>27</sup>

Although numerous indications cast suspicion on the drug, traditional statistical procedures provided limited guidance for interpreting such interim results. Finding themselves in uncharted territory, the UGDP investigators requested that two outside statisticians review

<sup>25</sup> The study was initially funded for only five years and one criteria of eligibility was that patients enrolled in the study be expected to live at least that long. The investigators did not, therefore, expect to have a sufficient number of deaths to evaluate the effects of treatment on mortality. The primary outcomes were the neurological, retinal and vascular complications characteristic of diabetes. Substantial mortality differences began to appear only after the initial grant was renewed and the life of the study extended. Marks, Interview with Klimt, May 16, 1984. See also the renewal application [06875-06] from the UGDP investigators, on file at ERG, NIAID/NIH.

<sup>26</sup> Marks, Interview with Klimt, May 16, 1984. See also the studies cited in note 20.

<sup>27</sup> Marks, Interview with Klimt, May 16, 1984.



Klimt's findings.<sup>28</sup> Meanwhile, the mortality differences between patients on tolbutamide and placebo continued to mount. By early 1969, Klimt saw little point in continuing: even if the effect of tolbutamide were not as strong as the data suggested, it was no longer possible to demonstrate a benefit for the drug. Continued use of tolbutamide by the UGDP, in his view, was unethical. Other investigators, less sure of the findings, felt that the study should continue with the treatment, if for no other reason than to convincingly demonstrate tolbutamide's harmful effects to others.<sup>29</sup>

These differences of opinion erupted at a meeting of the UGDP's principal investigators called in June, 1969, to discuss the executive committee's recommendation that treatment with tolbutamide be discontinued. After two days of discussion, the majority voted 21-5 to stop

---

<sup>28</sup> Baseline differences in cardiovascular disease did not seem to account for the findings: mortality differences between tolbutamide and placebo patients appeared more pronounced for patients who entered the study without signs of cardiovascular disease than for those with prior cardiac problems. Analysis of the data as of September, 1968 found that mortality differences were greatest for patients who had been in the study longest and held up best for patients with good adherence to treatment. Nonetheless, the statistical procedures for analyzing such interim mortality results were hardly conclusive: alternative approaches applied to this data gave different indications. Klimt selected Jerome Cornfield and Byron Brown, Jr. as outside reviewers; Cornfield was currently doing work on the theory of interim analysis in conjunction with the Coronary Drug Project, on which he and Klimt were working. See Summary of the September 1968 UGDP Progress Report and the Resolutions Adopted at UGDP Investigators Meeting of October 4, 1968 and Jerome Cornfield to Max Miller, June 12, 1969. Jerome Cornfield papers, Mathematical Association of America Archives, University of Texas at Austin [hereafter cited as Cornfield papers]. NIH officials approved Klimt's request for outside consultants on November 20, 1968. LeMar Remmert to Christian Klimt, November 20, 1968 [DRG, NIAADK.]

<sup>29</sup> Marks, Interview with Klimt, May 16, 1984.

using the drug. The FDA and the three firms contributing drugs to the study were to be notified immediately of the decision.<sup>30</sup>

Stopping Tolbutamide: Evidence, Beliefs and Decisions

The UGDP's decision to terminate the use of tolbutamide before completion of the study was a reasonable one, based on the premise that no matter how long the study continued, it was unlikely (but not impossible) that it would demonstrate a net benefit for the drug. Given this assessment, a majority of the investigators agreed with Klumt that it was unethical to expose patients any longer to the possible risks associated with the drug. Their subsequent presentation to the scientific community succinctly conveyed this rationale:

...the findings of this study indicate that the combination of diet and tolbutamide therapy is no more effective than diet alone in prolonging life. Moreover, the findings suggest that tolbuta-

---

<sup>30</sup> See Minutes, Executive Committee, UGDP, May 28, 1969; Minutes, Principal Investigators Meeting, UGDP, June 5-6, 1969. These materials are in the possession of Thomas C. Chalmers, M.D. I am indebted to Dr. Chalmers for making them available to me. [Hereafter cited as Chalmers papers.] According to Angela Bowen, one of the dissident members of the group, the initial votes were more even, and it took extensive lobbying to bring the vote to the 21-5 result. [Interview, Bowen]. The minutes do not contain counts of any previous votes, though they document extensive discussion among participants. The possibility that tolbutamide was not toxic remained a real one for some participants. Jerome Cornfield, one of two outside statistical consultants, was reluctant to "unequivocally" conclude that the drug was toxic, although Cornfield made it clear that he personally would refuse the drug. He nonetheless recommended that treatment with tolbutamide be discontinued; see Cornfield to Max Miller, June 12, 1969. Cornfield papers. After the FDA expressed doubts about the strength of the finding, Max Miller, the UGDP's chairman, again raised the possibility of continuing the tolbutamide component of the study to establish a more definitive finding. Miller appears, however, largely to have been looking for NIAMD to take an official stance on the decision. Minutes, FDA-UGDP Meeting, June 16, 1969; Dr. Remmert [Diabetes Program Director, NIAMD] to Donald Whedon [Director NIAMD] June 16, 1969; Chalmers papers.

mide and diet may be less effective than diet alone or than diet and insulin at least insofar as cardiovascular mortality is concerned. For this reason, use of tolbutamide has been discontinued in the UGDP.<sup>31</sup>

It is difficult to imagine a more cautiously worded or carefully reasoned conclusion. Even so, agreement entailed several collateral beliefs, which proved not to be universally shared:

- 1) That the conclusion regarding the potential risks of using tolbutamide was, on balance, likely to be correct;
- 2) correspondingly, that alternative explanations of the heightened cardiovascular mortality in the tolbutamide group were implausible;
- 3) that the study had adequately demonstrated the lack of efficacy, as well as the potential hazards, of tolbutamide; and
- 4) that the study had appropriately measured the potential benefits of using the drug.

In the majority's opinion, the rigorous experimental design of the UGDP contributed greatly to their belief in these premises and their confidence in the decision. A minority of the UGDP investigators, however, strongly questioned the strength of the evidence presented, on the basis of doubts about these underlying premises.

In particular, investigators whose own experience with the drug had been favorable thought that differences in medical management among the participating clinics might just as well explain the level of mortality in the tolbutamide group. The tolbutamide mortality was heavily concentrated in four of the clinics, leading to the inference that differences in medical care or patient risk factors, not differen-

---

<sup>31</sup> UGDP, "Study of the Effects of Hypoglycemic Agents on Vascular Complications in Patients with Adult-Onset Diabetes. II. Mortality Results," Diabetes 19 (1970) Supplement 2, 814.

oes in the drug, accounted for the finding.<sup>32</sup> The possibility that alternative explanations existed for the excess cardiovascular mortality, coupled with an underlying belief in the benefits offered by the drugs, led these participants to quarrel with the decision. Yet had it not been for the events which accompanied the announcement of the UGDP decision the following spring, these differences of opinion might never have led to such a vociferous controversy.

### Tolbutamide: Birth of A Controversy

Between June, 1969 and May, 1970, the UGDP investigators were preparing their data for presentation at the American Diabetes Association's (ADA) annual meeting. Meanwhile, knowledge of their suspicions regarding tolbutamide was confined to a small group of specialists in diabetes, drug evaluation, statistics and clinical research.<sup>33</sup> Several weeks before the UGDP presented its results at the June, 1970 meeting, the Washington Post and the New York Times publicly announced the

---

<sup>32</sup> Not all the doubts initially expressed came from those who dissented from the ultimate decision. One investigator requested a re-analysis of the autopsy data, to see if the differences in cardiovascular deaths would be sustained on closer examination. See Minutes, Principal Investigators Meeting, UGDP, June 5-6, 1969, Chalmers papers and Harry M. Marks, interviews with Robert Reeves and Angela Bowen, November 10, 1981.

<sup>33</sup> Apart from the UGDP investigators and their consultants, individuals aware of the findings included Alvin Feinstein, recruited as a consultant for Upjohn Pharmaceutical Co. to review the findings early in 1969; representatives of Upjohn, members of the FDA's Medical Advisory Board who heard a presentation on the UGDP's findings from the FDA's Charles Anello and Edwin Ortiz in June 1969; Donald Whedon, director of NIAMD, and other NIAMD officials; and various specialists in the treatment of diabetes, including Robert Bradley, subsequently chair of the Committee for the Care of the Diabetic which was highly critical of the UGDP study.

study's findings.<sup>34</sup> The FDA quickly issued a summary statement, announcing their intention to revise the labeling for tolbutamide and other sulfonyurea drugs.<sup>35</sup>

For the 26 UGDP investigators, the decision to terminate tolbutamide had been difficult. Yet the difficulty they faced paled in comparison with the next step: reassessing tolbutamide for the 800,000 patients estimated to be using the drug.<sup>36</sup> The premature announcement of the UGDP findings left physicians around the country with few answers for diabetics calling to ask about the safety of their current medication.<sup>37</sup> Abstracts of the study published in Diabetes at the end of May added only a few paragraphs of information concerning its design and startling conclusion. Max Miller, chairman of the UGDP group, while deploring the premature notice of the findings, felt it "inappropriate" to comment further before the study received peer review.<sup>38</sup> Like their

<sup>34</sup> Morton Mintz, "Antidiabetes Pill Held Causing Early Death," Washington Post May 21, 1970, 1, 7. The Post article was followed by one in the New York Times, the following day: Harold M. Schmeck, Jr. "Scientists Wary on Diabetes Pill," New York Times May 22, 1970, 56. Of the two, the initial Post article was the more alarmist. Given the number of individuals with some knowledge of the study at this point, it is not surprising that the individual responsible for the leak has never been identified. For more on the circumstances surrounding the FDA announcement, see below, pp.

<sup>35</sup> "FDA Statement. Friday, May 22, 1970," Diabetes 19 (1970), Supplement 1, 467.

<sup>36</sup> Morton Mintz, "Antidiabetes Pill Held Causing Early Death," Washington Post May 21, 1970, 1, 7.

<sup>37</sup> "The Tolbutamide Debate," Medical World News 12 (January 8, 1971), 39.

<sup>38</sup> Thaddeus E. Prout and Martin G. Goldner, "The University Group Diabetes Program: The Effects of Hypoglycemic Agents on Vascular Complications in Patients with Adult Onset Diabetes. 2. Findings at Baseline. 3. Course and Mortality," Diabetes (1970), Supplement 1,

patients, most physicians remained reliant on the various press reports which appeared over the summer of 1970.<sup>39</sup> Criticisms of the study in the medical press were equally fragmentary and even less informative.<sup>40</sup>

Evaluation of a complex study like the UGDP customarily progresses at what seems like a glacial pace, in the relative obscurity of specialist communities. The experts can consider themselves fortunate if anyone pays immediate attention to what they eventually decide. In the UGDP's case, however, attitudes and opinions were formed in the glare of publicity, and exchanged in the popular media alongside discussion in more sheltered forums. More than one disputant discovered an unexplored gift for polemic along the way. The early notoriety given the UGDP did much to upstage and short circuit the customary processes of peer review. The widespread publicity announcing the study imparted urgency to each deliberation: prompt action was called for to resolve the uncertainties of patients and physicians alike.

In the ensuing debate, much of the subtlety of the UGDP's reasoning concerning the use of tolbutamide was lost. Their initial decision had been premised on a complex calculus of risk and benefit, issuing in the judgment that in the absence of benefit, any evidence of risk needed to

---

374-375. This supplement, paginated to follow the June issue, arrived at the Francis Countway Library of Medicine of Harvard University on May 26, 1970. For Miller's remark, see "Statement of Chairman of UGDP [May 21, 1970]," Diabetes 19 (June, 1970), 467-468.

<sup>39</sup> Harold M. Schmeck, Jr. "Diabetes Drug Use Backed By Council," New York Times June 15, 1970, 42; Harold M. Schmeck, Jr., "Pills for the Diabetic: Dilemma for Doctors," New York Times June 22, 1970, 1, 25.

<sup>40</sup> Robert L. Reeves, "Erroneous Interpretation of Data," Northwest Medicine 69 (July, 1970), 870; "The Tolbutamide Controversy," JAMA 213 (August 3, 1970), 861.

be heavily weighted. By and large, beliefs about the possible benefits of tolbutamide remained in the background of subsequent discussions, despite the repeated efforts of UGDP supporters to center the debate on this issue. The spotlight focused instead on the critics' claims that the UGDP's conclusion about excess mortality associated with tolbutamide was invalid, bringing issues about experimental design and inference to the fore.<sup>41</sup>

By December, 1970, when the UGDP published extensive details of the study's design and findings, their opponents' basic arguments were already well staked out. Echoing concerns of the UGDP's internal dissenters, critics raised questions about the analysis and distribution of cardiovascular disease among patients in the study. Despite the randomization procedure, the tolbutamide group appeared to be sicker, particularly with regard to cardiovascular risk factors. While none of these baseline differences appeared to be statistically significant, questions remained as to their cumulative effect. Particular concern was raised about omitted risk factors such as smoking, not taken into account when the study was designed.<sup>42</sup> The second basic challenge to

---

<sup>41</sup> The focus on tolbutamide mortality and on the UGDP's lack of validity are emphasized in Statement on the Treatment of Diabetes, telegraph to the Commissioner of Food and Drugs, December 1, 1970, from assembled critics of the UGDP. FDA Docket 75N-0062. Vol. 2.

<sup>42</sup> On the distribution of cardiovascular risk factors, see [Alvin Feinstein], An Analytic Critique of the UGDP Protocols [June, 1969], 19-20; *idem*, A Supplemental Critique of Recent UGDP Reports [August, 1969], 4-6; and E. Keith Borden [Upjohn Company], Preliminary Evaluation of the UGDP report of May 9, 1969 [June, 1969], 2, 7-9. On the question of omitted risk factors, see Feinstein, Supplemental Critique, 4. All documents cited are in Cornfield papers. These criticisms were echoed and elaborated in Feinstein's published article: "Clinical Biostatistics VIII. An Analytic Appraisal of the University Group Diabetes Program (UGDP) Study," Clinical Pharmacology and Therapeutics 12 (1971), 172-173; 177-178; 185-189. See also Stanley Schor, "The University Group Diabetes Program. A Statistician Looks at the Mortality Results," JAMA 217 (September 20, 1971), 1671-1673.

the tolbutamide finding concerned the distribution of deaths within the study. Deaths appeared to be concentrated in a minority of clinics, raising questions about the selection or management of patients in these cities.<sup>43</sup>

Proponents of the UGDP took remarks about baseline inequalities within the study populations as fundamental criticisms, challenges to what is sometimes termed the "internal validity" of a study. Evidence of such inequalities, even in a subset of the clinics, might be taken as grounds for believing that the randomization procedure had "broken down", allowing physician bias in patient assignment to enter. In their initial report, the UGDP examined this issue at length. Of ten cardiovascular risk factors, there was a statistically significant difference between the tolbutamide and placebo groups for only one--serum cholesterol. Given the number of factors recorded, chance alone might well be responsible for that difference.<sup>44</sup> As for variations in mortality by clinic, the UGDP had looked extensively for differences in population characteristics or cardiovascular risk factors which could explain these: among the factors they examined, they found none of statistical significance.<sup>45</sup> Subsequent reviews came to similar conclusions: the variations in mortality among the clinics were no more than might be expected by chance, and the known differences in risk

---

<sup>43</sup> Borden, Preliminary Evaluation of the UGDP, 2-6; Feinstein, "An Analytic Appraisal," 177-178; S. Schor, "A Statistician Looks," 1673-1674 [All references in n. 42].

<sup>44</sup> UGDP, "II. Mortality Results," (n. 31) 799-803.

<sup>45</sup> UGDP, "II. Mortality Results," (n. 31) 807-809, 823-826.



factors could not explain the increased mortality in these clinics.<sup>46</sup>

For statisticians who viewed the UGDP favorably, demonstrating the integrity of the randomization provided equal reassurance about the role of risk factors like smoking, on which no data was collected. Critics' arguments about omitted risk factors showed a misunderstanding of the ability of randomization "to achieve approximate comparability with respect to all variables, whether known or not." The basic rationale for introducing randomized allocation in medicine was that so many of the factors which might influence outcome are unknown. Since there was no evidence that the randomization procedure had failed, the critics' objections seemed unsound.<sup>47</sup>

Critics attached equal or greater importance, however, to a set of issues which were more difficult to articulate and examine within the language and framework of statistical inference. First, they argued that the fixed dosage of tolbutamide used in the study may have been inappropriately high. Their argument was a physiological one, based on experimental and clinical experience with the drug. High initial

---

<sup>46</sup> Jerome Cornfield, "The University Group Diabetes Program. A Further Statistical Analysis of the Mortality Findings," JAMA 217 (September 20, 1971), 1676-1687.

<sup>47</sup> Ibid., 1676. As Paul Meier has pointed out, much of the argument over the UGDP concerned the fundamental question of the purpose of randomization: one ground holding with Bradford Hill that the aim is to make the groups "as nearly equal as possible" and the other, following Fisher, to "provide a firm basis" for the use of the statistical tests customarily applied. Adherence to the former view, Meier suggests, set the UGDP critics on the prowl for variables which might be unevenly distributed in the two groups, neglecting the contention that such baseline variations rarely alter the robustness of the analysis. See Meier, "Statistics and Medical Experimentation," (n. 3) 519. What he neglects to mention is that his colleague Cornfield's account sounds as much or more like Hill's than like Fisher's.

dosages might lead to a large rate of secondary failures, patients whose blood sugar was excessive after several years of treatment because they no longer responded to the drug. If so, the cardiovascular mortality seen might be the result of using tolbutamide inappropriately, rather than an inherent property of the drug itself.<sup>48</sup>

Similarly, critics maintained that the high rate of vascular complications in the study suggested that many of the patients were sicker at the onset than the UGDP investigators assumed. "All they are really saying is that high-risk patients die sooner than low-risk patients." Critics were arguing that many of these patients ought not to have been on oral hypoglycemics in the first place, and that, consequently, the findings were no test of the drug's safety.<sup>49</sup>

The UGDP response was that the fixed dosages and diagnostic criteria employed approximated those used in the majority of clinical practices in the country; by implication, the findings could reasonably be extrapolated to these patients.<sup>50</sup> In short, they read both criticisms as challenges to the study's external validity. What the critics were implying, however, was not that the UGDP had drawn the wrong conclusion but that it had conducted the wrong study.

---

<sup>48</sup> Minutes, Ad Hoc Committee Meeting on UGDP, May 21, 1970, FDA Docket 75N 0062 v. 7.

<sup>49</sup> Robert Bradley, quoted in "The Tolbutamide Debate," Medical World News 12 (January 8, 1971), 38.

<sup>50</sup> Thaddeus Prout, Genell L. Knatterud, Curtis L. Meintert, and Christian Klint, "The UGDP Controversy. Clinical Trials versus Clinical Impressions," Diabetes 21 (October, 1972), 1037-1038.

Methodenstreit

The issue is whether or not medicine is to become a science or remain an art. - Max Miller

The issue is not between medicine as a science or medicine as an art, but whether or not we are to accept at face value questionable conclusions from inadequate data." -Robert Bradley

For most onlookers, the UGDP controversy was concerned with the appropriate treatment of diabetes. To a small handful of participants, the debate raised questions concerning the appropriate role of statistical methods in clinical investigation. Few of the diabetologists with reservations about the UGDP's findings were equipped to challenge the study's distinguished consultants on points of experimental procedure and analysis. But for Yale University's Alvan Feinstein, the UGDP offered an unusual opportunity to voice his reservations about the growing influence of statisticians in epidemiological research. An advocate of the importance of clinical expertise, Feinstein attributed the UGDP's difficulties to an overconfidence in statistical procedures and a neglect of "biologic logic" and "clinical judgement" in designing and interpreting the study. Initially invited to review the UGDP by the manufacturer of tolbutamide, Feinstein apparently found the study an irresistible example of all that was going awry in clinical investigation.<sup>51</sup> The UGDP, he argued, was not so much clinically invalid as clinically irrelevant.

---

<sup>51</sup> In addition to three, somewhat redundant, published papers on the UGDP, Feinstein has used the study as an example in numerous methodological writings. See "An Analytic Appraisal," (n. 42) 167-191; "The Persistent Clinical Failures and Fallacies of the UGDP Study," Clinical Pharmacology and Therapeutics 19 (1976), 78-93; "How Good is the Statistical Evidence Against Oral Hypoglycemic Agents," (n. 5) 71-94.

To be clinically relevant, according to Feinstein, an experiment had to answer two kinds of questions: 1) How are the patients in the study like my patients? How are they different? 2) How are patients' lives affected by the choice of treatment "a over treatment "b"? To be clinically meaningful, the answers to these questions had to be provided in terms the physician could recognize and readily interpret.

The UGDP, in Feinstein's assessment, had failed to record interpretable information about three crucial parameters of diabetes treatment: 1) the initial severity of the disease (including measures of co-morbidity); 2) the rates of non-fatal complications associated with each treatment; and 3) the iatrogenic impact of insulin treatment on patient's comfort and well-being. In the absence of such information, the UGDP was of limited use to clinicians seeking to reassess their current practice, or apply the UGDP findings to their patients.<sup>52</sup>

The UGDP, Feinstein maintained, was a product of its times. Working in the late 1950's, the principal investigators, in an effort to avoid the flaws in experimental design which had compromised earlier controlled studies, had gone overboard. In a misguided quest for objectivity, they had overlooked the very data needed to interpret their findings.<sup>53</sup>

For example, the UGDP had taken great care to define the laboratory data by which eligibility for the study was determined but failed to provide any information about those rejected from the study. How were clinicians able to decide whether their own patients resembled the

---

<sup>52</sup> Feinstein, "An Analytic Appraisal," (n. 42) 169-175.

<sup>53</sup> Feinstein, "An Analytic Appraisal," (n. 42) 175-176, 187-189.

wheat or the chaff?<sup>54</sup> More important, the characterization of those who were accepted was inadequate and incomplete. The UGDP reported in detail on the rates of angina pectoris in the study but what exactly did the investigators mean by angina, and did the operational definition differ from clinic to clinic?<sup>55</sup>

The UGDP's aspirations to methodological rigor were, Feinstein alleged, a tragic flaw. The investigators had decided that, whenever possible, evidence of complications, was to consist of objective measures of pathology and/or functional impairment--physical evidence which could be interpreted by individuals blind to the status of patients, interpretations which could, in turn, be checked by others, to control for the effects of inter-observer variation. Many of these measures were introduced on a wide scale for the first time in this study. For example, to assess nerve conduction, the investigators relied on a biothesiometer, an instrument to measure nerve impulse conduction, rather than on the traditional multi-dimensional clinical assessment of touch, sensitivity to pain and vibration with which most clinicians were familiar.<sup>56</sup> Several of these measures proved difficult

---

<sup>54</sup>. See [Alvin Feinstein], An Analytical Critique of the UGDP Protocols (n. 42), 2-4; Idem., "An Analytic Appraisal," (n. 42) 171. One of the eligibility criteria was "inability" to follow the protocol, but "inability" was not further defined. In a chronic disease like diabetes, where patients must permanently adopt some form of treatment, this information is potentially quite valuable to the clinician trying to extrapolate research studies to his own patients.

<sup>55</sup> Feinstein, "An Analytic Appraisal," (n. 42) 175.

<sup>56</sup> For a description of the UGDP's measures, see UGDP, "I. Design, Methods and Baseline Results," (n. 22) 755-757.

to interpret.<sup>57</sup> Perhaps most important, the UGDP did not systematically collect data on "softer" patient-reported complications which matter a great deal in the choice of management strategies for the diabetic: episodes of 'hypoglycemic' shock, infection rates, reduced episodes of weakness, fatigue, headaches.<sup>58</sup> In summary, the UGDP was not able to provide satisfactory data on many of the intermediate (not-fatal) outcomes of diabetes which are a consideration in choosing therapy.

Ironically, Feinstein shared with the UGDP investigators the belief that the central issue in the debate was the purported benefits of tolbutamide; where they differed was on the question of how well the UGDP had measured those benefits. In forming judgments about the risks of using tolbutamide, Feinstein argued, physicians must consider the potential benefits it offers in comparison to insulin, with its hazards and inconveniences. As a result of efforts to accommodate the methodologists' notions of science, the clinicians in the UGDP had failed to provide the kind of information physicians needed to make therapeutic decisions. For Feinstein, the UGDP was an object lesson in the dangers of sacrificing "clinical wisdom and scientific judgment" to "rigid doctrines of statistical design" or the "conveniences" offered by computerized data processing.<sup>59</sup>

---

<sup>57</sup> UGDP reported difficulties in distinguishing "true" abnormalities from artifacts in their photographic records of eye examinations; the rates of abnormalities reported were unusually high. UGDP, "I. Design, Methods and Baseline Results," (n. 22) 766-767. Apart from one preliminary report, the biothesiometer data was never reported.

<sup>58</sup> Feinstein, "An Analytic Appraisal," (n. 42) 173-174.

<sup>59</sup> Feinstein, An Analytic Critique (n. 42), 25.

In the heat of the immediate debate over the mortality associated with tolbutamide, Feinstein's more fundamental criticisms of the UGDP's approach to clinical investigation went unnoticed.<sup>60</sup> For other critics of the study, the issue became not so much "what does the UGDP suggest I ought to do with my patients" but "should the FDA make suggestions about what I should do with my patients on the basis of the UGDP?" Following the FDA's decision, in the fall of 1970, to alter the package labeling which accompanied tolbutamide, critics of the study precipitated a fifteen year long legal and political campaign to reduce the agency's confidence in the UGDP. Arguments about the scientific merit of the study were consequently conducted against the background of an impending regulatory action.

---

<sup>60</sup> In his 1971 defense of the UGDP, Jerome Cornfield elected to reply only to criticisms which challenged the internal validity of the UGDP. Feinstein's remarks about unknown baseline inequalities were regarded as gnostic utterances which were unintelligible within the framework of statistical inference. See "A Further Statistical Analysis," (n. 46), 1676, 1682.

### III. The Politics of Authority

In June of 1969, the UGDP investigators had notified the FDA of their decision to discontinue use of tolbutamide in the study.<sup>61</sup> So far as the regulatory agency was concerned, the UGDP's reasoning about the drug's lack of benefit was irrelevant. To justify the oral hypoglycemics' claims to efficacy, advocates need not demonstrate that the drugs reduced cardiovascular complications, but merely that they lowered blood sugar. No one was disputing that claim. The regulatory issue was the mortality finding, and the FDA's initial response was reserved. The number of tolbutamide deaths was small, and of borderline statistical significance. Additional numbers would make a more persuasive case.<sup>62</sup> The UGDP investigators stood firm in their initial decision to withdraw the drug.

While the UGDP investigators worked on their report, news of their decision reached physicians at the Joslin Clinic, which had pioneered in the use of the oral hypoglycemics. The Joslin researchers were

---

<sup>61</sup> The UGDP investigators may have notified the FDA as early as November, 1968, of impending problems with the drug. [Summary of the September 1968 UGDP Progress Report and the Resolutions Adopted at UGDP Investigators Meeting of October 4, 1968. Cornfield papers.] By June, 1969, discussions with the FDA about the decision to stop tolbutamide were under way. [see references following note.] The FDA, however, is unable to locate any documents concerning the UGDP prior to May, 1970, and appears to date its awareness of the UGDP from March, 1970. See Federal Register 40 (July 7, 1975), 28587.

<sup>62</sup> Minutes. FDA-UGDP Meeting, June 16, 1969; [LeMar F. Remmert] Diabetes Program Director, Extramural Programs, NIAMD to Director, NIAMD [Donald Whedon], June 18, 1969. Chalmers papers. The FDA's Medical Advisory Board was equally tentative about the UGDP's provisional findings and advised that the FDA await publication of the study results before acting. FDA, Bureau of Medicine, Medical Advisory Board, Minutes. June 26-27, 1969, 6-7. Box 6, Harry Dowling papers, MS C 372. National Library of Medicine.



skeptical of the reputed findings. On May 21, 1970, Henry Simmons, newly appointed head of the FDA's Bureau of Drugs, convened a meeting of the UGDP investigators with their critics.<sup>63</sup> On arriving at the FDA, the visiting diabetologists were greeted by an announcement of the UGDP's findings on the front page of the Washington Post. While their meeting proceeded as scheduled, the FDA's deliberations were cut short by the unexpected publicity.<sup>64</sup> The following day, the agency announced their intention to revise the labeling for the oral hypoglycemic drugs, meanwhile cautioning physicians that the drugs should only be "used only in patients with symptomatic adult onset diabetes mellitus who cannot be adequately controlled by diet alone and who are not insulin dependent (i.e. do not require insulin)."<sup>65</sup>

The FDA's sudden action surprised the UGDP's critics, but it was hardly irrevocable. Through the summer of 1970, the agency continued to negotiate the precise wording of the revised package labeling for the suspect drugs with interested parties.<sup>66</sup> On October 22, 1970, FDA officials announced their intention to caution doctors against using

<sup>63</sup> Present at the meeting were Christian Klint and Jerome Cornfield, statistical experts from the UGDP; Stanley Schor, a statistician critical of the findings; Robert Bradley and George Cahill, from the Joslin Clinic; a group of other prominent diabetologists (T.S. Danowski, Fred Kruger, Albert Winegrad); and Drs. Finkel, Simmons, Ortiz and Anello from the FDA. Minutes, Ad Hoc Committee Meeting on UGDP, May 21, 1970. FDA Docket # 75N-0062, v. 7.

<sup>64</sup> Harry M. Marks, Interview with Robert Bradley, May 20, 1985; The FDA notes of the meeting suggest a "consensus favoring support of conclusions of the UGDP". Minutes, Ad Hoc Committee Meeting on UGDP, May 21, 1970. FDA Docket # 75N-0062, v. 7.

<sup>65</sup> "FDA Statement, May 22, 1970" Diabetes 19 (June 1970) 467.

<sup>66</sup> Memorandum of Conference, July 1, 1970. Records of the Food and Drug Administration.

oral hypoglycemics for mild diabetics, except in cases where insulin is unacceptable and diet does not work. However moderate these guidelines, they did not satisfy the UGDP's critics. Followed by similar recommendations from the AMA and ADA, the FDA's new pronouncement precipitated a crisis for proponents of the oral hypoglycemics.<sup>67</sup> The agency's ruling raised the prospect of malpractice suits for physicians who continued to use the drugs routinely.<sup>68</sup> The critics responded by holding a national press conference, announcing the formation of a Committee on the Care of the Diabetic (CCD) to persuade the FDA to revise its warning notice.<sup>69</sup>

Over the next twenty months (December 1970-July 1972), the CCD and the FDA conducted an intricate war of words over the precise nature of the labeling changes indicated by the UGDP. To FDA officials, the proposed changes in labeling were relatively minor and backed by the weight of scientific and medical authority.<sup>70</sup> Although the CCD's grievances were numerous and varied, their primary concern was that the FDA not unduly "expose" the practicing physician to "medicolegal

---

<sup>67</sup> Harold M. Schmeck, Jr., "FDA Cites Doubt on Diabetes Pill," New York Times October 22, 1970, 1, 54; "Doctors Cautioned by AMA Council About Diabetes Pill," New York Times, October 23, 1970, 17; "New Caution Issued on Drug in Diabetes," New York Times, October 28, 1970, 30.

<sup>68</sup> Robert F. Bradley to Henry Simmons, February 26, 1971. FDA Docket 75N-0062. Vol. 2.

<sup>69</sup> "Forty Diabetes Experts Deplore Warning on Drug by F.D.A.," New York Times, December 2, 1970, 19; "In Boston: A Diabetes Tea Party Hits FDA," Medical World News (December 18, 1970), 13-14

<sup>70</sup> "2 FDA Officials Defend Position As A Consensus," Medical Tribune (November 23, 1970), 1, 20.

redress".<sup>71</sup> Adequate protection, in their view, required the FDA to officially acknowledge the differences of opinion in the medical community concerning the strength of the UGDP study. What may have seemed to CCD representatives a matter of minor consequence posed major difficulties for the regulators.

Shortly before learning of the UGDP case, the FDA had completed a major revision of their policies regarding the scientific evidence acceptable in agency procedures. Eight years after the passage of the 1962 Drug Amendments, the FDA had finally issued regulations describing its requirements for the "well-controlled clinical investigations" which by law were to be the basis of new drug approvals.<sup>72</sup> The new standards had been the subject of a protracted legal struggle regarding the FDA's authority to apply them to drugs, like tolbutamide, which had been approved prior to the 1962 amendments.<sup>73</sup> The criticisms being offered of the UGDP, however reasonable to some diabetologists, did not meet the criteria of "substantial evidence" as defined by the regulations; in the FDA's judgment, "an undiluted and unencumbered warning"

---

<sup>71</sup> Robert F. Bradley to Henry Simmons, February 26, 1971. FDA Docket 75N-0062. Vol. 2. Since the FDA was still drafting the wording of the labelling to be published in the Federal Register, their approach was also in flux, with new initiatives prompting new commentary from the CCD and other onlookers. John Jennings to Charles C. Edwards, March 22, 1971. [On file with Food and Drug Administration].

<sup>72</sup> Federal Register 35 (May 8, 1970), 7250-7053.

<sup>73</sup> The cases, which involved the FDA's right to remove approval from pre-1962 drugs without a hearing unless the manufacturer could produce two studies which met the agency's standards of "substantial evidence," are reviewed in Richard A. Merrill and Peter Barton Hutt, Food and Drug Law. Cases and Materials. (Mineola: The Foundation Press, 1980), 374-375. Earlier efforts to review the status of pre-1962 drugs under the new law had encountered substantial political and legal difficulties.

was "fully warranted by the available evidence."<sup>74</sup> Translated into practical terms, this meant that differences of opinion over the UGDP would not be mentioned in the new labeling.

The FDA's decision to proceed with the labeling was based less on a detailed consideration of the legal situation, however, than on a belief that the weight of scientific and medical authority endorsed the UGDP. While negotiating with the UGDP's critics over the wording of the proposed warning, FDA officials were also seeking to assure themselves that the UGDP's supporters would "not pull the rug out from them regarding the package insert."<sup>75</sup> Among the scientific authorities they consulted were officials at the National Institutes of Health (NIH).

NIH officials had more than a passing interest in the developing controversy over the UGDP. Donald Whedon, director of NIAMD which had funded the study, had also participated in the UGDP's deliberations to discontinue the drug.<sup>76</sup> With an active program of support for clinical trials, senior NIH officials were also concerned about the profession's

---

<sup>74</sup> This is the FDA's paraphrase of a letter from FDA Commissioner Charles Edwards to Neil L. Chayet [Counsel for the CCD], June 5, 1972, as cited in the Federal Register 40 (July 7, 1975), 28588. The full letter appears in U.S. Senate. Select Committee on Small Business. Subcommittee on Monopoly. Hearings on Competitive Problems in the Drug Industry. May 9-10, June 21, and July 1979. Part 22, 8796-8800. 92nd Congress. Second Session.

<sup>75</sup> Thomas C. Chalmers [NIH] to Deputy Director for Science, NIH and Associate Director for Program Planning and Evaluation, NIH, March 23, 1971. [Chalmers papers].

<sup>76</sup> Minutes, Principal Investigators Meeting, June 5-6, 1969. [Chalmers papers]. In response to questioning at this meeting, Whedon had indicated that he thought the decision to discontinue tolbutamide was warranted, and that the investigators could even recommend having the drug taken off the market.

response to the UGDP. When the initial controversy failed to subside, NIH's Director, Robert Marston asked Thomas Chalmers, director of the NIH's Clinical Center, to review the debate.<sup>77</sup>

A physician with an interest in and passion for the use of clinical trials, Chalmers nonetheless produced a balanced assessment of the controversy. In particular, he attempted to distinguish between the investigators' decision to discontinue tolbutamide in the study and the claim "that tolbutamide actually causes an increased death rate." The former was based on the belief that "this study could not possibly have missed a favorable effect" of the oral hypoglycemic drugs. A simple calculation of sample size would demonstrate the reasonableness of this decision. Evaluating the latter claim would require an extensive analysis of the UGDP's procedures and an examination of other studies.<sup>78</sup> Chalmers' carefully drawn distinction was wasted on NIH director Marston, who reported to the FDA "that the conclusions of the study are valid" and that "the technical objections have been satisfactorily rebutted," both by the UGDP's coordinating center and their "consulting biostatisticians". The FDA's "proposed package insert on tolbutamide," Marston reassured Commissioner Edwards, "seems to us to be a fair representation of the state of our knowledge."<sup>79</sup>

---

<sup>77</sup> Max Miller to Jerome Cornfield, March 5, 1971. Chalmers papers.

<sup>78</sup> Thomas C. Chalmers, The Controversy Over Oral Anti-Diabetic Agents [Memorandum for Robert Q. Marston]. March 11, 1971. Chalmers papers.

<sup>79</sup> Robert Q. Marston to Charles C. Edwards, April 5, 1971. Chalmers papers.

Armed with a favorable report from NIH and the other authorities they consulted, FDA officials proceeded to formulate their warning statement on oral hypoglycemics, while the CCD proceeded to the courts.<sup>80</sup> Meanwhile debate among specialists in diabetes over the merits of the UGDP's study continued. To the beleaguered UGDP investigators, it was not the continuation of the debate but its tone which was disturbing. Increasingly, critics seemed "more interested in making in debater's points than in constructive suggestions for further analysis."<sup>81</sup> In the Spring of 1971, NIH officials began looking for an "august" and impartial body to review the UGDP controversy. After exploring several options, they invited the Biometric Society to form a committee to assess "the validity and the conclusions of the UGDP."<sup>82</sup>

---

<sup>80</sup> The FDA sought no further legal test of their authority, but the CCD's lawyers did. In July, 1973, they obtained the first of a series of restraining orders preventing the FDA from issuing new labeling on tolbutamide, and for the rest of the decade, the CCD's lawyers occupied the courts with aspects of the UGDP case. Federal Register 40 (July 7, 1975), 28589. The legal issues in these suits were complex and like the terms of the controversy itself, shifting over time as old ground was lost or new targets of opportunity arose. At some times they rested on issues of administrative procedure and due process; at other times on the scope of the FDA's authority and the evidentiary standards required of it. While litigation effectively prevented the FDA from taking any final action, they did not prevent the agency from entering into new rounds of hearings and negotiations in the interstices between the resolution of one suit and the filing of another.

<sup>81</sup> Jerome Cornfield to Max Miller, January 11, 1971. Cornfield papers.

<sup>82</sup> Before deciding on the Biometric Society, NIH officials considered the Institute of Medicine, the National Research Council, the American Statistical Association as possible forums for reviewing the UGDP. Thomas Chalmers, draft letter [not mailed] to Dr. [Robert] Glaser, Institute of Medicine from [Robert Marston], Director NIH, March 23, 1971; Minutes, July 21, 1971 [Thomas Chalmers, Peter Armitage, Byron Brown, Peter Bennett, Max Miller]; Thomas Chalmers, "Proposal for Review of UGDP," August 6, 1971; Minutes, August 17, 1971 [Thomas Chalmers, Donald Whedon, Peter Armitage, Max Miller]. All materials from Chalmers papers.

An international organization of statisticians specializing in medical research, the Biometric Society seemed to NIH's Chalmers the most appropriate body to review the controversy:

We believe that the basis for any dispute about the conclusions should lie in the consideration of the technical details of design, execution, and statistical analyses of the [UGDP and contrasting] studies. This critique should be carried out by biostatisticians who are experienced [sic] in the field of clinical trials.<sup>83</sup>

The UGDP controversy, NIH officials reminded the statisticians, had implications "much broader than the clinical problem of long-term therapy in diabetes"; it could have a "distinct effect" on the NIH's approach to the planning and funding of comparable studies.<sup>84</sup> After assuring themselves that they would have complete access to data from the UGDP and a free hand in formulating their report, members of the Biometric Society agreed to review the controversy.<sup>85</sup>

NIH's formal charge to the reviewing committee was limited: to evaluate the "scientific quality" and "in particular ... the biometrical aspects" of the UGDP and "other controlled trials of oral hypoglycemic

<sup>83</sup> Thomas Chalmers to Bertold Schneider [President, Biometric Society], September 14, 1971. [Materials relating to the organization and activities of the Biometric Society committee were provided through the courtesy of Professor Marvin Zelen, Harvard School of Public Health, a member of the reviewing panel. Hereafter cited as Zelen papers.]

<sup>84</sup> Thomas Chalmers to Bertold Schneider, September 14, 1971; Minutes, Committee for Assessment of Biometric Aspects of Controlled Trials of Hypoglycemic Agents, May 18, 1972. Zelen papers.

<sup>85</sup> See Marvin Zelen to Berthold Schneider, November 16, 1971; Thomas Chalmers to Peter Armitage, Rodolfo Saracci, Berthold Schneider, Colin White, Marvin Zelen, February 8, 1972. Zelen papers.

drugs."<sup>86</sup> Within the limits of this mandate, the committee's effort was extraordinary: they scrutinized in detail the UGDP operating procedures, extensively reanalyzed the data, and undertook to review not only the published critiques of the study but to interview each of the UGDP's principal critics.<sup>87</sup> But in preparing their report, the review committee operated under constraints of time, money and, most especially, competence.

The panel divided objections to the UGDP into two broad categories: those which might affect physicians' willingness to generalize the UGDP's overall results to their own patients and those which might affect the validity of the UGDP's specific conclusions regarding the possible toxicity of tolbutamide. The former, they decided, were more a matter of medical than statistical expertise. The latter consequently became the principal subject of their review.<sup>88</sup> In assessing the UGDP's conclusions regarding tolbutamide, the committee concentrated on two areas: they carefully examined the UGDP's operating procedures, to assure themselves that the randomization had not broken down; and they re-analyzed the UGDP data to determine whether specific factors mentioned by the critics might singly or by interaction account for the findings. As befitted a blue ribbon panel of experienced statisticians,

---

<sup>86</sup> Robert Q. Marston to Colin White, June 9, 1972, cited in "Report of the Committee for the Assessment of Biometric Aspects of Controlled Trials of Hypoglycemic Agents," JAMA 231 (February 10, 1975), 585.

<sup>87</sup> For a report of the Committee's activities, see Ibid.

<sup>88</sup> See, for example, the committee's remarks on the appropriateness of the UGDP's selection criteria for patients in the trial: "Report of the Committee for the Assessment of Biometric Aspects of Controlled Trials of Hypoglycemic Agents," (n. 86), 591-592, 599.



both efforts were diligent and ingenious, the committee's analyses improving in several respects on the UGDP's original handling of the data.<sup>89</sup> But in translating the critics' objections into the categories of statistical inference— external and internal validity—and into the mathematical procedures necessary to measure the possibility of interaction effects, the committee lost any opportunity to resolve the controversy.<sup>90</sup>

The committee's conclusions were, like the UGDP's original report, circumspect: "On the question of cardiovascular mortality due to tolbutamide...we consider that the UGDP trial has raised suspicions that cannot be dismissed on the basis of other evidence presently available."<sup>91</sup> The thoroughness of their review reinforced this judgment:

We find most of the criticisms leveled against the UGDP findings on this point unpersuasive. [While]... some reservation about the

---

<sup>89</sup> Among the problems handled adroitly by the committee were the role of 1) clinic effects, 2) of baseline cardiovascular risk, and the effects of 3) duration of and 4) adherence to treatment on cardiovascular mortality. The committee's analysis of the first two factors was more thorough and more robust than the UGDP's initial tests, while the latter two issues had not been addressed directly in the UGDP's original analysis.

<sup>90</sup> The committee was abetted in this by the near-inarticulateness of the UGDP's critics when discussing their objections to the study, which focused predominantly on charges that the UGDP had altered its measures of cardiovascular risk midway through the study. One has the impression in reading the critics' subsequent remarks on the Biometric Society's report that they literally did not comprehend how the analyses developed by the committee worked, and to what a considerable degree they addressed the majority of published criticisms of the UGDP.

<sup>91</sup> "Report of the Committee for the Assessment of Biometric Aspects of Controlled Trials of Hypoglycemic Agents," (n. 86), 599.

conclusion that the oral hypoglycemics are toxic must remain... we consider the evidence of harmfulness moderately strong.<sup>92</sup>

To the critics, however, the Biometric Society review failed to address a fundamental issue: whether the UGDP had in the first instance collected the appropriate data to evaluate the risk of cardiovascular disease in patients receiving tolbutamide.<sup>93</sup> In statistical terms, the data being so carefully and thoughtfully analyzed did not contain sufficient information. But none of the critics knew enough about statistics to frame their objections in this way, and none of the statisticians knew enough about medicine to recognize that the specific criticisms being broached might be discussible in such terms.<sup>94</sup> The result was an impasse: to critics, the Biometric Society's report was "predictable" but unconvincing; to proponents, the critics' persistence placed them beyond the pale, among those "more interested in making debater's points than in constructive suggestions" for new analyses and research.<sup>95</sup>

---

<sup>92</sup> Ibid.

<sup>93</sup> Robert F. Bradley, Henry Dolger, Peter H. Forsham, Holbrooke Seltzer, "Settling the UGDP Controversy?" JAMA 232 (May 26, 1975), 813-815; James M. Moss, "The UGDP Scandal and Cover-up," [Letter], JAMA 232 (May 26, 1975), 806-088; Feinstein, "The Persistent Clinical Failures and Fallacies of the UGDP Study," (n. 51) 89-90.

<sup>94</sup> This might not have resolved the debate, of course, since the statisticians might well have demanded that the critics put up or shut up: produce some plausible set of data or live with the modicum of doubt implied by the UGDP. Indeed, their report's concluding appeal to the critics to conduct a controlled study demonstrating the benefits of oral hypoglycemics suggests such a stance. However, had this issue been raised explicitly and intelligibly at the outset, the scope of the committee's entire review might have been differently conceived.

<sup>95</sup> Jerome Cornfield to Max Miller, January 11, 1971. Cornfield papers.

### Knowledge and Interests

The release of the Biometric Society's report marks the point of diminishing returns in the scientific debate over the UGDP.<sup>96</sup> Arguments about the truth of the UGDP's findings gave way to arguments about the motives of those who doubted, or supported, the study. To proponents, physicians' continued use of the oral hypoglycemics was explicable only in psychological or economic terms, by

...the strong desire of both physicians and patients for a way to treat diabetes that does not involve injections, and ...a natural reluctance to accept any possibility that the drugs might be harmful. This [attitude] has been fostered by one-sided presentations of the controversy by one or more of the so-called throwaway medical journals so widely read by physicians.<sup>97</sup>

Given the frequency with which diabetics, regardless of what treatment they receive, die of cardiovascular disease, physicians in private practice could not be expected to notice the marginally increased risk

<sup>96</sup> Neither the study's critics nor the UGDP's supporters seemed willing to sponsor a new trial, although the NIH had considered this in the early stages of the debate. While the study's critics attached considerable importance to the publication of an analysis by Charles Kilo and Joseph Williamson, the article, only followed through on an earlier observation that the placebo group in the UGDP seemed unusually healthy, and that the differences in mortality between placebo and tolbutamide patients were due to the unusual healthiness of the former, rather than the iatrogenic disease of the later. See "The Achilles Heel of the University Group Diabetes Program," JAMA 243 (February 1, 1980), 450-457.

<sup>97</sup> Thomas Chalmers, "Settling the UGDP Controversy," JAMA 231 (February 10, 1975), 624; See also the remarks of John K. Davidson, "The FDA and Hypoglycemic Drugs," JAMA 232 (May 26, 1975), 853 and the testimony offered by Chalmers and Max Miller, the UGDP's study chairman, in U.S. Senate. Select Committee on Small Business. Subcommittee on Monopoly. Hearings ..on Competitive Problems in the Drug Industry, Part 25, 10798-10780, 10801. 93rd Congress. 2nd Session.

associated with the use of tolbutamide.<sup>98</sup> But if the average physician's inability to accept the grim conclusions regarding tolbutamide could be regarded as a failure of imagination, the reluctance of critics to accept the one study large enough to detect such a risk—the UGDP—could have no such benign explanation. In the view of the UGDP supporters, their critics' persistence was due to the corrupting influences or the marketplace on the evaluation process:

...one of the things that this 'controversy' has brought out in the last 5 years, I guess—it seems longer—is the incredible way in which a group of physicians teamed up with industry to attack the only scientific evidence there is on the use of these agents, at a time when we sorely need it.<sup>99</sup>

After five years of controversy, the position taken by the UGDP investigators had hardened. Christian Klimt, the UGDP's statistician, was quite convinced the oral hypoglycemics were "toxic".<sup>100</sup> To his clinical colleagues, the issue at stake was whether "'clinical impression,' anecdotal stories and wishful opinions" or "substantial evidence from adequate, well-controlled investigations" were to be the basis of therapeutic practice.<sup>101</sup> The reluctance of the UGDP's critics, after five years of inconclusive debate over the failings of

<sup>98</sup> See the testimony of Marvin Zelen in U.S. Senate. Select Committee on Small Business. Subcommittee on Monopoly. Hearings ..on Competitive Problems in the Drug Industry, Part 28, 13271. 94th Congress. 1st Session; and Jerome Cornfield, Transcript. FDA Hearings on Oral Hypoglycemic Drugs. August 20, 1975, 176. FDA Docket 75N-0062.

<sup>99</sup> Thaddeus Prout, Transcript. FDA Hearings on Oral Hypoglycemic Drugs. August 20, 1975, 84. FDA Docket 75N-0062.

<sup>100</sup> Testimony of Christian Klimt, U.S. Senate. Select Committee on Small Business. Subcommittee on Monopoly. Hearings ..on Competitive Problems in the Drug Industry, Part 25, 10767. 93rd Congress. 2nd Session.

<sup>101</sup> Max Miller to Hearing Clerk, FDA, August 21, 1975. FDA Docket 75N-0062.

the UGDP, to produce evidence of comparable rigor seemed to the study's advocates silent but eloquent testimony of bad faith. Meanwhile, the UGDP remained, according to its defenders, "the only scientific evidence" on the use of the oral hypoglycemic drugs.<sup>102</sup>

This formulation of the problem recalls that of the early 20th century therapeutic reformers: there are the ignorant, the cupid, and the rational men of science, among whom the UGDP's partisans naturally classed themselves. To at least one of the UGDP's critics, such claims to a monopoly on scientific rationality represented an extreme form of hubris:

Supporting the UGDP contentions are those who believe that a statistical approach to clinical trials, and particularly the assignment of treatment by randomization, is the single most important desideratum in evaluating therapy. On the other side are those who regard randomization as merely a useful component of the evaluation process and who believe that the clinical validity of a clinical trial has primary scientific importance.<sup>103</sup>

But to their other critics, the issue of the best approach to scientific inference from clinical trials was secondary. To the CCD, in particular, the paramount issue was whether "the FDA, the National Institutes of Health, or any other administrative group" had a right to dictate medical practice on the basis of a single, controversial study. Such decisions were best left, in their view, to the profession at large.<sup>104</sup>

---

<sup>102</sup> Thaddeus Prout, Transcript. FDA Hearings on Oral Hypoglycemic Drugs. August 20, 1975, 84. FDA Docket 75N-0062.

<sup>103</sup> Feinstein, "How Good is the Statistical Evidence?" (n. 5), 74.

<sup>104</sup> Bradley, et. al. "Settling the UGDP Controversy?" (n. 93), 813-815.

Despite their animus against medicine "by administrative fiat," the attitude of CCD members towards the practicing physician was ambivalent. Most physicians, they acknowledged, used the oral hypoglycemic drugs poorly and were in need of better advice to use them more appropriately. Such guidance, however, should come from the profession's self-constituted authorities, the specialists in diabetes, and in the form of education, not instruction. In ruling that the issue was not "who decides," but the nature of the evidence on which decisions are made, the FDA sided with the UGDP. Yet the agency's decision was far less radical than its critics implied. For the physician who continued to use the drugs inappropriately, the FDA had no sanctions in mind.<sup>105</sup>

If the CCD's opinion of the practicing physician was ambivalent, so was that of the UGDP proponents towards the FDA. Initially, the UGDP investigators stood aloof from the FDA proceedings, distinguishing their limited conclusions about tolbutamide from the FDA's more general indictment of the remaining hypoglycemic drugs. As controversy over the UGDP developed, however, they came to see the FDA ruling as vindicating their study, and any hesitations on the FDA's part as a sign that the agency was confusing "evidence with influence."<sup>106</sup> The FDA's ruling had become a measure of the scientific community's judgment. Either the agency issued a warning to physicians, or it did

---

<sup>105</sup> See the testimony of Alexander Schmidt, FDA Commissioner, in U.S. Senate. Select Committee on Small Business. Subcommittee on Monopoly. Hearings ..on Competitive Problems in the Drug Industry, Part 28, 13307. 94th Congress. 1st Session.

<sup>106</sup> Thaddeus Prout to Julian Santangelo [FDA], undated. Chalmers papers.

not; the idea that the FDA might have issued a warning and acknowledged uncertainty about the study met with studied incomprehension.<sup>107</sup>

### Social Decisions and Private Judgments

Any conclusions to be drawn from the UGDP experience are potentially as controversial as the study itself. To the UGDP's advocates, the controversy demonstrated the ability of a vocal minority to undermine confidence in the conclusions of a responsibly conducted study, and to frustrate the FDA's efforts to protect the health of diabetic patients. To the study's critics, the controversy represented the usurpation of clinical expertise and physician autonomy in favor of statisticians and the dictates of bureaucrats. Certainly, the disputant's political and intellectual commitments made it difficult for either to concede, or at times even to comprehend, the arguments of the other. This inability to reach agreement only reinforced each party's convictions that greater issues than the management of diabetes were at stake: the intellectual integrity of the evaluation process for the UGDP's advocates, the clinical autonomy of the physician for the study's

---

<sup>107</sup> A measure of the situation may be taken from the colloquies between Paul Meier, a statistician member of the Biometric Society review panel and Senator Gaylord Nelson. Although not focused around any specific regulatory action, these exchanges are instructive: Meier's insistence that the UGDP was well conceived and executed and that its conclusions were uncertain was regarded by Nelson either as unintelligible or as typical of scientific agnosticism, the view that no study is good enough for making policy, and a stance which Meier explicitly repudiated. Yet in the context, the idea that someone could affirm the UGDP's scientific quality and explicitly acknowledge the uncertainty of its conclusions was literally incomprehensible. See U.S. Senate. Select Committee on Small Business. Subcommittee on Monopoly. Hearings ..on Competitive Problems in the Drug Industry, Part 28, 13261-13263, 13267-13271. 94th Congress. 1st Session.

critics. Yet merely to observe that ideological commitments played a role is not sufficient: the question remains, how did they gain the upper hand?

The polarization of the debate has much to do with the institutional context in which the controversy developed. Once news of the UGDP findings was made public, the customary process of seeking consensus through peer review would not work. The advisory process depends on the ability to negotiate compromises in private; concessions are easier to make when they are not publicly acknowledged as such. But the CCD began the regulatory debate by insisting on public redress for damages which, in their view, had already been done; the UGDP representatives came, over time, to much the same position. Only a public act of vindication by the FDA would begin to repair the harm done by the lengthy delays in the re-labeling of the drugs. To give up the effort to relabel the drugs would be to leave the matter to the judgments of individual physicians who, as even the CCD acknowledged, showed few signs of using the drugs responsibly.

The UGDP's proponents began with the presumption that both the study's results and its implications for practice would be deliberated among individuals who shared a common understanding of the issues, but who might nonetheless reach different conclusions. As the controversy progressed, they discovered that the UGDP's critics disagreed not only with the study's specific conclusions but with the intellectual framework in which these issues were being discussed. In accepting representatives of the Biometric Society as the most authoritative body to adjudicate the dispute, they only confirmed the critics' view of the



decision making process as partial and coercive, a judgment which has nothing to do with the reasonableness of that body's inferences per se.

This statement of the problem suggests that the choice of forum for interpreting the UGDP's conclusions was as much at issue as the validity of the conclusions themselves. But to address this complaint fully would have required both parties to examine the merits of various institutional arrangements for translating the results of experimental findings into clinical practice. Such an examination, calling for statisticians and clinicians alike to move away from the comfortable grounds provided by their scientific expertise towards the uncharted territories of politics, might well have proved difficult. The UGDP's critics would have had to defend their unexamined partiality for the intellectual autonomy of physicians, in the face of their own admission that many physicians manage diabetes poorly. For the UGDP's defenders, it would have meant exploring the ambivalence within the statistical tradition towards such autonomy.

Statistical concepts were introduced into medicine with the expectation that they would not only improve the quality of therapeutic experimentation, but would also contribute to improving therapeutic practice. Yet formal accounts of the properties of different experimental designs and statistical tests do not, and cannot, articulate the means by which the results of well conducted studies should reform therapeutics. Associated with the identical conceptual apparatus are two widely divergent visions of the social means for bringing rational order to contemporary therapeutic practice.

On one account, members of the medical community, accepting the intellectual discipline of statistics as two generations before they had assimilated that of chemistry, adopt a more rigorous approach to evaluating evidence. Formal aspects of experimental design—randomization, power, alpha and beta levels—become incorporated into medicine along with other criteria. Medical practice improves, on this view, through improvements in methodology which elevate the standards by which evidence is judged. Once physicians learn to reason statistically, procedures which are found wanting by these more rigorous standards of evidence would no longer be accepted by the medical community. An alternative view, equally intent on elevating the standards of therapeutic evaluation, proposes to bypass the step of reforming the judges by putting RCTs in place as the gateway through which all practices must pass before being admitted to clinical practice. The problem is thereby reduced to a simple decision: which therapies do or do not pass the test? Individuals who do not keep pace with the scientific standards being promulgated must nevertheless fall in step.

Each model represents virtues lacking in the other. The first is tolerant, respectful of differences in values and in judgments, but it offers no mechanisms for recognizing intellectual deviance and accepts a slow and uneven rate of change. The second paradigm has the potential to bring about quicker and more uniform changes in practices, but may find it more difficult to distinguish deviance from honest dissent. The choice between them is fundamentally an extra-scientific problem. The task of the next chapter is to examine what a political account of the problem can contribute that the sciences of statistics and medicine alone cannot.

Chapter Seven: Science and Politics

"...though the knowledge science has to offer is always more than technical knowledge, what it has to offer to politics is never more than a technique." <sup>1</sup>

In the two preceding chapters, I have examined the introduction of statistical methods into therapeutic research, and the conflicts which arose in one particular instance of applying these methods. In the present chapter, I want to consider some recent proposals that randomized experiments (RCTs) should play a more central role in shaping health policy towards medical practice.

Over the past fifteen years, numerous academic commentators have proposed that RCTs be used to evaluate new medical technologies, surgical procedures and other innovations in medical practice.<sup>2</sup> Recently, economists have joined an existing chorus of statisticians and therapeutic reformers in suggesting that controlled experiments might provide a scientific basis for allocating medical resources.<sup>3</sup> To call such position papers 'policy proposals' stretches the term somewhat. Apart from appeals for increased funding of more rigorous technology assess-

<sup>1</sup> Michael Oakeshott, "Rationalism in Politics," in his Rationalism in Politics and Other Essays (New York: Methuen, 1981), 22.

<sup>2</sup> Drugs are already subject to such restrictions, under the 1962 Drug Amendments.

<sup>3</sup> For advocacy of this position, see Jean Pierre Boissel, "Clinical Trials: Aims and Four Puzzling Points," in J.P. Boissel and C.R. Klint, eds. Essais Controles Multicentres. Principes et Problems (Paris: Institute Nationale de la Sante et de la Recherche en Medicine, 1979), 21-28; A.L. Cochrane, Effectiveness and Efficiency. Random Reflections on Health Services (Abingdon: Nuffield Provincial Hospitals Trust, 1971); M.F. Drummond and G.L. Stoddard, Economic Analysis and Clinical Trials. Quantitative Studies in Economics and Population. Research Report No. 34. (Hamilton: McMaster University, 1982).

ments, the institutional details of such proposals remain quite vague. Exactly who is to decide which experiments are to be conducted, and how the results are to be used to direct "policy" goes unstated.<sup>4</sup> Apart from legislation requiring that manufacturers demonstrate the efficacy of new drugs "by means of adequate and well-controlled investigations," federal programs have yet to define the precise role of RCTs in regulating medical practice.<sup>5</sup>

That RCTs do not yet play a decisive role in determining our policies toward medical practices and technologies does not mean that they are incapable of doing so. Nor should the lack of institutional detail in current policy proposals regarding RCTs deter us from considering their merits. But the vagueness of such proposals suggests that we begin by examining the more general claim: that the methodology of controlled experiments offers a 'scientific' basis for allocating medical resources.

In the final analysis, I will argue, the political issues raised by such proposals are crucial to our assessment of any proposed reforms. But first I want to examine the rationale offered by those I have

---

<sup>4</sup> On the planning of clinical trials, see James P. Kahan, C.R. Neu, Glenn T. Hammons and Bruce J. Hillman, The Decision to Initiate Clinical Trials of Current Medical Practices R-3289-NCHSR. (Santa Monica: Rand Corporation, 1985).

<sup>5</sup> The language is that of the 1962 Drug Amendments, as cited in Peter Temin, Taking Your Medicine. Drug Regulation in the United States (Cambridge: Harvard University Press, 1980), 127. To date, the promise that controlled technology assessments direct the policies of "third-party payors" and federal health insurance programs appears to be honored largely in the breach. For a detailed analysis, see Stan N. Finkelstein, Kenneth A. Issacson, and John J. Frishkopf, "The Process of Evaluating Medical Technologies for Third Party Coverage," Journal of Health Care Technology 1 (Fall, 1984), 89-102.

characterized as "methodological reformers," whose arguments regarding the superiority of RCTs are both logically and historically prior to any specific policy proposals. Ultimately, their's is an intellectual brief and the claims they have on our attention rest on what we end up thinking about the adequacy of their ideas regarding the relation between therapeutic experimentation and therapeutic reform.<sup>6</sup>

At bottom, arguments that RCTs should play a central role in the evaluation of medical practices and in the formulation of health policy towards those practices rest on the same ground: the strength of the inferences which can be drawn from well designed, well conducted RCTs. Despite its costs and inconvenience, "...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 RCTs."<sup>7</sup> On the face of it, such an argument seems intuitively plausible. Therapeutic practices ought to be based on scientific considerations, and if the conclusions from RCTs are inherently more reliable than

---

<sup>6</sup> The approach taken here is premised on the notion that theoretical reflection on social institutions begins with an understanding of the premises and purposes of those institutions, as articulated by those who participate in them. Rather than burden this chapter with an overlong theoretical preface justifying this approach, I will simply refer the reader to those who have put the case for it much better than I. See Charles Taylor's essays on the social sciences, collected in his Philosophy and the Human Sciences. Philosophical Papers. Volume 2. (Cambridge: Cambridge University Press, 1985), and Alisdair MacIntyre's After Virtue (Notre Dame: University of Notre Dame Press, 1981). Robert E. Goodin's Political Theory and Public Policy (Chicago: University of Chicago Press, 1982), while it favors a more utilitarian line of reasoning than I do here, makes an explicit case for employing normative analysis of this sort to the sorts of problems which currently go under the rubric of "public policy".

<sup>7</sup> John C. Bailar III, "Introduction," in Stanley H. Shapiro and Thomas A. Louis, eds. Clinical Trials. Issues and Approaches (New York: Marcel Dekker, 1983), 1.

those from any other source, then they deserve special consideration, both in medical practice and health policy.

As a matter of formal logic, it is difficult to quarrel with proponents of RCTs. Randomization not only makes it more likely that an experiment will distinguish the effects of treatment from independent variations in the conditions of patients; perhaps more importantly, it provides a basis for determining how likely it is that the treatment is as good (or bad) as an experiment indicates.<sup>8</sup> These are powerful arguments. Yet a moment's reflection should give pause. Why should method per se play such a key role in therapeutic research? Elsewhere in science, an experiment's formal validity is at best a secondary consideration. What matters is the congruence between an experiment's findings and a body of other evidence, theoretical and empirical. If methods are considered at all, it is the substantive aspects of method which matter: which reagents and buffers were used? what kind of a detection device? how were the cultures plated? were there dramatic shifts in ambient temperature during the study? and so on.<sup>9</sup>

---

<sup>8</sup> The first of these views is emphasized by Sonja McKinley, "Experimentation in Human Populations," Milbank Memorial Fund Quarterly 59 (1981), 308-323; the second by David Byar in "The Necessity and Justification of Randomized Clinical Trials," in H.J. Tagnon and M.J. Staquet, eds. Controversies in Cancer. Design of Trials and Treatment (New York: Masson Publishing Co., 1979), 31-53. The importance of recognizing the distinction between the two is stressed in Paul Meier's interesting discussion "Statistics and Medical Experimentation," Biometrics 31 (1981), 339-345.

<sup>9</sup> Apart from gross errors of logic, it is difficult to recall an instance in the history, as opposed to the philosophy of science, where formal method matters as much as it seems to in the application of statistics to medicine. For a useful account of how scientists do criticize details of experimental procedure, see Michael Mulkey and Nigel Gilbert, "Accounting for Error: How Scientists Construct Their World View When They Account for Correct and Incorrect Belief," Sociology 16 (1982), 165-193. In practice, as opposed to formal theory, good statisticians often excel in identifying substantive details of experimental procedure which might affect conclusions. See the discussion of rat cage placement in Stephen Lakagos and Frederick Mosteller, "A Case Study of Statistics in the Regulatory Process. The FD & C Red Dye No. 40 Experiment," Journal of the National Cancer Institute 66 (January, 1981), 197-212.

The link to policy is similarly problematic. In any other policy arena, the first question might be: what is the problem to which RCTs are a solution? The operative assumption has been that the policy problem is to determine which medical practices are effective and which are not. But this statement of the problem comes, largely unexamined, from the community of therapeutic reformers. While it accurately reflects the viewpoint of many physicians in academic medicine, it may not be as persuasive to other concerned parties: consumers, providers of insurance or regulatory authorities.

Since the current rationale for using RCTs is based principally on the claim that they provide a scientific basis for judging therapeutic practices, I want to begin by examining the relation between method and science implicit in discussions of RCTs. What is the view of science put forth by proponents of RCTs, and how compelling are the justifications they offer for basing scientific judgments and policy decisions on randomized experiments? In the second half of this chapter, I want to explore in greater detail the institutional presuppositions of policy arguments about RCTs: what exactly are we committing ourselves to when we endorse one or another view of experimental inference as a means for establishing public policy?

I: Method and Science

If the adequacy of a method is not measured by its usefulness to the purpose of science, if on the contrary the use of a method is made the criterion of science, then the meaning of science as a truthful account of the structure of reality, as the theoretical orientation of man in his world, and as the great instrument for man's understanding of his own position in the universe is lost.<sup>10</sup>

The sciences have retained one characteristic of philosophy: the illusion of pure theory. This illusion does not determine the practice of scientific research but only its self-understanding.<sup>11</sup>

Most advocates of RCTs embrace a purely methodological theory of science. By a purely methodological theory, I mean one in which a demarcation line is drawn between the scientific and the non-scientific, with the use of the designated method constituting the measure of whether a particular piece of work falls on one side of the line or another. Unlike some such theories—e.g. logical positivism—the philosophical rationale for randomized experiments has not been fully articulated by its proponents. One must therefore rely on a process of historical and logical reconstruction to establish the argument being set forth.<sup>12</sup>

---

<sup>10</sup> Eric Voegelin [1952], The New Science of Politics. An Introduction (Chicago: University of Chicago Press, 1983), 5.

<sup>11</sup> Jurgen Habermas, Knowledge and Human Interests (Boston: Beacon Press, 1971), 315.

<sup>12</sup> To proponents of RCTs who protest at being fitted in this Procrustean bed, I suggest they make clearer, and more explicit, the philosophical basis for their advocacy of clinical trials, as some among them have done. [See the brief discussion in Meier, "Statistics and Medical Experimentation," (n. 7).] If the present work stimulates a few outraged citizens in the biostatistical community to correct the following exposition, and in the process, clarify their own stance, it will have accomplished enough.



The principal claim made for RCTs is that randomized experiments offer a superior method for measuring therapeutic effects. The case for their methodological superiority rests partially on empirical and partially on theoretical grounds. The empirical argument is one which comments on the defects of non-randomized studies. Having long suspected researchers in non-randomized studies of bias in the assignment of treatments, advocates of RCTs have recently begun amassing evidence that a serious imbalance can be found in many non-randomized clinical trials, with the healthier patients being selected for the experimental treatment.<sup>13</sup> Accordingly, they argue, the favorable view such studies produce of innovative treatments cannot be trusted. Yet as proponents of RCTs would be the first to admit, the past performance of any treatment offers an imperfect guide to its current merits. By analogy, the past defects of non-randomized studies should not necessarily be taken as an indictment of their present or future value. While documenting the flaws of particular non-randomized studies provides some comfort to proponents of RCTs, it ultimately provides no greater weight than the observation that particular randomized studies were flawed in design or execution. Ultimately, the case for the methodological superiority of RCTs is a theoretical one.

---

<sup>13</sup> See, for example, Henry Sacks, Thomas C. Chalmers, Harry Smith, Jr., "Randomized versus Historical Controls for Clinical Trials," American Journal of Medicine 72 (February, 1982), 233-240, and the broader discussion of selection biases in "David L. Sackett, "Bias in Analytic Research," Journal of Chronic Disease 32 (1979), 51-63. Chalmers' work, which focuses on the ways in which doctors in case control studies bias these studies by picking the winners raises the following question: if doctors are so good at guessing who will benefit from treatments, why do we need RCTs in the first place?

As we saw in chapter five, statisticians have offered two distinct theoretical rationales for randomization. The initial, and strongest case, was put by R.A. Fisher: "the physical act of randomization is necessary for the validity of any test of significance."<sup>14</sup> In any experiment where the treatments are randomized, the researcher has a measure of how likely it is that the results are due to chance. The value of Fisher's approach is twofold: one, it provides a logically valid basis for experimental inference; two, it gives the experimenter an intelligible and precise benchmark for assessing the likely truth of his findings. To anyone interested in placing experimental inference on a more logical and trans-subjective basis, the merits of Fisher's approach are obvious.

What Fisher's theory does not address are the extra-logical and non-statistical considerations involved in experimental inference. First, on what basis do we chose a significance level? Do we want to take the chance of being wrong one in twenty times or one in a hundred? There is nothing in the formal theory of probability dictating a unique answer to this question.<sup>15</sup> Second, on what basis do we integrate the findings from a given experiment with the relevant body of theoretical and empirical knowledge? Suppose, for example, agronomists find that nitrogen additives had less effect than had been expected in their

---

<sup>14</sup> R.A. Fisher, The Design of Experiments (London: Oliver and Boyd, 1935), 51.

<sup>15</sup> Michael Cowles and Caroline Davis, "On the Origins of the .05 Level of Statistical Significance," American Psychologist 37 (May, 1982), 553-558.

experimental crops. Do we conclude that such additives offer less benefit than had previously been thought or, as some scientists believe, that the concentrations of nitrogen used in the experiment were sub optimal? For Fisher, that is a question which requires a different experiment—one planned to compare the effects of nitrogen at various concentrations. At any rate, such speculations have no immediate bearing on the inferences to be drawn from the experiment presently under discussion. <sup>16</sup>

But what bearing do these niceties of logic and statistical theory have on the status of clinical trials? A good deal, I wish to argue. For once one abandons the claim that randomized experiments provide a "uniquely superior" means of "quantifying uncertainty", then both the premises on which a plausible defense of RCTs rests and the role we assign them in achieving a rational therapeutics shifts. If statistical theory cannot dictate a decision rule for interpreting the results of therapeutic experiments, then a different, more fallible, source of intellectual authority must be found.

---

<sup>16</sup> While statisticians have been active in devising experimental designs which can incorporate provisional findings from a clinical trial into the subsequent rates of patient assignment within that same trial, they have somewhat less to say about how to handle information from other studies going on at the same time. The question is hardly academic. During the 1970's, NHLBI sponsored a number of parallel trials studying the effects of aspirin on myocardial infarction. Some individuals in the coordinating statistical office had access to information about the provisional findings of both studies. While it is certainly believable that they kept the findings of study a in confidence when discussing the course of study b, it is somewhat less credible that they were able to ignore their own knowledge of the parallel study when formulating their judgments about the future course of the complementary experiment.

At present, there are two views of the place RCTs should have in therapeutic research. The first view is dependent on what I have characterized as a methodological theory of science. Its proponents regard RCTs as a gateway through which all new innovations must pass before being admitted to clinical practice. Theories about how a drug works to intervene in a disease or pharmacological studies of drug action are secondary to a direct test of therapeutic results. The truth is found in what works.<sup>17</sup> RCTs, on this account, are simply a technology by which other technologies are to be judged. Appealing as this stance is to reformers who think current progress in improving therapeutics is too slow, by elevating RCTs to a standard above criticism it does as much to retard rational progress in therapeutics as to advance that cause.

Consider the case of a thoughtfully planned and conducted RCT whose design has been carefully reviewed by a panel of expert clinicians and statisticians. The study over, the investigators find that the new drug is somewhat less effective than had been hoped in preventing recurrences of a specific disease. What is then a reasonable response to critics who, skeptical of the study's findings, disagree with the investigators' definition of the eligible population, the details of the therapeutic regimen, the indices of prognosis employed, the measures of outcome chosen, and so on? In the context of a pluralistic

---

<sup>17</sup> Some authors point to the emphasis on clinical outcomes as an important virtue of RCTs, one which prevents evaluations of medical treatments from being perturbed by changing scientific fashions in pathophysiology and pharmacology: "...the controlled trial produces the same answer irrespective of hypothetical mechanisms." See David Spodick, "The Randomized Controlled Clinical Trial," American Journal of Medicine 73 (1982), 421.

scientific community, these differences of opinion are accommodated through debate and discussion about the suitability of the means (the experiment) to the ends at hand (the clinical practice). To treat such a study instead as binding is to foreclose any such discussion.

A purely methodological theory is no less limited when the planning rather than the interpretation of experiments is under consideration. One difficulty is the inability of such theories to incorporate practical and policy considerations alongside discussions of experimental method. The tendency is to distinguish between the theoretical considerations which dictate a reliance on 'appropriate method,' and 'practical' matters of cost, convenience, and feasibility.<sup>18</sup> While few advocates of RCTs would argue that these practical considerations are negligible, they have no proper place in evaluating what path science ought to take, as opposed to their uncomplimentary role in dictating compromise with scientific ideals. In effect, such counsels of perfection discourage any systematic thinking about the relation between the best and the actual.<sup>19</sup>

The alternative position, despite its espousal of randomized experimental designs, does not place its faith in method alone. The

---

<sup>18</sup> On the separation of theoretical and practical reasoning in modern theories of science more generally, see Jurgen Habermas, "On Theory and Practice in Our Scientific Civilization," in his Theory and Practice (Boston: Beacon Press, 1973), 253-282. Habermas' discussion of practical reasoning encompasses far more than a consideration of instrumental elements such as cost and feasibility. For more on this point, see part II, below.

<sup>19</sup> For some recent systematic approaches to experimental design, see Frederick Mosteller and Milton C. Weinstein, "Toward Evaluating the Cost-Effectiveness of Medical and Social Experiments," in Jerry A. Hausman and David A. Wise, Social Experimentation (Chicago: University of Chicago Press, 1985) 221-249.

scientific community provides the missing link between the formal logic governing the application of statistical analysis to experimental data and the evaluation of individual studies. Once physicians learn to reason statistically, it is hoped, procedures which are found wanting by these more rigorous standards of evidence would no longer be accepted by the medical community.<sup>20</sup> This is a theory of progress, but one in which progress is measured by judging the actions and choices of individuals and institutions, not purely by establishing accordance with a preset methodological standard.

Such an account has much to recommend it. In rejecting a purely methodological criterion of science, it allows for the possibility of rational disagreement about the adequacy of a given study, as distinct from irrational opposition to its conclusions. It explicitly acknowledges that measures of therapeutic value are plural, not singular, and that any one study may fail to capture important dimensions of therapeutic effect. And it opens the door to a more pluralistic discussion of the aims of medical science and the best means for meeting them. To those made uncomfortable by dogmatism, the appeal of an explicitly anti-dogmatic view of statistical methodology may be obvious. But before we wholeheartedly embrace pluralism, let us reconsider the virtues of methodological dogmatism.

What methodological dogmatism has to offer is a clear cut criterion for distinguishing between good and bad decisions: is the decision based on evidence from an RCT? The individual adopting a more multi-

---

<sup>20</sup> Donald Mainland, "Medical Statistics-Thinking Vs Arithmetic," Journal of Chronic Disease 35 (1982), 413-417.

dimensional view of science has no such single, compelling standard to offer. In the case of the UGDP, for example, the strongest case a statistician adopting an anti-dogmatic stance could make was that the study had shifted the burden of proof regarding the merits of tolbutamide, a thoughtful view, but not one conducive to resolving therapeutic controversies.<sup>21</sup> Does the dogmatic position have anything to offer other than the convenience of a rule for making decisions?

One conceivable rationale for the dogmatic position rests on what I have characterized as the non-Fisherian argument for randomization. There are numerous factors, other than therapy, which might influence the outcome of a study. Only randomization can ensure that any important factors, known or unknown, have been equally balanced between the group receiving treatment and the control group. Might this argument provide a justification for the universal use of randomization in therapeutic experiments?

When faced with critics of an individual study who allege that, in this particular instance, the randomization "didn't work", advocates of this position have seemed little better off than exponents of Fisher's view. At best they have urged that critics empirically examine the groups to test for significant differences; at worst, they have haughtily asserted that critics "misunderstand" the functions of randomization. But they might conceivably reply: yes, it is possible we are wrong in this instance. But in the long run, we will be right a

---

<sup>21</sup> See the testimony of Paul Meier in U.S. Senate. Select Committee on Small Business. Subcommittee on Monopoly. Hearings ..on Competitive Problems in the Drug Industry, Part 28, 13261-13263, 13267-13271. 94th Congress. 1st Session.

lot more often if we randomize patients all the time, and abide by the conclusions of those studies, because in general RCTs are less likely to mistakenly credit innovative treatments with making patients better. We cannot offer a foolproof method for avoiding mistaken inferences but we can offer you a reliable procedure for making fewer mistakes overall. If we are to realize this long run benefit, however, we must acknowledge RCTs as a basic standard in therapeutic evaluations.

To the best of my knowledge, no one has attempted to argue such a utilitarian case for employing randomized studies as a minimal, universal standard for therapeutic evaluations.<sup>22</sup> Given the elective affinities between utilitarianism and the espousal of controlled therapeutic experiments, the lack of a fully articulated rationale on

---

<sup>22</sup> The closest to a formal philosophical discussion of these issues I know of is Lockwood's and Anscombe's criticism of the utilitarian rationale for urging participation in such studies, which does not quite fit the bill; moreover, they seem to underestimate the force of arguments that the answers one gets from RCTs are qualitatively superior to the answers one gets from other studies, a claim which needs to be addressed in deciding whether one study design is more or less ethical than another. See Michael Lockwood and G.E.M. Anscombe, "Sins of Omission? The Non-Treatment of Controls in Clinical Trials," Proceedings of the Aristotelian Society (1983), 207-222.

C.S. Peirce, as usual a half-century ahead of everyone else, does sketch out the utilitarian view of scientific inference: "That which determines us, from given premises, to draw one inference rather than another is some habit of mind, whether it be constitutional or acquired.- The habit is good or otherwise, according as it produces true conclusions from true premises or not; and an inference is regarded as valid or not, without reference to the truth or falsity of its conclusion specially, but according as the habit which determines it is such as to produce true conclusions in general or not." C.S. Peirce, "The Fixation of Belief," in Justus Buchler, ed., Philosophical Writings of Peirce (N.Y. Dover Publications, 1955), 8.



utilitarian premises is somewhat surprising.<sup>23</sup> My own view is that ultimately the logic of a utilitarian argument would undermine the rationale for a purely methodological defense of RCTs, as criteria of feasibility, cost, and effectiveness came to be applied to randomized experiments themselves.<sup>24</sup> But whatever the eventual inadequacy of such a justification, it at least points in the right direction. In the case of medicine, the rationale for adoption of randomized experiments depends as much on social and political questions as on matters of epistemology and the philosophy of science.

With this contention in mind, I want to re-examine the arguments over randomization on a different basis, one which emphasizes the institutional relations between therapeutic experimentation and therapeutic reform.

## II: Methods and Institutions

For much of this century, a small medical elite has maintained that the evidence on which the majority of physicians base their therapeutic practices is woefully inadequate. Therapeutic reformers introduced statistical concepts and procedures into medicine in the belief that, by improving the quality of the evidence on which therapeutic evaluations were made, these methods would improve physicians' therapeutic deci-

---

<sup>23</sup> On the link between utilitarian frameworks and statistical methodologies, see Martin S. Pernick, "A Calculus of Suffering. Pain Professionalism and Anesthesia in Nineteenth Century America (New York: Columbia University Press, 1985) 93-124.

<sup>24</sup> Even narrowly construed, the utilitarian defense has nothing to say about the basis for a decision in the case where multiple RCTs arrive at conflicting conclusions, as in a number of trials dealing with the prevention of secondary myocardial infarction.

sions. Public policy in the area of drug regulation and medical technology assessment has taken its lead from these reformers: better evaluations mean better practices.

The claim that RCTs contribute to improving medical practice by improving medical science is central to the arguments of contemporary therapeutic reformers. One way of evaluating this claim seems straightforward enough: do evaluations conducted according to the canons of statistical methodology influence physicians' therapeutic decisions? The evidence here is both limited and equivocal. Thus, several studies report that physicians have adopted (or abandoned) therapeutic practices in a manner consistent with the findings of controlled experiments. In at least some of these instances, however, adoption (or abandonment) of the practice preceded publication of the critical studies.<sup>25</sup> In other instances, researchers report a substantial lag between publication of the reference studies and changes in practice, leaving it up to the individual authors to decide whether these cases argue for or against the hypothesis.<sup>26</sup>

---

<sup>25</sup> Thus physicians began treating mild hypertension aggressively before results of a clinical trial advocating that practice were available, and abandoned the use of "gastric freezing" apparatus to treat ulcers before definitive trials indicating its lack of efficacy were available. On hypertension, see Sally Guttmacher, Michael Teitelman, Georganne Chapin, Gail Garbowski and Peter Schnall, "Ethics and Preventive Medicine: The Case of Borderline Hypertension," Hastings Center Report (February, 1981), 15-17. On gastric freezing, see H.V. Fineberg, "Gastric Freezing: A Study of the Diffusion of a Medical Innovation," in Institute of Medicine, Medical Technology and the Health Care System (Washington: National Academy of Sciences, 1979), 173-200.

<sup>26</sup> Contrast the conclusions on the impact of the University Group Diabetes Program study offered by Thomas C. Chalmers, "The Impact of Controlled Trials on the Practice of Medicine," Mount Sinai Journal of Medicine 41 (1974), 753-9; and Stan N. Finkelstein, Stephen B. Schechtman, Edward J. Sondik and Dana Gilbert, "Clinical Trials and Established Medical Practice: Two Examples," in Edward B. Roberts, et. al. Biomedical Innovation (Cambridge: MIT Press, 1981), 200-215.

Certainly, in interpreting such evaluation studies it would be helpful if we knew not only whether or not physicians used the drug or technology in question, but why they chose to practice in this way. Yet few evaluation reports provide independent data on whether physicians knew of the reference studies whose influence is being measured, and even fewer ask whether or not physicians agreed with the findings.<sup>27</sup> Influence, or the lack thereof, is more often presumed than established. More imagination is certainly needed in designing studies which would illuminate the complex interplay of knowledge and circumstance in shaping physicians' therapeutic decisions.<sup>28</sup> But even improved studies of therapeutic choice would be unable to tell us how to regard the claim that RCTs provide an authoritative basis for formulating our public policies toward medical practices.

One answer to the question has the merits of simplicity: RCTs are authoritative because they are conducted by authorities. To the social scientist steeped in the culture of relativism, such a

---

<sup>27</sup> Most of the studies surveyed by the Institute of Medicine's Committee for Evaluating Medical Technologies rely on time trends to study the effects of clinical trials on subsequent practice. See Harvey V. Fineberg, "Effects of Clinical Evaluation on the Diffusion of Medical Technology," in Institute of Medicine, Committee for Evaluating Medical Technologies in Clinical Use, Assessing Medical Technologies (Washington: National Academy Press, 1985), 187-194.

<sup>28</sup> The report by Fineberg and his colleagues that anesthesiologists were aware of and agreed with the need for certain precautions in the use of specific anesthetic technologies, yet did not personally employ these precautions deserves more examination. See H.V. Fineberg, R. Gabel and M. Sosman, "The Acquisition and Application of New Knowledge by Anesthesiologists: Three Recent Examples," Anesthesiology 48 (1978), 430-438. The research strategy followed by Arabella Melville and Roy Mapes on physicians' awareness of, and responses to, warnings regarding practolol suggests one line of attack on this problem. See their "Anatomy of a Disaster: The Case of Practolol," in Roy Mapes, ed. Prescribing Practice and Drug Usage (London: Croom Helm, 1980), 121-144.

verdict is welcome, if somewhat predictable. It leaves the rest of us, however, with a problem. If we are to regard the use of RCTs in evaluating medical practices as something more than an administrative convenience, then we need to ask: what kind of a decision making process are we committing ourselves to when we advocate an expanded role for RCTs, and what are we getting in return?

The following discussion is based on two premises. The first is an old and familiar dictum: in selecting a particular forum for making social decisions, we are also choosing who will decide and what decisions will be made.<sup>29</sup> Discussions of the optimal methods for therapeutic experimentation could benefit from a generous dose of such political realism. At the same time, I will argue, discussing the institutional politics of methodological reformers is a fruitless exercise unless the discussion provides us with some criteria for evaluating those politics. My second premise, accordingly, is that thinking about politics in this case means not only thinking about the institutions which develop and enforce our policies toward medical practices, but the criteria by which to judge both policies and institutions.

With these considerations in mind, I want to return to the distinction presented previously, between advocates of a "dogmatic" and "anti-dogmatic" view of statistical methods, and examine more closely the anti-dogmatic position. What it offers, in place of a reliance on method, is dependence on the scientific judgment of the medical

---

<sup>29</sup> The point is developed at length in E.E. Schattschneider, The Semi-Sovereign People (New York: Holt, Rhinehart and Winston, 1960).

community. Given that the outcome of therapeutic interventions is both multi-causal and multi-dimensional, proponents of the anti-dogmatic view are in effect arguing that judgments about therapeutic merit are too complex to be encompassed by a single decision rule or methodological criterion. The statistical measures offered by a properly randomized study provide one important measure of the evidence concerning a treatment. But this information is only one piece of data about the effects of this therapy: it ought to be evaluated in the context of other information about the mechanisms of disease and treatment, other studies of the therapy, and other measures of therapeutic outcome.

Such a view has much in common with traditional notions of therapeutic reform. Implicit in this line of reasoning is the assumption that we ought to delegate judgments about therapeutic practices to the community of experts. What we think about this argument will depend, at least in part, on how well we think such delegation has served us in the past. In examining this question, it will prove useful to draw a distinction between two aspects of the problem which are not always clearly distinguished: do these experts make "correct" decisions? do they make the "correct" decisions for us?

In the culture of biomedicine, language about the "correctness" of therapeutic evaluations usually translates to an interest in evaluating the efficacy of the treatments being studied. The advantage of relying on RCTs, proponents argue, is that given the complexity of factors determining therapeutic outcomes, a well designed randomized experiment can do a better job of estimating the effects of treatment than an individual or group of individuals relying on unaided judgment. Given

what we know about the inability of most individuals to interpret and evaluate quantitative data, this argument seems plausible.<sup>30</sup> If we regard randomized experiments as an aid to scientific and clinical judgement, rather than a substitute for it, then such studies do offer a more exacting measure of treatment effect than traditional clinical studies.

The more telling objection made by advocates of clinical judgment is that therapeutic outcomes, in addition to being multi-causal, are multi-dimensional. Symptoms, physiological or biochemical processes, and functional status are all plausible measures of response to treatment. For many diseases, reasonable agreement about a measure of effect can be obtained, despite the plurality of outcome measures. In the case of acute infections, for example, the observation that therapeutic results can be measured clinically, by reductions in mortality or relief of symptoms, or serologically, by measuring infectivity or immunity, is academic. For other conditions, the problem is more difficult. Some treatments heal lesions but do not relieve symptoms. Are the patients "healed"?<sup>31</sup> Other treatments reduce mortality but only at the cost of substantial impairment for the survivors.

Advocates of clinical judgment cite such dilemmas in arguing that particular RCTs have not always captured the relevant therapeutic

<sup>30</sup> Amos Tversky and Daniel Kahnemann, "The Framing of Decisions and the Psychology of Choice," Science 211 (1981), 453-458.

<sup>31</sup> See the paper by Walter I. Peterson, Richard A.L. Sturdevant, and Harold D. Frankl, et. al. "Healing of Duodenal Ulcer With An Antacid Regimen," New England Journal of Medicine 297 (August 18, 1977), 341-345 and the ensuing correspondence, Robert H. Palmer, "Antacid Therapy of Duodenal Ulcers," New England Journal of Medicine 297 (December 8, 1977), 1294.

outcomes. But the corollary of this argument, that the medical community is adequately equipped to make such Hobson's choices, is not necessarily convincing. Evidence is mounting that in such difficult situations, the choices physicians make on behalf of patients are not necessarily the choices patients would make for themselves. In a series of studies, Barbara McNeil has asked former and prospective patients to evaluate life extending but disabling medical treatments. Their expressed preferences do not support the traditional measure of medical benefits captured in indices such as the five year survival rate, customarily used in randomized studies of cancer treatment.<sup>32</sup>

Such findings suggest that proponents of RCTs, in emphasizing the need to improve the amount of reliable knowledge concerning the effects of medical treatments have, at best, only partially grasped the problem.<sup>33</sup> Their conception of the task of therapeutic reform assumes either that the benefits of treatment are unequivocal or that the profession can reliably act as the patient's agent in choosing among the available measures of outcome. It is precisely these assumptions which the findings of McNeil and her colleagues call into question.

---

<sup>32</sup> Barbara J. McNeil, R. Weichselbaum, S.G. Pauker, "Fallacy of the Five Year Survival Rate in Lung Cancer," New England Journal of Medicine, 1397-1401; "Correspondence: Fallacy of the Five Year Survival Rate," New England Journal of Medicine 300 (1978), 927-928; Barbara J. McNeil, R. Weichselbaum, S.G. Pauker, "Speech and Survival: Tradeoffs Between Quantity and Quality of Life in Laryngeal Cancer," New England Journal of Medicine 305 (1981), 982-987.

<sup>33</sup> Earlier in the discussion, I referred to the tension between those who interpret this task to consist of finding out which treatments "work," and those who will remain uneasy unless we know why and how they produce their effects. Even at its extremes, this dispute remains a family quarrel. Whatever their differences, both camps share in the belief that the path to therapeutic reform lies in improving the quality of the information on which therapeutic practice is based.

But if expert physicians and lay persons do not agree on the value of medical benefits in some instances, why should that disagreement affect our procedures for evaluating new medical practices? Methodological reformers might argue that by conflating inferences about the value of particular treatments with critical judgments about clinical decision making, I have confused problems of medical science and medical practice. After all, better information about the effects of medical procedures potentially benefits patients as well as physicians.

The difficulty, I would suggest, stems from the belief that factual information about the benefits of medical treatment can be produced without regard for the way in we choose to value these benefits. The routine separation of judgments about truth and value called for by this belief is illusory.<sup>34</sup> Nor is the illusion merely conceptual: our institutions for allocating medical goods presuppose the unequivocal character of medical benefits.

In an earlier era of medical technology, the allocation of medical goods presented a relatively straightforward problem in social engineering: ensuring that effective procedures reached those who would benefit from them. The standard measure of benefit, accepted by public and profession alike, was unequivocal: lives saved or deaths averted. The tasks assumed by public policy followed on this definition of the situation: subsidizing the distribution of goods deemed effective by

---

<sup>34</sup> For an excellent exposition of this point in another policy context highly dependent on statistical analysis, see Joel Yellin's essay on environmental policy, "Who Shall Make Environmental Decisions," American Statistician 37 (November, 1983), 362-366.



the medical profession to individuals who, for various reasons, were unable to pay or save for the goods themselves.

Advances in medical technology, it has been argued, have transformed this situation. We can no longer expect the dramatic breakthrough discovery, the technology or group of technologies, which will radically alter our expectations of life.<sup>35</sup> Medical innovations operate at the margin and at the margin, we are told, there is little to gain.<sup>36</sup> Our existing policies are, according to this argument, obsolete: new institutions and principles for distributing medical goods must be found.<sup>37</sup>

An unusual consensus has arisen regarding these premises, at least among social scientists. Much of the ensuing debate has centered on the appropriate roles of governments and markets in allocating health benefits. But regardless of their ideological coloration, present discussions of health policy concur in the assessment that the medical profession has overestimated the benefits of many medical practices.<sup>38</sup>

---

<sup>35</sup> There are some, such as Thomas McKeown, who question whether such breakthroughs ever existed; McKeown's challenge is directed at the mythologies which arose around the first and second generation antibiotics.

<sup>36</sup> Victor Fuchs, Who Shall Live? Health Economics and Social Choice (New York: Basic Books, 1974), esp. 3-29.

<sup>37</sup> See Fuchs, Who Shall Live? (n. 36), 143-151; Lawrence D. Brown, "The Scope and Limits of Equality as a Normative Guide to Federal Health Care Policy," Public Policy 26 (Fall, 1978), 482-532.

<sup>38</sup> The issue with which I am concerned is largely orthogonal to contemporary debates about market improvements and more planning. Thus, both advocates of improving medical markets and the opportunities for consumerism, such as Peter Temin, Professor of Economics at the Massachusetts Institute of Technology, and proponents of a larger public role for technology assessment, such as Harvey Fineberg, Dean of Harvard School of Public Health, are in agreement on this point.

Yet the challenge thereby offered to medical authority is a limited one, premised on the assumption that physicians misjudge the amount, but not the nature of medical benefits.<sup>39</sup> By offering improved procedures for estimating the benefits of medical practices, methodological reformers promise to return us to an earlier state, in which the problem for public policy is limited to deciding how to subsidize the distribution of effective goods.

My argument is also about changes in the capacities of medical technology, but it emphasizes a different aspect of that technology. In the current era, medical advances determine not only whether we live or die but, increasingly, how we will live and die.<sup>40</sup> Ordinary citizens, those who are potential patients and the friends and relatives of such patients, have as much to say on these aspects of medical goods as any expert. As presently designed, however, our institutions give citizens little voice in these matters. Their's is the consumer's choice: pick a brand or shop around.<sup>41</sup>

Under these circumstances, the task for public policy is accordingly more complex. Our present difficulties are both conceptual and institutional. It is not simply a matter of redesigning the institutions which allocate medical goods or reexamining the principles on which we

---

<sup>39</sup> See, for example, the discussion by Richard J. Zeckhauser and Donald S. Shepard, "Where Now For Saving Lives?" Law and Contemporary Problems 40 (Fall, 1976), 5-45.

<sup>40</sup> Alonzo Plough, Borrowed Time: Artificial Organs and the Politics of Extending Life (Philadelphia: Temple University Press, 1986).

<sup>41</sup> The choices some consumers face are even more limited than that. See, for example, Rudolf Klein, "Models of Man and Models of Policy: Reflections on Exit, Voice and Loyalty Ten Years Later," Milbank Memorial Fund Quarterly 58 (Summer, 1980), 426-428.

subsidize their distribution. It is more a question of establishing institutions in which our traditional concepts of the benefits of medicine can be deliberated.

Why deliberated? First, because in numerous instances professional and lay concepts of the benefits to be gained by increasing our technological capacities in medicine are at odds. These differences need to be articulated and explored, preferably in a context where the professionals do not hold the upper hand. Second, because there is no archimedean point from which the benefits of new technological capacities can be calculated, once and for all. Different communities, with different traditions, will value these benefits differently.<sup>42</sup> Thirdly, because most of us have yet to examine what we think about these matters, much less to discuss our thoughts with others. At best we have opinions, grounded in fears and hopes. Deliberation involves scrutiny of opinions, in the light of experience and reason as well as hopes and fears.

Calls for deliberation evoke images of travelling Chataquas. While a road show for enlightenment is not what I have in mind, the concern is legitimate. What good is talk when power or money is at stake? And talk of different communities with differing traditions makes many of us nervous, perhaps rightly so.<sup>43</sup> While I do not have a detailed

---

<sup>42</sup> Alasdair MacIntyre, "Patients as Agents," in S.F. Spicker and H.T. Engelhardt, Jr. eds. Philosophical Medical Ethics: Its Nature and Significance (Dordrecht: D. Reidel Publishing Company, 1977), 197-212.

<sup>43</sup> This is, after all, the language of state's rights and segregation. In matters of economic and political power, it evokes the status quo ante: before the Johnson Administration, that is. Anyone who does not understand this point should re-read Grant McConnell's Private Power and American Democracy (New York: Knopf Publishing Company, 1966).

institutional proposal to put forth, elaborating these and other objections to my argument may offer some criteria for thinking about the institutional details of various proposed reforms.

The most obvious objections to any call for increased participation in public affairs are the questions of capacity and interest. "What makes you think that most citizens are either capable or interested in evaluating new medical technologies? These are tasks for experts, if not because their knowledge better qualifies them, then because they are more highly motivated to participate in these decisions." Interest, however, is largely a function of how issues are framed. Opportunities to review a hospital's application to purchase a new cardiac catheterization laboratory are likely to bring out the professional consumers. Reframe the issue as one of deciding between buying such a piece of equipment and investing in a program for home health visits for the elderly, and a different sort of constituency shows up. Indeed, one measure of any program for therapeutic reform is how well it succeeds in posing issues which will capture the general interest in medical affairs, and not merely bring out the specialized publics.<sup>44</sup>

What about capacity? It is by no means novel to argue that the specialized knowledge and skills needed to evaluate the benefits of medical procedures are beyond the capacity of most citizens.<sup>45</sup> But it

---

<sup>44</sup> Harry M. Marks, Representative Health Planning: Who Represents Whom? (Boston: Harvard School of Public Health, 1979).

<sup>45</sup> The argument goes back at least as far as John Dewey's The Public and Its Problems (Athens: Ohio university Press, 1976). Even so fervent an advocate of participation as Amy Guttmann allows for substantial amounts of bureaucratic "execution" of public health policies on the grounds that specialized expertise is necessary in certain aspects of these matters. See Amy Guttmann, Liberal Equality (Cambridge: Cambridge University Press, 1980), 211-212.

requires no special skills to have experienced the management of disease and illness in our modern medical institutions. What we lack, under our present institutional arrangements, are the opportunities to articulate these experiences to others and to examine our policies toward medical practices in light of these experiences. I do not wish to belittle the issue of capacity, but to suggest that our current presumption equating capacity with specialized technical knowledge and skills is mistaken. The capacity needed here is the ability to imagine the possible consequences of present decisions and policies for our futures, both individual and collective, and the talents to communicate these images to others.<sup>46</sup> Like the level of public interest, such capacity is as much an outcome of our policies as a pre-requisite for designing workable policies.

A good deal of my argument about the possibility of participatory decision making here depends on which decisions we regard as subject to participation. Our present policies have diminished participation both in the decisions to deploy new technological capacities in medicine and in the decisions to utilize these capacities in particular cases. The former class of decisions is left largely to the specialist communities, the federal regulatory authorities, and the corporations. The second

---

<sup>46</sup> On the importance of talk in the democratic process, see Benjamin Barber, Strong Democracy. Participatory Politics for a New Age (Berkeley: University of California Press, 1984), 173-198.

class of decisions are left largely to the individual physician.<sup>47</sup> In the first instance, the consequences of the decisions are too remote and too abstract; in the second instance, too immediate and personal, to allow for effective public participation.

What is lacking is a level of decision making which leaves the participants interested in and capable of taking both personal experience and social consequence into account. Such decisions would be public regarding, in the sense that they depend on a recognition that both our predicaments and opportunities are held in common.<sup>48</sup> This seems like a problem in institutional design, but it is equally a matter of political choice.<sup>49</sup> Whether or not we develop suitable democratic institutions

<sup>47</sup> While there is some empirical evidence that individuals decide whether or not to "consume" medical care, there is little evidence of participation in the decision as to what medical care (procedures, drugs) they will receive. See Irving K. Zola, "Culture and Symptoms: An Analysis of Patients' Presenting Complaints," American Sociological Review 31 (1966), 615-63; John E. Wennberg, Benjamin Barnes and Michael Zubkoff, "Professional Uncertainty and the Problem of Supplier Induced Demand," Social Science and Medicine (1982), 811-824.

<sup>48</sup> On the notion of the public employed here, see Manfred Stanley, "On the Mystery of the Commons: On the Indispensability of Civic Rhetoric," Social Research 50 (Winter, 1983), 860-863. J. David Greenstone's discussion of procedural and substantive concepts of the public in Grant McConnell's work is also germane: see his "The Transient and the Permanent in American Politics: Standards, Interests and the Concept of 'Public'," in J. David Greenstone, ed. Public Value and Private Power in American Politics (Chicago: The University of Chicago Press, 1982), 12-19.

<sup>49</sup> It would be ludicrous in any case to specify in advance the size of the jurisdiction which would meet this criteria: neighborhoods, communities, regions and even nations could conceivably fit the bill. On the case for nations as communities, see Benedict Anderson, Imagined Communities. Reflections on the origins and Spread of Nationalism (London: New Left Books, 1983) and Richard M. Titmuss, The Gift Relationship (New York: Vintage Books, 1971).

for valuing medical goods depends in part on whether we as citizens are willing to accept responsibility for the ensuing decisions.<sup>50</sup>

Who makes, and who enforces, the decisions made regarding medical practices are key questions. But in speaking of decisions, I do not wish to lose sight of deliberation. Our technocratic culture places a high premium on making decisions and, especially in the medical realm, on making the "right" decisions. Yet unless we create institutions which enable us to reflect on our decisions and to communicate those reflections to one another, we do not have an independent means to judge the "rightness" of these decisions. The value of our decisions, on this argument, is historical, in the sense that our experience of past and present decisions provides us with a means to assess our future choices.<sup>51</sup> But without a public forum for deliberating our interpretations and valuations of the present and past, we have no way of incorporating this experience into our collective decisions: it remains private (and often unarticulated, even to ourselves).<sup>52</sup>

All political institutions must somehow resolve the tension between deciding and deliberating, between acting and judging. Institutions which do not permit us to act, in the name of deliberation, are as

---

<sup>50</sup> At least part of the case for delegating our policies for valuing and allocating medical goods to the medical community rests on the contention that the rest of us are unwilling to accept this responsibility.

<sup>51</sup> Charles Taylor, "Social Theory as Practice," in Charles Taylor, Philosophy and the Human Sciences. Philosophical Papers (Cambridge: Cambridge University Press, 1985), 2: 91-115.

<sup>52</sup> On the distinction between private and public judgments, see Michael J. Sandel, Liberalism and the Limits of Justice (Cambridge: Cambridge University Press, 1982), 179-183.

damaging as those which exclude most opportunities to reassess our prior decisions.<sup>53</sup> Our present institutions for valuing and allocating medical goods have, on my argument, gone too far in the latter direction. What about current proposals for reforming these institutions?

I want to examine two recent proposals for reform: the first an appeal to establish an Institute for Health Care Evaluation; the second a call for community based review and determination of Medicare budgets and benefits. I choose these two examples not because they capture the full range of current efforts to evaluate medical practices, but because both represent self-conscious attempts to adopt the traditional programs of therapeutic reformers to contemporary political realities, and because they invoke the antipodes of contemporary political reform: corporatist and communitarian.<sup>54</sup>

My purpose in examining these proposals is twofold: first, to ask what we might expect from these particular reforms if instituted and second, to explore the uses and limitations of the preceding analysis in thinking about particular reform proposals. In each case I propose to focus on the following questions: 1) what kinds of decisions about medical practice will the proposed bodies be responsible for? 2) Who decides, and what rights and responsibilities go along with participation? 3) What criteria and standards for evaluating the benefits of new

---

<sup>53</sup> Ronald Beiner, Political Judgment (Chicago: University of Chicago Press, 1983), 148-149.

<sup>54</sup> For a comprehensive review of programs in medical technology assessment, see Clifford S. Goodman, "Appendix A: Profiles of 20 Technology Assessment Programs," in Institute of Medicine, Committee for Evaluating Medical Technologies in Clinical Use, Assessing Medical Technologies (Washington: National Academy Press, 1985), 255-489.



medical practices are proposed? 4) What opportunities, if any, do the proposed reforms create for deliberating the standards and criteria to be used in these evaluations?

Professor John Bunker and his colleagues at the University of California have proposed that health insurers pool funds to subsidize an Institute for Health Care Evaluation.<sup>55</sup> The proposed institute will collect and evaluate reports on medical practices. While it will not confine itself to data generated from RCTs, the Institute will sponsor or conduct controlled trials of common medical and surgical practices, and it will issue reports on its findings to corporate sponsors. However, the Institute will have no "responsibility for making policy (e.g. deciding whether to reimburse physicians for a new procedure)."<sup>56</sup>

The proposed Institute has much in common with the programs of traditional therapeutic reformers. Its principal function is to provide information about the (medical) benefits of new technologies and practices, information which will be gathered (and assessed) from various sources. Like the long-lived AMA Council on Pharmacy and Chemistry, it will pronounce judgment on the quality of evidence which underlies its recommendations, but take no responsibility for seeing its judgments translated into action. Unlike the Council, however, its members will include not only physicians and scientists but insurers and "public" representatives: employers, employees and consumer

---

<sup>55</sup> John P. Bunker, Jinnet Fowles, and Ralph Schatfarzick, "Evaluation of Medical Technology Strategies. Proposal for an Institute for Health-Care Evaluation," New England Journal of Medicine 306 (March 18, 1982," 687-692.

<sup>56</sup> Bunker, et. al., "Evaluation of Medical Technology," (n. 55) 689.

"representatives".<sup>57</sup> As members of the Institute, third party payors will have a say in deciding which topics are studied intensively and presumably they, in negotiations with the buyers of insurance, will be responsible for making decisions about reimbursement coverage on the basis of the data provided.<sup>58</sup>

Its sponsors regard the Institute as an innovative response to contemporary political realities, most notably the eagerness of the Reagan administration to leave negotiations about the scope of health care benefits to the "private" sector, and its reluctance to set national standards for medical practices. By providing the buyers and sellers of health insurance with "high quality" information about the benefits of new medical practices, the Institute's sponsors propose, in Wildavsky's phrase, to speak truth to power. Yet in seeking to adapt the traditional practices of therapeutic reformers to new political realities, the proposed Institute may be combining the worst of several worlds.

First, the Institute bears no responsibility for the decisions which result from its assessments or even how those assessments are presented and interpreted. Under these circumstances, two scenarios seem possible. One possibility is that the Institute's reviews would be used to legitimate rationing decisions which third party payors would

---

<sup>57</sup> "The medical profession would be represented by a combination of researchers, academicians, and clinical practitioners." Bunker, et. al., "Evaluation of Medical Technology," (n. 55) 688.

<sup>58</sup> Bunker, et. al., "Evaluation of Medical Technology," (n. 55) 688-689.

like to make, but are unable to defend at present.<sup>59</sup> The second, and more likely outcome, is that in the competition among insurers the demand for coverage of particular benefits may overwhelm all but the most negative evaluations.<sup>60</sup>

If the professional and scientific members of the Institute are excluded from subsequent decisions to provide coverage for specific medical practices, corporate sponsors are not excluded from decisions about what to study and how to study it. The Institute's sponsors allude to a system of "checks and balances" which will ensure the integrity of the evaluation process. Both public and professional representation are expected to offset any undue influence on the standards and criteria of evaluation.<sup>61</sup> But the system of representation here is "corporatist", and as in all corporatist systems, only organized "publics" need apply.

How one regards these two aspects of the proposed Institute (its lack of responsibility for decisions made in its name or its corporatist

---

<sup>59</sup> There is even some intimation that this might be the case. See Bunker, et. al., "Evaluation of Medical Technology," (n. 55) 690.

<sup>60</sup> The recent policy debate on competition has focused almost exclusively on price competition and/or the effects of competition on total expenditures; there is little empirical data on the degree of product competition, e.g. competition over benefits.

<sup>61</sup> Although the proposal speaks of "checks and balances," it is not clear exactly what these constitutional arrangements are, or how they will work. Each party to the Institute will have one vote. The authors of the proposal also describe evaluation as occurring under "local control and responsibility"; what they appear to mean by this, however, is control by hospital and medical school Institutional Review Boards. See Bunker, et. al., "Evaluation of Medical Technology," (n. 55) 688, 690. On the performance of IRB's in policing the quality of research designs, see Mary Giammona and Stanton A. Glantz, "Poor Statistical Design in Research on Humans: The Role of Committees on Human Research," Clinical Research 31 (December, 1983), 572-578.

system of representation) might depend on one's general analyses (and valuations) of such organizational forms. Whether such corporatist systems are more effective than parliamentary-bureaucratic systems in making (and enforcing) decisions, and whether they damage the prospects for more democratic forms of representation and participation are the subject of some debate.<sup>62</sup> But my arguments have less to do with the desirability, in general, of enhancing participation and deliberation, than with a specific observation about the need for more participation and deliberation in evaluating medical technology.<sup>63</sup>

If one accepts the analysis of medical technology presented above—that we need to reassess not merely the benefits of new medical technology, but the ways in which think about these benefits—then the proposal for a new Institute is inadequate in several respects. It is a closed system: the interested parties are identified at the outset, and their negotiations are not open to public scrutiny. Both features exclude the possibility of creating new interests in these matters, and both severely restrict the opportunities for any collective reassessment of the criteria and standards used in evaluating the benefits of new technologies. That the rigor and scope of the Institute's assessments

---

<sup>62</sup> See Claus Offe, "The Attribution of Public Status to Interest Groups: Observations on the West German Case," in Suzanne D. Berger, ed. Organizing Interests in Western Europe: Pluralism, Corporatism and the Transformation of Politics (Cambridge: Cambridge University Press, 1981) 123-158 and Charles F. Sabel, "The Internal Politics of Trade Unions," ibid., 209-244.

<sup>63</sup> In other words, whether or not we regard participation and deliberation as public goods, necessary to the health of the polity, is not immediately germane to the argument I am making, which is about the need, under the present historical circumstances, for more participation and deliberation in allocating and evaluating medical goods.

are themselves subject to private negotiations further limits the opportunity for public consideration of these matters. And given the decision driven nature of the enterprise—shall insurers reimburse providers for a specific procedure?—questions about the impact of medical practices on our daily lives are likely to get short shrift in these private negotiations.

The second of our two proposals also works through the reimbursement system. Professor John Wennberg, of Dartmouth Medical School, has recently suggested a move towards community based financing and review of medical services. Though motivated by the wish to make therapeutic reform relevant to the contemporary political rhetoric of "competition," Wennberg's proposal has the potential to engender more democratic control over the valuation and allocation of medical goods.<sup>64</sup>

Wennberg's scheme rests on a body of analytic work suggesting that populations in neighboring communities experience radically different rates of hospitalization for elective surgery, differences which appear to be more closely associated with the clinical "styles" of local practitioners than with any identifiable risk factors in the population

---

<sup>64</sup> Wennberg's proposals are set out in two articles: John E. Wennberg, "Should the Cost of Insurance Reflect the Cost of Use in Local Hospital Markets," New England Journal of Medicine 307 (November 25, 1982), 1374-1381; *idem.*, "Dealing With Medical Practice Variations: A Proposal for Action," Health Affairs 3 (Spring, 1984), 6-32. Both published proposals emphasize organized consumer participation; the first article also stresses the participation of insurance companies and the second that of organized medicine. While earlier, unpublished versions of Wennberg's analysis were more communitarian, the emphasis placed on democratic participation in the following discussion is entirely my responsibility.

itself.<sup>65</sup> These "local" variations in hospitalization appear to be responsible for much of the differences in per capita expenditures between local communities. In Massachusetts, for example, residents of Boston "spend" \$923 per capita for Medicare inpatient services, while residents of nearby Milton "spend" only \$444.<sup>66</sup> So long as insurance premiums are set on a regional or statewide basis, the resulting income transfers from communities whose physicians practice conservatively to those with high utilization rates are substantial, as much as \$1.75M per year in Maine, for example. Such subsidies permit, and possibly encourage, existing practice patterns to flourish unexamined.<sup>67</sup>

Wennberg proposes that insurers set premiums which are more closely related to the practice and expenditure patterns of these local communities. His interest, however, is not in writing more actuarially sound insurance policies, but in calling public attention to local styles of medical practice and their effects. The procedures for which there is the greatest variation are often those whose value is hotly contested within the medical profession. One motive in Wennberg's proposal is to promote the demand for more rigorous evaluation of medical practices within the medical community.<sup>68</sup> However, his emphasis on community based analysis and review of medical practice

---

<sup>65</sup> More recently, Wennberg has reported on variations in medical, as opposed to surgical, practice. See his "Dealing With Medical Practice," (n. 64) 12-14, 18-19.

<sup>66</sup> Wennberg, "Cost of Insurance," 1380.

<sup>67</sup> Wennberg, "Cost of Insurance," (n. 64) 1375-1377.

<sup>68</sup> Wennberg, "Dealing With Medical Practice," (n. 64) 24-26.



The Libraries  
Massachusetts Institute of Technology  
Cambridge, Massachusetts 02139

Institute Archives and Special Collections  
Room 14N-118  
(617) 253-6688

This is the most complete text of the  
thesis available. The following page(s)  
were not included in the copy of the  
thesis deposited in the Institute Archives  
by the author:

pg 271

of medical practices be publicly reviewed creates at least the possibility for discussing aspects of that experience which are not readily captured in epidemiologic statistics or therapeutic experiments.

Finally, unlike reviews driven by the need to decide whether or not to cover a particular medical practice, the scope of these deliberations is open ended. Public review and discussion constitutes the starting point of policy in this instance, and not merely a mechanism for enacting policies already established elsewhere.<sup>70</sup> Whether or not the potential for participatory deliberation is realized depends in large part on how proposals of this sort are implemented: the degree to which, for example, mobilizing general interests is accepted as an intrinsic and desirable element in such reforms.<sup>71</sup>

What kinds of objections might be made to proposals for community based review of medical practices? One objection is that such reforms are too utopian to possibly work.<sup>72</sup> Another objection, ironically, stems from the fear that such proposals will work, and that the resulting decisions will be made on the basis of inferior information, or worse, not on the basis of information at all.<sup>73</sup> Both objections

<sup>70</sup> As was the case, for example, under the old health planning law, PL 93-641. See Marks, Representative Health Planning (n. 44).

<sup>71</sup> Wennberg himself alternates between suggestions to rely on community elites and other parties with established interests and more participatory approaches.

<sup>72</sup> Peter Buck, personal communication.

<sup>73</sup> Communities will act, according to this argument, on the basis of their hopes for what medical treatment can accomplish rather than recognize its limitations, or will ignore the question of medical benefit entirely, focusing on the financial gains (or losses) which further investment in medical goods entails. In either case, the methodical but abstruse reasoning involved in the experimental evaluation of medical practices would have little or no voice.



are not so much diagnoses of a problem as symptoms of a disorder: the mutual isolation of science and politics.

Unlike the narrower problem of assessing the value of medical practice, the difficulty of incorporating scientific knowledge in political deliberations transcends the particular institutions in which decisions about medical practice take place. And it is a problem for which we lack, at present, even the glimmer of a possible solution. The best I can do is to articulate the dilemma, as it affects our current thinking about therapeutic reform.

### III: Science and Politics: Closing Reflections

At least part of what I have been arguing here is that the relations between political and scientific truth are too important to foreclose by insulating scientific institutions and decisions from the considerations which arise in the political communities affected by those decisions.<sup>74</sup> But it seems equally important not to insulate political institutions and decisions from the considerations which arise in the realm of science. If scientists have something to learn from politics, then perhaps politics has something to learn from science.

There is a long-standing tradition in liberal political theory (and institutions) that knowledge about our social institutions is developed by the central organs of government and locally applied.<sup>75</sup> The democratic theorists whose notions of participation I have relied on

---

<sup>74</sup> This I take to be the essential point of Don Price's meditation on The Scientific Estate (Cambridge: Harvard University Press, 1965).

<sup>75</sup> John Stuart Mill, Considerations on Representative Government (New York: The Liberal Arts Press, 1958), 224-226.

here are hostile not merely to the idea that the federal government has a privileged relation to knowledge but that anyone does. When it comes to claims about knowledge, they imply, we are all equals. The further implication drawn from their arguments is that all truths are equal.<sup>76</sup>

The target at which these democratic theorists are aiming is the unwarranted appropriation of expertise to justify the control of social decisions by elites. But in furthering their attack, they have obliterated a much wider target: the possibility that science can contribute at all to evaluating our institutions and decisions. The hostility evidenced toward the claims of science hardly furthers the democratic aim of giving citizens the capacity to deliberate those decisions and policies which rely in part on specialized knowledge.<sup>77</sup> The ability to assess the achievements as well as the limitations of that knowledge is an essential part of that capacity.

The difficulty remains of establishing institutions in which claims about knowledge and arguments about the value of that knowledge for social decisions can be deliberated. In that sense, my argument concludes with an appeal for more, not less, science in these matters.

---

<sup>76</sup> The hostility of democratic theorists to notions of rational truth (including science) comes out particularly strongly in the relativism of Benjamin Barber's Strong Democracy (n. 46). Other theorists, such as Charles Taylor are more inclined to emphasize the differences in epistemological status between the social and the natural sciences. On the problems with such dualism see Hilary Putnam, "Beyond Historicism," in idem., Realism and Reason. Philosophical Papers. Volume 3 (Cambridge University Press, 1985), 287-303.

<sup>77</sup> It is not necessary to agree with Mill's premise that the central government is more enlightened and knowledgeable to agree with his observation that what is "wanted is the means of making ignorance aware of itself and able to profit by knowledge...." Mill, Considerations (n. 75), 228.

Only when the rest of us are in a position to comprehend and evaluate the claims made by methodological reformers will we be in a position to make use of their advice. So long as we can only choose between deferring to their authority or ignoring their counsels, then we left choosing between medical science and medical politics. And that is an unfortunate choice.

### A Note on Sources

A work such as the present is dependent on three sorts of materials: printed sources, manuscript sources and works of historical and social commentary. Most of my obligations to the literature and my assessments of it are recorded in the footnotes to the text. What follows is a brief additional commentary and guide.

#### A. Secondary Literature

There is little to go on in reconstructing the intellectual and institutional history of the medical sciences in the 20th century. The thematic essays collected in Morris J. Vogel and Charles E. Rosenberg, eds. The Therapeutic Revolution. Essays in the Social History of American Medicine (Philadelphia: University of Pennsylvania Press, 1979), while thought provoking for the disciplines and settings they touch on, were not very helpful for the events and developments chronicled here. For the perennial questions of who did what, who knew whom and who was where when, I have relied on A. McGehee Harvey's invaluable compendium, Science at the Bedside. Clinical Research in American Medicine, 1905-1945 (Baltimore: The Johns Hopkins University Press, 1981). While a few figures on the periphery of academic medicine escaped Harvey's attention, they are thankfully few in number. Harvey's accounts are usefully supplemented by two old standards: Who Was Who in America and Cattell's American Men & Women of Science, each of which has gone through numerous editions.

For finding out more about the evolution, intellectual and social, of pharmacology, I am especially indebted to two doctoral theses cited in the text: Jonathan Liebenau, Medical Science and Medical Industry, 1890-1929: A Study of Pharmaceutical Manufacturing in Philadelphia (University of Pennsylvania, 1981) and John Patrick Swann, The Emergence of Cooperative Research Between American Universities and the Pharmaceutical Industry, 1920-1940 (University of Wisconsin, 1985). Forthcoming publication will make both theses more widely available. Despite its title, The Therapeutic Revolution (Philadelphia: University of Pennsylvania Press, 1979) edited by Charles Rosenberg and Morris Vogel, says very little about 20th century therapeutics, or the ancillary sciences most closely connected with it.

My sense of the 19th century background to 20th century therapeutics comes largely from Martin S. Pernick's A Calculus of Suffering. Pain, Professionalism and Anesthesia in Nineteenth Century America (New York: Columbia University Press, 1985) and from the various essays of John Harley Warner: "The Nature Trusting Heresy: American Physicians and the Concept of the Healing Power of Nature in the 1850's and 1860's," Perspectives in American History 11 (1977-1978), 291-324; "Physiological Theory and Therapeutic Explanation in the 1860's: The British Debate on the Medical Use of Alcohol," Bulletin of the History of Medicine 54 (1980), 235-257; "The Selective Transport of Medical Knowledge: Antebellum American Physicians and Parisian Medical

Therapeutics," Bulletin of the History of Medicine 59 (Summer, 1985), 213-231. Warner's book length study, The Therapeutic Perspective (Cambridge: Harvard University Press, 1987) appeared too late for me to make use of it. Neither author should be considered responsible for the remaining insensitivities to 19th century developments which this manuscript manifests despite their instruction.

The institutional side of American medical research and politics is not in much better shape, despite the greater amount of attention paid to it in recent scholarship. The reader seeking to know the who, what and when of organizational politics in medicine will still rely heavily on Rosemary Stevens's American Medicine and the Public Interest (New Haven: Yale University Press, 1971). Despite the great interest in the "machinations" of the American Medical Association in the first half of the 20th century, none of the present books on the AMA tell much about its internal politics, and in particular about the relations between the AMA and the academic medical community prior to World War II, or about the relation between the AMA leadership and its rank and file. Trying to place a political-scientific operation like the AMA's Council on Pharmacy and Chemistry in its proper place is accordingly quite difficult. Despite paying a great deal of attention to the physics and engineering community, historians of science have not paid much attention to the politics of medicine and biology during and after World War II. The near contemporary accounts cited in chapters 3 and 4 remain useful guides to those of us not especially interested in the history of the atomic bomb project. See especially, Irwin Stewart Organizing for War. The Administrative History of OSRD (Boston: Little Brown and Company, 1948.) NIH politics are seen mostly from the Congressional side in Steven Strickland's pioneering Politics, Science and Dread Disease (Cambridge: Harvard University Press, 1972) and the evolution of the NIH internal bureaucracy, its relation to the academic medical community and to other loci of medical research policy remain obscure.

On the FDA, James Harvey Young's work, especially The Medical Messiahs. A Social History of Health Quackery in 20th Century America (Princeton: Princeton University Press, 1967) retains a commanding presence. Peter Temin's Taking Your Medicine. Drug Regulation in the United States (Cambridge: Harvard University Press, 1980) picks up interest where Young leaves off, in the 1940's. The multiple editions of Charles Wesley Dunn and Vincent Kleinfeld, Federal Food, Drug and Cosmetic Act. Judicial and Administrative Record (New York: Commerce Clearing House) provide an indispensable guide to the actions of the FDA and the relevant court rulings.

The reader interested in the history of contemporary statistics is by and large at a total loss. The various collections memorializing and celebrating senior statisticians provide a useful path of entry to events, persons and conceptual developments but the outsider in search of an overview here will not find it.

## B. Published Sources

For those interested in finding their way into the medical literature, the easiest way in for the period prior to 1930 are the Index Catalogues of the Surgeon-General, Series 1-3, which detail both the book and periodical literature. For the periodical literature, the annual Index Medicus--and its various incarnations--Current Guide to the Medical Literature, etc.--are indispensable if at times overwhelming. It takes any novice some time and ingenuity to think of the possible topic headings under which medical librarians may have catalogued items of interest and relevance. I have supplemented these guides with systematic reading in several of the relevant journals: especially, The Journal of Clinical and Laboratory Investigation and the Journal of Experimental Medicine for the 1910's and 1920's; Proceedings of the American Drug Manufacturers for the 1930's and 1940's; Biometrics for the 1950's and 1960's; Clinical Pharmacology and Therapeutics for the 1960's and 1970's; Controlled Clinical Trials for the 1970's and 1980's. As most historians know, a systematic read through the table of contents of key journals turns up items and issues left unnoticed by the most creative user of indexes and on-line data bases.

## C. Manuscript Sources.

The bulk of the work has been done from manuscript sources. For someone who cut his teeth on the French archival system, the anarchy of the American "system" takes some getting to use to. Even finding whether someone's papers exist, much less where they are, can be a major undertaking. The general catalogues, such as the National Union Manuscript Collections, are more useful for tracing the papers of Congressmen and other national institutions than for work on the interstices of academic science and government bureaucracies.

The weekly minutes Bulletins of the AMA's Council on Pharmacy and Chemistry are on file at the archives of the AMA in Chicago, and are excellently indexed and thorough in reporting the daily workings of the Council. Relatively few additional manuscript collections of key Council members have turned up, and some, such as Torald Sollman's papers in the Cleveland Health Sciences Library, are surprisingly thin on these matters.

For my accounts of rational therapeutics and medical experimentation, I have relied heavily on manuscript collections of academic physicians. The collection at the Francis Countway Library of Medicine at Harvard Medical School, under the direction of Richard Wolfe, has become one of the central repositories for studying 20th century medicine, if not because of the eminence of the Harvard faculty, then because Dick Wolfe has salvaged so many of their papers. Such work remains frustrating, because so many key figures still have not held on to their professional correspondence or made it available. The archives at New York Hospital retain the records of some Cornell Medical School faculty; other New York institutions (e.g. Columbia College of Physicians and Surgeons)

don't seem to hold on to their heritage. The collections at Johns Hopkins Medical School and the National Library of Medicine, while extensive, were less useful for the particular topics in 20th century medicine surveyed here.

The records of the National Institutes of Health and the Food and Drug Administration are divided between records donated to the National Archives (Record Groups 90 and 443 for NIH and Record Group 88 for the FDA), and records which remain in the custody of the agencies, although physically on deposit at the Federal Records Center in Suitland, Maryland. For records still owned by the agencies, which date generally from the 1960's for NIH and go back to the 1930's for the FDA, permission must be obtained in advance from the relevant records officer at NIH or the Freedom of Information Office at the FDA. In each case, the agency retains detailed records of what has been sent to Suitland, records which are also on file at Suitland. It is generally difficult to figure out if the records one seeks are available, and where they are without one or more provisional trip to these officers to consult the transmittal records and make an official request. Although individuals in both bureaucracies are extremely helpful, they are overworked, and not overly familiar with historical materials. Dig ye must.

The National Archives also contain the records of the Office of Scientific Research and Development (RG 227), the War Production Board and the Bureau of the Budget, each of which provided valuable if obscure information about the development of drug research. The National Archives collections on wartime, prewar and postwar research are supplemented by the extensive manuscript holdings of the National Research Council at the National Academy of Sciences. NAS and NRC groups were sufficiently engaged in enough projects in research and policy to make this well organized archive worth a check on one's list.

Statisticians, alas, do not seem to have papers or else hold them as too precious to part with. Neither the papers of Jerome Cornfield or of William G. Cochran shed much light on the activities of either individual with regard to therapeutic experiments. The papers of other prominent figures in these developments remain in unknown or in private hands.