TREATMENT OF EIGHT NIH CONSENSUS DEVELOPMENT CONFERENCES IN THE BIOMEDICAL LITERATURE

Grace M. Carter, Geoffrey M. Anderson, David E. Kanouse, John D. Winkler

September 1986

N-2504-NIH

Prepared for

The National Institutes of Health
The research described in this report was sponsored by the National Institutes of Health under Contract N01-OD-2-2128.

The Rand Publications Series: The Report is the principal publication documenting and transmitting Rand's major research findings and final research results. The Rand Note reports other outputs of sponsored research for general distribution. Publications of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.
TREATMENT OF EIGHT NIH CONSENSUS DEVELOPMENT CONFERENCES IN THE BIOMEDICAL LITERATURE

Grace M. Carter, Geoffrey M. Anderson, David E. Kanouse, John D. Winkler

September 1986

N-2504-NIH

Prepared for

The National Institutes of Health
PREFACE

This Note examines the treatment of eight consensus development conferences, sponsored by the National Institutes of Health (NIH), in the medical literature during a period of three or four years following each conference. The examination was carried out as part of a study of the NIH Consensus Development Program, sponsored by the National Institutes of Health. This Note was prepared under contract with the Office of Medical Applications of Research, National Institutes of Health.

This work should interest individuals and agencies concerned with health technology assessment and with the transfer of information about medical technologies to health care practitioners. It should also interest those concerned with scientific information systems and the dissemination of scientific knowledge.

Other publications based on RAND's study of the NIH Consensus Development Program include:


SUMMARY

In 1977, the National Institutes of Health (NIH) created the Consensus Development Program, administered by NIH's Office of Medical Applications of Research (OMAR), to improve the assessment and dissemination of information about medical technologies. This program brings together panels of scientists, medical practitioners, and informed laypeople to review research findings and competing viewpoints on selected technologies. Each panel's findings are presented in the form of a consensus statement. Consensus statements and supporting materials are disseminated through reports in medical journals and through publication of conference summaries to target groups that include practicing physicians and other health professionals, the biomedical research community, and the public.

As part of an evaluation of the consensus program, the research reported here examines how eight NIH consensus development conferences (CDCs) were discussed in the biomedical literature during the three or four years following each conference. We describe areas of publicly expressed agreement and disagreement with the conference recommendations and identify other papers that may have influenced physicians concurrently with the conference. We also report on the extent of coverage of each conference in the professional literature, the frequency and types of citations to conference findings, and the influence of conference findings and recommendations on later scientific and professional reports. These aspects of dissemination may ultimately affect physicians' practices. One likely route for such an effect is through a small set of physician "opinion leaders," who influence a much larger set of practicing physicians.

We examined the professional literature surrounding eight consensus development conferences: antenatal diagnosis, primary treatment of breast cancer, steroid receptors, estrogen use for postmenopausal symptoms, thrombolytic therapy, Pap smear, cesarean birth, and coronary artery bypass graft surgery (CABS). For each of these topics, we performed systematic online literature searches in medical and
scientific journals that address medical researchers and clinicians. First, we searched *Index Medicus* for descriptions and reports of each consensus conference ("primary articles"). Next, we searched the citation index of the Institute for Scientific Information (ISI) biomedicine database for "secondary articles" that cited the primary articles.

For each of the eight conferences, at least one of us read the collection of primary literature. We classified each article according to a coding scheme that characterizes the dissemination of the conference findings and their apparent effect. We noted where primary articles were published, their relationship to the conference (conference statements, articles based on presentations made at the conference, and commentaries), and whether individuals connected with the conference wrote the articles. For secondary articles, we noted how widely the primary articles were cited, the types of citing articles (e.g., research reports, literature review, and commentary), and the types of journals in which the citing articles appeared (i.e., basic science or clinical journals).

Consensus statements from seven of the eight CDCs were published in national medical journals; the only exception (antenatal diagnosis) received extensive coverage in appropriate specialty journals. Statements were also published in regional medical journals and in specialty journals. For three of the conferences, additional articles based on conference presentations were published in national medical journals or prominent specialty journals. For several conferences, commentaries (not direct products of the panel process) appeared in various specialty journals and in *Science*.

Statements for all eight conferences were published in journals with the potential to reach a large and appropriate audience, as were papers based on conference presentations when they were published at all. Commentaries had a more uneven distribution.

The number and pattern of citations differed widely from conference to conference. The conference on steroid receptors received the largest number of citations, followed by the one on coronary artery bypass graft surgery. The Pap smear conference received the fewest citations. Notably, those conferences that published proceedings (rather than just
the summary statements) -- steroid receptors, CABS, and estrogen use -- received dramatically more citations than did the other five conferences we studied. Our conclusion is that journal publication of conference proceedings enhances the visibility of the consensus conference to the relevant research community, increasing its potential effect on the reporting of subsequent research.

We also found that the types of articles and the kinds of journals citing the various conferences differed in interesting ways. Generally, articles reporting the results of original research accounted for most of the citations. Nearly all conferences were viewed by citing authors as having contributed in some way to the research literature in their field. Review articles, which clinicians may read to keep informed of advances in medicine, accounted for a substantial fraction of the citations to the primary articles. Antenatal diagnosis and thrombolytic therapy received a large proportion of review-based citations. Published commentary, which may highlight controversial issues, accounted for few citations to the conferences.

We observed large differences in conferences that published proceedings with respect to the type as well as the amount of citations received. Not surprisingly, supplementary conference materials (which are often reports of original research) receive recognition in research articles. However, these materials also receive numerous citations in review and commentary articles, which may be of greater interest to clinicians.

Our investigation of the kinds of journals in which citations appeared revealed sensible patterns according to topic. Most citations appear in specialty journals, followed by clinical journals with a national readership, basic science journals, and regional clinical journals. The type of journal covering a topic generally corresponded to the audience of interest; for example, the conference on steroid receptors, a biochemical topic, was cited primarily in basic science journals.

To summarize the effects of the conference (as recorded in the literature), we categorized the citations according to which aspect of the conference is cited and whether the context of the citation affirms the judgment of the panel or reflects continuing controversy over the
issue. All conferences except two showed some, usually small, evidence that the conference statement had been persuasive in at least some of its clinical recommendations. Citations to half the conferences also showed that controversy continued after the conference over at least some clinical recommendations. In two other cases, scientific controversies over the subject of clinical recommendations clearly continued in the literature, although these controversies were not discussed in terms of conference recommendations.

When conferences failed to produce lasting consensus on a controversy, the reasons for this failure appear to be tied to the state of science in the particular area. This makes generalization from our studies problematic, because no single factor explains many cases. Rather, the results in different cases are consistent with various standard hypotheses used to explain delays in the diffusion of innovation—including the absence of compelling scientific fact, difficulties in interpreting scientific facts, the state of competing technologies, and the absence of compelling need.

One conclusion does arise strongly from both the quantitative and qualitative analyses. Most citations incorporate specific points from a cited article either as background for current research or as part of an argument for a particular interpretation of findings. The point that is referenced must have scientific credibility if it is to serve this purpose. Thus, anything that enhances the scientific credibility of a conference recommendation will increase its potential effect on the literature. One way to increase both credibility and potential effect is to provide the rationale or supporting documentation for conference recommendations. Indeed, the three conferences in our sample that published speaker presentations have by far the highest numbers of citations per primary article. In addition to publishing supplements recording the conference's scientific presentations, it might be worthwhile to encourage panel members or the chairman to write articles reviewing the rationale for conference decisions.

Our results also suggest that when the rationale for conference decisions is clearly spelled out, the benefits of a conference may outlast the consensus on its clinical recommendations. In such cases, the conference may provide a useful framework within which the relevant
facts may be viewed—even facts discovered long after the conference. The clearest example of this is provided by the conference on estrogen therapy, whose clearly articulated rationale has been adopted by several subsequent reviewers, even though new evidence quickly made the conference recommendations themselves obsolete.

Finally, despite the short time period during which we tracked citations to these conferences, we find evidence that they prompted research into clinically important areas and provided a framework within which others could make clinical decisions or interpret new research findings.
ACKNOWLEDGMENTS

We are grateful to RAND colleagues Laurel Brodsley and Mark Chassin, who helped us review the literature on coronary artery bypass surgery, and to Diane Alexander, who supervised or carried out the numerous searches this project required. Sandra H. Berry and Robert H. Brook offered useful suggestions during the early phases of this research. Barbara Quint and Roberta Shanman provided superlative library assistance. Rick Eden furnished extremely helpful comments on early drafts of the entire manuscript and helped us shape the stylistically diverse sections into a (more or less) coherent whole. Richard Rettig reviewed the manuscript in its later stages and made many helpful suggestions. Linda Freeman, Vale Dry, and Dolly Hardy provided expert text-processing assistance throughout numerous drafts.
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREFACE</td>
<td></td>
<td>iii</td>
</tr>
<tr>
<td>SUMMARY</td>
<td></td>
<td>v</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td></td>
<td>xi</td>
</tr>
<tr>
<td>I.</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II.</td>
<td>METHOD</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Role of Citations</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Search Procedures</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Categorization</td>
<td>9</td>
</tr>
<tr>
<td>III.</td>
<td>ANTEnatal DIAGNOSIS</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Primary Literature</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Secondary Literature</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Conclusions</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>19</td>
</tr>
<tr>
<td>IV.</td>
<td>TREATMENT OF PRIMARY BREAST CANCER</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Primary Literature</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Secondary Literature</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Conclusions</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>31</td>
</tr>
<tr>
<td>V.</td>
<td>STEROID RECEPTORS IN BREAST CANCER</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Special Report in the <em>New England Journal of Medicine</em></td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Citations Concerning Clinical Correlates</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>The Conference Supplement in <em>Cancer</em></td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>Conclusions</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>43</td>
</tr>
<tr>
<td>VI.</td>
<td>ESTROGEN USE AND POSTMENOPAUSAL WOMEN</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Primary Literature</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Secondary Literature</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>Conclusions</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>63</td>
</tr>
<tr>
<td>VII.</td>
<td>THROMBOLYTIC THERAPY</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Primary Literature</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Secondary Literature</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Conclusions</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>74</td>
</tr>
<tr>
<td>VIII.</td>
<td>CERVICAL CANCER SCREENING: THE PAP SMEAR</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Primary Literature</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>Secondary Literature</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>Conclusions</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>Bibliography</td>
<td>81</td>
</tr>
<tr>
<td>IX. CHILDBIRTH BY CESAREAN DELIVERY</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>------------------------------------</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>Primary Literature</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>Secondary Literature</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>Conclusions</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>Bibliography</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>X. CORONARY ARTERY BYPASS SURGERY</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>Primary Literature</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>Secondary Literature</td>
<td>97</td>
<td></td>
</tr>
<tr>
<td>Circulation Supplement</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Conclusions</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Bibliography</td>
<td>102</td>
<td></td>
</tr>
<tr>
<td>XI. SYNTHESIS AND CONCLUSIONS</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>Overview of the Primary Literature</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>Patterns in the Secondary Literature</td>
<td>113</td>
<td></td>
</tr>
<tr>
<td>Other Measures of Influence</td>
<td>127</td>
<td></td>
</tr>
<tr>
<td>Conclusions</td>
<td>132</td>
<td></td>
</tr>
<tr>
<td>REFERENCES</td>
<td>135</td>
<td></td>
</tr>
</tbody>
</table>
### TABLES

2.1. Classification of Professional Literature .................................. 10

5.1. Citations to Cancer Supplement Articles on the Relationship Between Receptor Levels and Response to Hormone Therapy, 1981 through 1983 .............................................. 40

5.2. Citations to Cancer Supplement Articles on Other Clinical Correlates, 1981 through 1983 .................................................... 41

5.3. Citations To Cancer Supplement Articles On Analytical Methods 1981 Through 1983 .......................................................... 42

10.1. Citations to Circulation Supplement Articles, 1982 through June 1984 .............................................................................. 101

11.1. Classification of Primary Articles ............................................. 110

11.2. Types of Journals Publishing Consensus Statements and Conference Presentations ....................................................... 111

11.3. Published Commentaries on Consensus Conferences by Type of Journal ........................................................................... 113

11.4. Number of Citations to Primary Articles ..................................... 115

11.5. Comparison of Citation Rates to Conference Publications With Indicators of the Size of the Field ........................................ 116

11.6. Number of Secondary Articles by Type ....................................... 117

11.7. Number of Secondary Articles Appearing in Various Types of Journals ................................................................. 119

11.8. System for Classifying Conference Effects Found in the Citing Literature ................................................................. 120

11.9. Types of Conference Effects Found in the Literature ............. 122

11.10. Summary of Conference Effects ................................................. 123
I. INTRODUCTION

In 1977, the National Institutes of Health (NIH) created the Consensus Development Program to improve the assessment and dissemination of information about medical technologies. This program brings together panels of scientists, medical practitioners, and informed laypeople to review research findings and competing viewpoints on selected technologies. Each panel's findings are presented in the form of a consensus statement. Consensus statements and supporting materials are disseminated through reports in medical journals and through publication of conference summaries to target groups that include practicing physicians and other health professionals, the biomedical research community, and the public.

To assess the effectiveness of the Consensus Development Program, NIH has contracted with The RAND Corporation to conduct a study of how consensus conferences have affected the knowledge, attitudes, and practices of health care professionals. The study analyzes data from several sources, including a survey of physicians' knowledge, attitudes, and practices concerning selected technologies, and a longitudinal audit of hospital records aimed at detecting changes in practices over time. The specific goals of this assessment are:

1. To analyze the consensus conference statements with respect to both type and content, as a basis for understanding their potential and actual effect;
2. To trace the process by which consensus conference conclusions have disseminated into the medical and research communities and to the public at large;
3. To assess physicians' knowledge and attitudes concerning selected medical technologies that have been the subject of recent conferences, to assess physicians' awareness of conference findings, and to determine the degree of correspondence between conference recommendations and physicians' practices;
4. To measure changes over time in selected hospital-based practices that have been the subject of consensus conferences;

5. To identify other sources of information about these technologies that may have a concurrent effect on knowledge, awareness, and practice; and

6. To identify and measure the relative importance of physicians' major sources of information about new technologies, with the goal of understanding more about the dissemination process.

This Note reports work on the second and fifth goals. We describe how eight Consensus Development Conferences (CDC) were discussed in the medical and scientific literature.\(^1\) This review provides qualitative information that is essential for interpreting the findings of other parts of the study that directly assess physicians' knowledge and practice habits. It identifies areas of publicly expressed agreement and disagreement with the conference recommendations, and it identifies other papers that may have affected knowledge, attitudes, and practice concurrently with the conference.

Further, the study reports on the extent of coverage of each sample conference in the professional literature, on the characteristics of the scientific and clinical community participating in the discussion, and on the role played by the conference's findings and recommendations within the process of scientific communication on the topic.

Section II describes our methods for locating the relevant literature and categorizing its treatment of consensus conferences. The eight case studies of how consensus conferences are discussed in the literature form the core of this Note (Secs. III through X). The final section (Sec. XI) synthesizes the findings across conferences and relates the findings to conference management and dissemination strategies.

\(^1\)Winkler et al., 1986, describe how these conferences were discussed in the popular press. References to methodology will be found on p. 139; literature references to specific conferences will be found at the ends of the relevant sections.
II. METHOD

A major purpose of our study is to assess how well consensus conference findings have been disseminated to various target audiences, which include practicing physicians, the biomedical research community, and the public. Professional journals are an important information channel for two of those audiences. For that reason, publication of consensus conference findings—and subsequent published references to those findings—are an important part of the dissemination process. The professional periodical literature also records (and broadcasts) how the professional community responds to the conferences. Thus, it can be used to trace both the dissemination process and audience response.

Because we are interested in the effect as well as the extent of dissemination, the analysis we report here uses qualitative as well as quantitative methods. We examine the breadth and depth of the coverage of the conference, the frequency of citations to conference publications, and the frequency with which authors discuss a controversy concerning conference recommendations.

We analyzed the literature for the eight CDC topics that are covered by either the physician survey or the audit of hospital-based practices in our larger study. These are: antenatal diagnosis, primary treatment of breast cancer, steroid receptors in breast cancer, estrogen use in postmenopausal women, thrombolytic therapy, Pap smear, cesarean delivery, and coronary artery bypass graft surgery (CABS).

Some of our data derive from the scientific practice of citation. In particular, we use citations to help identify the literature that discusses each conference and to measure how much each conference has affected the literature. This strategy reflects our understanding of how successful conferences may shape scientific debate and how citations record patterns of scientific influences.
ROLE OF CITATIONS

There is a long and varied literature that discusses the role of citations in scientific writing;¹ we will not endeavor to review it here. Rather, we shall briefly discuss how conference statements might be expected to influence the literature, why we believe that tracing citations should enable us to discover such influences, and the limitations of the methodology.

The consensus conference's main purpose is to help translate current scientific knowledge into medical practice, not to advance the state of science itself. In keeping with this goal, the conference statement is aimed primarily at clinicians rather than the biomedical research community. However, the research community is an important secondary audience; and because the consensus conference acts in many ways as a research conference, it might also bring about any of the following:

1. An evaluation of the scientific evidence may result in a changed consensus about the scientific facts;
2. A review of theory and empirical evidence may lead to a new scientific model for interpreting scientific fact or a greater consensus about what the best model is;
3. A review of theory and of the known risks and benefits of various clinical interventions may lead to a new or more completely specified hypothesis about the relationship between intervention and clinical outcome, or a greater consensus about what that relationship is;
4. The evaluation of evidence and the relationship of evidence to clinical problems may uncover new research problems or underscore the importance of existing ones.

Each of the above may come about as a result of greater consensus on the issues. However, when the facts are ambiguous or when the existing

¹See, for example, Garfield, 1970; Moravcsik and Murugesan, 1975; and Gilbert, 1977.
models do not correspond very well to the empirical findings, detailed examination of the issues at the conference may result in increased dissension rather than consensus. (This should not be viewed as reflecting conference failure. In some cases, scientific thought needs to be turned around before further progress can be made.)

If consensus conferences do affect scientific thought, this might lead to an indirect effect on medical practice. One likely route for such an effect is through an influence on the thinking of those whom Coleman, Katz, and Menzel (1966) call "opinion leaders," who in turn influence their colleagues. Another route is through articles that review the meaning of scientific findings for medical practice.

Some of these effects may be detected by examining articles that refer to conference publications. Scientific authors cite other articles for many reasons. Among the most frequent are:

- As a shorthand way of incorporating the content of the referenced article into the citing article. This is probably the way that is most frequently used to describe experimental methodology. It is also often used in reference to theories and to the clinical relevance of scientific problems;
- To acknowledge the "owner" of scientific "property;"
- To increase the credibility of the citing article by appealing to authority or to evidence for certain propositions documented elsewhere;
- To correct the scientific record by disputing a statement appearing in the cited article.

Insofar as a conference synthesizes or validates new knowledge about medical science or medical practice, one might expect to find citations that serve any of the first three purposes listed above. Citations in the fourth category are somewhat different, in that they reflect disagreement with the citing article. The number of citations in this category may often understate the extent of disagreement, since disagreements are frequently not explicitly discussed in the literature; rather, those who disagree with the author are ignored in the reference
list (see MacRoberts and MacRoberts, 1984). However, our methodology was surprisingly successful in turning up various public disagreements.

Because our search focuses on peer-reviewed professional journals, our study is not well suited to the discovery of influence that proceeds through other routes. It is possible, for example, that a conference's influence might bypass this literature and instead show up directly in either "throwaways" or text books. As is discussed in more detail in Sec. XI, we did examine two major internal medicine text books and found no mention of any of our eight sample CDCs. Although a more exhaustive search might have uncovered more evidence of influence, results from a study we are preparing on a national survey of physicians suggest that physicians' main sources of printed information about consensus recommendations are journals and reports from NIH itself.

We have used citations here to objectively trace and measure the effects of the CDC on both the scientific literature and the indexed clinical literature. The latter offers physicians information on new scientific findings that may influence their practice.

The influence of a primary article on the subsequent literature is best measured by the number of citations it receives in relation to the size of the associated published literature. An article that receives several citations in a field of high publication activity may reach only a small fraction of the potential audience and therefore be less influential than one receiving a smaller number of citations in a less actively researched field.

We use two approaches to estimate the relative sizes of the literatures concerning each topic. First, we use the average citation rate for all articles in the medical field from which each topic is discussed. However, because each CDC focused on a more specialized topic within this larger medical field, our second approach defines the field more narrowly. Using the bibliographies of the secondary articles on each topic, we selected a review article covering the same topic as the consensus statement and examined the science citation index to see how frequently this "benchmark" review article was cited. To ensure that the benchmark review articles would be comparable to primary articles from the consensus conference, we limited benchmark articles to clinical reviews published in a journal of national distribution about
the same time as the primary article. To ensure comparability of citation counts, we counted only those citations that appeared within three years following publication of the benchmark review article.

SEARCH PROCEDURES

For each of these topics, we conducted a literature search in medical and scientific journals that address medical researchers and clinicians (the professional literature). We generally used systematic online searches; but on occasion, various topics warranted individualized treatment.

Our search involved two steps. First, we searched for articles describing the consensus conferences or reporting their findings. These are referred to in the following discussion as primary articles. Each CDC produced a Consensus Development Conference Summary Statement; these were reproduced in journals as well as by NIH. Certain conferences also commissioned papers, presentations, or task force reports that served as background for the consensus panel's deliberations; these could also be reprinted in journals. Statements and reference materials are direct products of the CDC and are easily defined. In some cases, NIH officials produced commentaries or supporting documents for the CDCs. These again can be taken as direct products of the consensus process. We also included in the primary literature other articles on the CDCs when they were published more or less concurrently with the direct products and when their primary purpose was to report and comment on the conference or its findings. Such commentaries were produced by panel members as well as others who attended the public meetings or who were exposed to the panel statements. For example, *Science* and *JAMA* frequently covered conferences and summary statements in great detail.

Primary articles were initially sought through an online search of *Index Medicus* using "consensus development" as the key words and covering the period between 1977 and 1982. Printed copies of *Index Medicus* were also examined to locate articles describing the results of the conferences, using the subjects of the conference (e.g., "breast

---

2 *Index Medicus* accesses the worldwide biomedical literature, including 3000 journals published in the United States and in 70 other countries covering research, clinical practice, administration, health care services, and policy issues.
cancer") as index terms. The primary literature for each conference consisted of all conference-related articles found by this procedure, along with the consensus statements and any task force reports or other conference-related materials printed and distributed by NIH itself.

As a second step, we searched the professional literature for articles containing references to the reports of consensus conferences and their findings. These articles, referred to as secondary articles in subsequent discussion, were found by searching the citation index of the Institute for Scientific Information (ISI) biomedicine database. The time periods examined depended on the date of the conference and on the number and variety of primary articles found. In general, we searched for citations to primary articles through the end of 1982. However, when supplements to conferences were published later (as was the case for steroid receptors and coronary artery bypass surgery), citation searches were extended through 1983. Other citation searches extended past 1982 on occasion, reflecting the iterative nature of the online searches and the changing parameters of the databases.

We experienced little difficulty in locating citations to primary articles that had a personal author (most conference presentations and many commentaries). Citations to articles with an institutional author had to be searched somewhat differently, however, because there is no widely accepted convention as to how such articles should be cited. A search keyed to the name of the journal and the volume and page number worked well in locating citations to published articles. This method could not be used for other types of primary literature, such as task

---

3ISI/BIO/MED document citations made to other articles in 1400 major biomedical journals published throughout the world.

4An article based on a consensus conference sponsored by NICHD, for example, might be cited as National Institutes of Health, U.S. Public Health Service, National Institutes of Child Health and Human Development, Department of Health and Human Services, or a combination of the above, with or without abbreviation. It might also be cited by referring to the name of a conference panel—e.g., "Cesarean Delivery Task Force."

5In the BIO/MED database built by ISI, all citations to articles by any institutional author are entered as if the author were "anonymous." Thus it was possible to key a volume, year, and page number when these were known.
force reports or the versions of the consensus statements published separately by NIH. For these, we used a combination of methods, including searching for key words (consensus development in combination with one or more words referring to the conference topic) and searches for the most likely forms of reference to the institutional author. These methods were not altogether successful, however. We doubt that we found all the important secondary articles referring to NIH consensus findings by citing printed material from NIH.

The time frame introduces another limitation into our study. The period during which we searched for citations (three to four years) is short compared with the amount of time required to mount a new research effort, obtain funding, conduct the research, and publish the findings. Thus we cannot fully assess the response of the scientific community to research recommendations from the CDCs. Nonetheless, we have observed several cases in which there was an apparent positive response and other cases in which studies ongoing at the time of the CDC appeared to satisfy the need for the research recommendation.

CATEGORIZATION

Each article retrieved from the professional literature was classified according to a coding scheme. For primary articles, we noted where consensus statements and reports of conference findings were published and whether NIH officials, members of NIH task forces, or CDC panel members wrote the articles. For secondary articles, we noted the types of articles that cited consensus conferences and how the citation was used. For both categories of articles, we noted the types of journals in which the articles appeared.

To categorize the journals in which articles appeared, we first distinguished among journals oriented toward basic science, journals oriented toward particular medical specialties, and general medical journals (Narin, 1976, provides this categorization). We then further subdivided the general medical group to distinguish journals that are national in scope (e.g., JAMA) from those that cater to a regional audience (e.g., Tennessee Medical Journal).\(^6\)

\(^6\)Despite its name, we classified the New England Journal of Medicine as a national general medical journal.
Table 2.1 lists the attributes coded for each article. This information allows us to summarize dissemination of the conference findings in the professional literature in several ways. The amount of coverage of the CDCs is indicated by the number and type of journals and articles reporting or using CDC findings. The breadth of the scientific and clinical community participating in the discussions is revealed by the number of individuals contributing to the literature, the variety of article formats and journals involved in the discussions, and the relationships (if any) of authors to the CDC. We also coded how the conference findings were cited--as a background reference for scientific research, as an accepted source of scientific findings, or as a guide to clinical practice. The degree of acceptance or controversy is reflected in whether citing articles agree or disagree with scientific findings or clinical recommendations made in the consensus statement.

Table 2.1

CLASSIFICATION OF PROFESSIONAL LITERATURE

1. Journal name and year

2. Authorship
   a. CDC panel member or conference speaker
   b. Other author

3. Journal type
   a. Basic science
   b. National/general medical
   c. Specialty medical
   d. Regional medical

4. Article type
   a. Research report
   b. Review article (clinical or basic science)
   c. Commentary (including editorials and letters)

5. Nature of citation (secondary literatures only)
   a. Background for research
   b. Scientific findings (agree or disagree)
   c. Clinical recommendations (agree or disagree)
III. ANTENATAL DIAGNOSIS

The late 1960s and the 1970s were marked by widespread development and application of techniques to make in utero diagnoses. These techniques fall into three broad categories: (1) those used to diagnose hereditary and congenital defects early in pregnancy, (2) those used to determine the level of fetal maturity late in pregnancy, and (3) those used to detect signs of fetal distress during labor and delivery.

Controversies over the safety and efficacy of these techniques persisted into the late 1970s. Questions were raised regarding the safety of first trimester amniocentesis and the reliability of various measures of fetal maturity. However, the most pressing question centered around the use of electronic fetal monitoring (EFM), a widely used but largely unproven technology.

These concerns led to the decision to hold a Consensus Development Conference on Antenatal Diagnosis. The conference, held on March 5, 1979, was divided into panels devoted to the study of the efficacy and safety of three obstetric techniques:

1. Predictors of hereditary disease and congenital defects. These include mid-trimester amniocentesis, fetoscopy, alpha fetoprotein measurement, and ultrasound;
2. Predictors of fetal maturity, including third trimester amniocentesis and ultrasound;
3. Predictors of fetal distress, including monitoring during labor.

With respect to the three technologies that could be used to predict hereditary disease and congenital defects, the panel concluded:

- Amniocentesis is a widely accepted clinical practice that should be offered to pregnant women who are at high risk (the panel clearly defined women at high risk);
• Ultrasonography is recommended as a technique for determining the position of the fetus and placenta before performing amniocentesis, but more clinical research is required before ultrasound can be relied on as a diagnostic tool in its own right;
• Fetoscopy and fetal blood sampling remain experimental and are not ready for general use at this time;
• Measurement of alpha-fetoprotein in amniotic fluid is accepted practice for detecting neural tube defects in selected pregnancies where such defects are likely to occur; however, measuring AFP in maternal serum cannot be recommended as a device for screening large numbers of pregnancies until pilot studies prove its efficacy.

With respect to predictors of fetal maturity, the panel concluded:

• Appropriate use of ultrasound and third trimester amniocentesis can eliminate the problems of prematurity and respiratory distress syndrome following scheduled cesarean delivery;
• The long-term effects of both techniques should be studied further;
• Ultrasound should not be used routinely for all pregnant women;
• Both ultrasound and third trimester amniocentesis should be used only when indicated: in high risk pregnancies, in cases where the woman has had a previous cesarean delivery, and in other cases where the physician needs information on fetal maturity;
• The L/S ratio, a measure of the relative phospholipid concentration in the amniotic fluid, is a more reliable measure of fetal lung maturity than the "shake test;"
• Postgraduate training in amniocentesis and ultrasonography should be provided to physicians who have no experience with these procedures;
There is a need for more accurate indicators of lung maturity;
Standards are needed for the manufacture of ultrasound machines
and for both technical and professional proficiency in doing
ultrasonography and amniocentesis.

With respect to predictors of fetal distress, the panel concluded:

- The use of EFM should be strongly considered in high risk
  situations, including: expected low birth weight, premature,
  postmature, or intrauterine growth-retarded babies; women with
  medical complications of pregnancy; the presence of meconium in
  the amniotic fluid; and situations where abnormal fetal heart
  rate is detected with a stethoscope.
- In the absence of identifiable risk factors, periodic
  auscultation of the fetal heart rate is an acceptable way to
  assess fetal condition.
- EFM should not be a substitute for clinical judgment.
- All hospitals and birthing centers providing maternity care
  should have the necessary trained staff and equipment to assess
  the status of each fetus during labor and to take appropriate
  action.
- Fetal scalp blood pH determination should be used as an adjunct
  to EFM.
- EFM should be carried out in a way that minimizes risk,
  including careful placement of the fetal scalp electrode and
  intrauterine pressure catheter, avoidance of unnecessary
  restriction of maternal mobility or prolonged supine position.
- Artificial rupture of the amniotic sac solely for internal
  electronic fetal monitoring should not be done on a routine
  basis.
- Additional research is needed on the effects of hypoxia on
  fetus and newborn, the identification of risk factors for fetal
  distress, the efficacy and risk of EFM in high risk patients,
  and the development of noninvasive methods of fetal monitoring.
The recommendations of the consensus conference were published in a summary statement that covered all three obstetric techniques and in task force reports that covered each technique separately. The recommendations were widely reprinted in the appropriate professional journals. With one minor exception the recommendations were published in the context of task force report summaries rather than consensus conference summaries. The bulk of the secondary literature dealt with the topic of predictors of intrapartum fetal distress.

PRIMARY LITERATURE

This conference and the conference on childbirth by cesarean delivery differed from the other conferences in the publication of task force reports along with consensus development conference statements. The reports summarized the issues related to each of the topics and made recommendations. Drafts of these reports were circulated to a wide audience before the panel meeting, and many individuals submitted written statements to the panel. After the panel had met, task force members reviewed and in some cases revised the recommendations made in the draft reports. The final task force report was published in August 1979, four months after the conference.

We located 15 primary articles. Only one of these (Connecticut Medicine, 1979) was a reprint of the overall CDC statement, while another seven were summaries or edited summaries of the various task force reports. The remaining seven can be classified as commentaries.

The summary and conclusion sections of the task force report on the predictors of intrapartum fetal distress were published concurrently in three journals (Journal of Reproductive Medicine, 1979; Journal of Pediatrics, 1979; American Journal of Obstetrics and Gynecology, 1979).

Benjamin K. Silverman, a member of the task force on predictors of intrapartum fetal distress, edited a series of four articles in Clinical Pediatrics. These articles presented the summary and conclusions from the task force reports. These articles were not exact verbatim summaries but contained comments and tables or figures that supported the task force statements.
The *Pediatric Annals* published a series of three articles that commented on predictors of fetal maturity and predictors of hereditary disease and congenital defects. These commentaries were prepared for the journal by the National Institute of Child Health and Human Development.

Four articles in two national journals discussed the conference proceedings and commented on the panel reports. The *British Medical Journal* published two of these commentaries. The first (Culliton and Waterfall, 1979a) deals with mid-trimester amniocentesis. After briefly reviewing the evidence on the sensitivity, specificity, and safety of mid-trimester amniocentesis, the authors suggest that there may be some problems as the use of the technology is expanded. These difficulties fall into two major classes: (1) as the use of the technology is expanded to centers not experienced in its application, the risks (abortion, bleeding) may be higher than noted in the original experimental centers; and (2) because few centers can perform analyses of amniotic fluid, they will not be able to meet the demand created by the widespread use of mid-trimester amniocentesis.

The main thrust of this article is that the NIH conference provides what amounts to a U.S. government seal of approval for amniocentesis as a mainstream technology. It also notes that the NIH "somewhat reluctantly" had moved into the area of consensus development and that consensus building differs greatly from regulating technologies.

The second article (Culliton and Waterfall, 1979b) deals with fetal distress and fetal maturity. The authors suggest that the panel approved EFM more equivocally than mid-trimester amniocentesis and conclude that the "lack of precision" in the panel statement regarding EFM is a consequence of the lack of a clear definition of fetal distress and its relationship to fetal monitoring and adverse fetal outcomes.

The discussion of fetal maturity highlights the link among premature delivery, respiratory distress syndrome, and neonatal deaths. It also states that the panel was much more direct in its recommendations regarding tests of fetal maturity than for EFM.
The authors also reiterate their point in the previous article that although the CDC panels were not designed to arbitrate or regulate modes of practice, the recommendations will soon be seen as "setting standards."

_JAMA_ published the other two commentaries. The first (Check, 1979a) deals with hereditary disease and fetal maturity, the second with EFM (Check, 1979b). Both articles briefly review the evidence examined by the panel and then describe the discussion that ensued. The articles report the content of discussions. The main comment is that although there tended to be agreement on the recommendations regarding hereditary disease and fetal maturity, "few people were satisfied with all the conclusions of the draft report" on EFM.

**SECONDARY LITERATURE**

We located 22 articles that refer to the primary literature. Four of these deal with hereditary disease and congenital defects (especially with the use of amniocentesis), four with fetal maturity, and 14 with fetal distress and the use of EFM. The greater effect of the findings on EFM probably reflects the rapid increase in and widespread use of this technique. It may also reflect the nature of the EFM report, which does not so much define clinical policy on the use of EFM as indicate the need for greater scientific knowledge of this widely used technology.

Seven of the 14 articles citing the CDC findings on EFM do so in the course of explaining the background for their own research, suggesting that the CDC may have had some role in prompting or shaping that research. It is also possible, of course, that the CDC findings are cited as justification for research that was undertaken for other reasons.

**Hereditary Disease and Congenital Defects**

Four papers refer to the primary literature on hereditary disease and congenital defects. Bantock, who was an author of a British Medical Research Council (MRC, 1978) study of the risks of amniocentesis that was dismissed by the CDC, argues that the MRC study's large sample size
afforded a more sensitive measure of risks than did the NICHD study on which the consensus panel relied (NICHD, 1979a-d) and that the MRC study's results also point to larger risks than the panel suggested (Bantock and Sutherland, 1979). Duff et al. (1981) is a review of antenatal diagnosis citing both the consensus conference and the MRC study; it reports risk levels suggesting that amniocentesis is fairly safe, consistent with the overall sense of the Consensus Statement. Both Duff et al. (1981) and McCormack, Breslin, and Coppola-McCormack (1980) cite the conference as a source of epidemiological data on birth defects. Ajl (1982) cites the conference as indicating a prominent role for fetal diagnosis in the future.

Both Bantock and Sutherland's (1979) article and Ajl's cite the Culliton and Waterfall (1979a,b) article as the source of their statements. Duff et al. (1981) and McCormack, Breslin, and Coppola-McCormack (1980) cite the Clinical Pediatrics papers, the list also citing the task force report.

Fetal Maturity

Four secondary articles are related to the topic of fetal maturity. Two cite the Clinical Pediatrics article on the topic (Thieme et al. 1983; Morrison, Douvas, and Gookin, 1982). The other two deal with the role of fetal maturity studies in cesarean deliveries. One (Shy, 1981) cites the NIH summary itself and points out that measuring fetal maturity should minimize iatrogenic prematurity and respiratory distress syndrome. Bottoms, Rosen, and Sokol (1980), who cite the task force report, also use the reference to support their contention that improved diagnosis of fetal maturity would help prevent premature delivery.

Fetal Distress and EFM

Of the 14 secondary articles dealing with EFM, 10 cite the task force report summary in the American Journal of Obstetrics and Gynecology, one cites the task force report in the Journal of Reproductive Medicine, and two cite the task force report itself. A letter in the British Journal of Medicine is the only citing article in the secondary literature on EFM that does not cite either the task force report itself or the task force report summaries.
Seven of the secondary articles present original research. In most of these studies the report is cited as background for the research. Six secondary articles review recent literature on EFM. One of these (Ingemarsson, 1981) notes that the report concluded that EFM may not be beneficial in low risk pregnancies.

Banta and Thacker (1979) produced an extensive review of the costs and benefits of EFM, noting that the report recommends that periodic auscultation and EFM are acceptable methods of monitoring low risk pregnancies. They are critical of the report's conclusion that EFM is the modality of choice in high risk pregnancies. Their criticism is based on the lack of a rigorous definition of high risk.

CONCLUSIONS

This conference produced a broad range of primary literature. Although only one summary of the conference conclusions was published (in a regional medical journal), verbatim task force reports or edited task force reports and various commentaries on the conference proceedings and conclusions yielded a body of additional material that appeared in a range of prominent specialty journals as well as major national medical journals. Although clinicians were unlikely to be exposed to the conference summary statement per se, many were probably exposed to the conference's conclusions through their encounters with summaries of the task force reports and commentaries on the conference.

The subject receiving the most attention in the primary literature was the prediction of intrapartum fetal distress, especially the role of electronic fetal monitoring. All three of the verbatim task force report summaries dealt with this subject, which was also emphasized by the commentaries in the national medical journals.

References to fetal monitoring also dominated the secondary literature. Of the 22 secondary articles, 14 dealt with fetal monitoring. Of those, 10 referred to the American Journal of Obstetrics and Gynecology task force summary.

Fetal monitoring is a widely used technique whose efficacy has been questioned and debated; the effect of the discussion of fetal monitoring on the primary and secondary literature could be a consequence of this interest.
Controversy on the risks of third trimester amniocentesis continued after the conference, with the task force report providing a focal point for at least some of the subsequent discussion.

The conference on antenatal diagnosis was interesting for two themes that developed in the primary literature: (1) that consensus conference statements might lead to government regulation of medical decisionmaking, and (2) that the conference statement was not specific enough in some of its recommendations. These themes suggest an underlying tension in the goals of the CDC dissemination effort that can make it difficult for a panel to produce a conference statement acceptable to all constituents within the medical profession. The statements are intended to inform and educate physicians, not to dictate practice. Yet to do an effective job of informing and educating, statements must offer specific guidelines, which may strike some as excessively prescriptive.

The conference was frequently cited in the secondary literature as justifying the author's subsequent research into the use of fetal monitoring. Inasmuch as both the task force summary and the Consensus Statement itself emphasized the need for such research, it is perhaps not surprising that they would frequently be cited in support of the topic's importance. It is possible the conference may have encouraged additional research, but, given the time that typically elapses between the initiation of research and publication of findings, we are unlikely to observe the results in near-term citation patterns.

BIBLIOGRAPHY

Primary Literature

A. Statements and Summaries


Because of the topical organization, some references appear more than once.

Only published summary of the consensus statement.

B. Summaries of Task Force Reports


First of a series of four edited summaries of the task force reports. Introduction includes a discussion of the consensus process and the effect of consensus recommendations. This article discusses echosonography, fetoscopy, and alpha-fetoprotein assay.


Second in series, featuring a discussion of amniocentesis.


Third in series, featuring a discussion of fetal maturity.


Fourth in series, deals with fetal distress.


Summary and conclusions of task force report on the predictors of intrapartum fetal distress. The most commonly cited primary article.


Summary and conclusions of task force report on the predictors of intrapartum fetal distress.

Summary and conclusions of task force report on the predictors of intrapartum fetal distress.

C. Commentaries


First in a series of three papers prepared by the NICHD. The papers review the consensus conference and task force report. The first paper deals with amniocentesis.


Second in series, deals with ultrasound, fetoscopy, and measurement of alpha-fetoprotein.


Third in series, deals with predictors of fetal maturity.


Commentary on amniocentesis and predictors of fetal maturity.


Commentary on electronic fetal monitoring.


Commentary on amniocentesis conference and task force report. Begins with a discussion of the consensus process and the assessment of medical technology.

Second in series of commentaries. This one deals with fetal distress and fetal maturity.

Secondary Literature

A. Hereditary Disease


Letter commenting on Culliton and Waterfall's (1979a) article.


A review of antenatal diagnosis of fetal abnormalities. Includes a set of indications for antenatal diagnosis and several references to the conference for epidemiological data.


B. Fetal Maturity


A review of fetal maturity study methods; mentions the conference recommendation for the use of laboratory tests.


C. Fetal Distress


Extensive review of EFM that is somewhat critical of task force recommendations.


Original research.


Original research.


Report on original research.


A letter that argues with Culliton and Waterfall's (1979b) conclusion that EFM has not been a major factor in the increasing cesarean section rate in the United States.


Original research.


Extensive research review.

Original research.


Original research.


Clinical review.


Clinical review.


Clinical review.


Clinical review.


Original research that discusses several of the clinical recommendations.

**Related Studies**


IV. TREATMENT OF PRIMARY BREAST CANCER

Dr. W. S. Halsted introduced what became known as the Halsted radical mastectomy for breast cancer during the 1890s. This procedure involves removal of the breast, axillary lymph nodes, and sections of the chest wall, including the pectoral muscles. It was a striking success and became the clear treatment of choice for over 60 years.

During the last 25 years, however, several influences have led some oncologists to the conclusion that a less radical treatment may be preferable. One of these influences was an evolution in the theoretical concepts of tumor biology. During Halsted's era, it was thought that tumors spread contiguously. The Halsted radical was designed to prevent disease recurrence by removing all the regions that were likely to harbor the tumor. More recently, some oncologists have come to believe that disease recurrence is more likely to be due to systemic dissemination before any therapeutic intervention and therefore that the removal of extensive tissue beyond the tumor is not effective therapy.

A Consensus Development Conference on "The Treatment of Primary Cancer: Management of Local Disease" was held at NIH on June 5, 1979. It was sponsored by the National Cancer Institute.

The conference addressed the following question:

• Are there clinical alternatives to radical mastectomy that minimize patient morbidity and do not decrease a patient's survival potential?

The discussion emphasized three categories of surgical techniques:

1. Radical mastectomy;
2. Total mastectomy, which removes axillary lymph nodes but not pectoralis muscles;
3. Lesser surgical procedures such as segmental mastectomy.
The panel also discussed radiotherapy as a primary treatment of local breast cancer.

Very briefly, the panel concluded that total mastectomy with axillary dissection should be the current treatment standard for women who have Stage I and selected Stage II breast cancer. The panel also recommended a two-step procedure in which a diagnostic biopsy precedes the discussion of therapeutic alternatives with the patient.

The panel produced a short Consensus Statement, "The Treatment of Primary Breast Cancer: Management of Local Disease," which summarizes its conclusions and recommendations concerning both clinical practice and research needs. The Consensus Statement was reprinted in complete form by one national and one regional journal. Several other journals--JAMA, a nursing journal, and Science--summarized the results.

However, the panel members were unable to reach a full consensus on the state of science in the area. Three groups of panel members published separate views on the conference, which are in effect minority reports. Moreover, the secondary literature shows continued controversy concerning the CDC's recommendations. Several authors took issue with the recommendation for a two-step procedure and with the CDC endorsement of the National Surgical Adjuvant Breast Project (NSABP) protocol.

**PRIMARY LITERATURE**

We located eight articles reporting the results of the Breast Cancer CDC, including several by NIH representatives or CDC panel members. Five are reports of the CDC; the remaining articles were written by conference participants.

The entire Consensus Statement was published in *New England Journal of Medicine* (NEJM) in August 1979 and in *Connecticut Medicine* in September 1979. The "From the NIH" column of JAMA, which reported on the conference a year after it was held, repeated the major conclusions and recommendations (both clinical recommendations and the need for further clinical trials of alternative treatments). A nursing journal (AORN Journal, 1979) described both the conference and panelist Fisher's comments on the conference. A brief report of the conference proceedings and major recommendations also appeared in Science (Holden, 1979).
Three groups of conference panel members chose to publish their separate views on the conference. One of these (Moxley et al., 1980) endorses the positions taken in the Consensus Statement and explains the reasoning behind them. The other two take up stands on either side of the panel position. Because of the public controversy, the silence of the other five panel members (Durant, Hellman, Kushner, Pierquin, and Veronesi) leaves their positions in doubt.

The middle ground is stated in a JAMA article by Moxley et al. (1980) who provide a long and very detailed discussion of the scientific evidence behind the conference recommendations, with which they are in clear agreement. They review the results and status of recent clinical trials and come to the following conclusions. (1) Total mastectomy with axillary dissection should be the treatment standard. (2) The ideal management of breast tumor would include a diagnostic biopsy followed by a discussion of therapeutic alternatives with the patient. (3) Lesser surgical procedures should be performed only within the confines of prospective clinical trials. But this article is signed only by Moxley (the panel chairman) and the three NCI participants on the panel (Allegra, Henney, and Muggia).

Urban (1980), taking the traditional surgical position, dissents strongly from the Consensus Statement. His minority report argues the following:

1. The recommended two-step procedure is unnecessary in the majority of cases and potentially dangerous;
2. "To initiate a study evaluating the effectiveness of partial mastectomy without radiation therapy" (which is the treatment given to one cohort in the NSABP protocol endorsed by the Consensus Statement) "is unlikely to add to our information or to benefit the patients undergoing this trial;"
3. The recommendations that the total mastectomy with axillary dissection should be the treatment standard should apply only to "Stage I cases and early selected Stage II cases."

(Emphasis added; the word "early" is an addition to the conference wording.)
Urban also objects that the Moxley article omits factual data presented at the CDC meeting.

At an opposite extreme, Fisher (1979) who was heavily involved in the NSABP since its inception, appears to go further than the Consensus Statement in his endorsement of breast conserving operations.

The disagreement between Fisher (1979) and Urban (1980) centers on two related scientific issues:

1. Whether the multicentricity of breast cancer is a sufficient deterrent to the performance of breast-conserving operations;
2. Whether the disease spreads by means of a local or systemic process.

Fisher states that "by repudiating the radical mastectomy, the consensus statement, wittingly or unwittingly, also rejects the principles that had provided the scientific basis for the operation. For that reason, if no other, the report is of singular importance in the annals of oncology." Urban accepts this interpretation of the statement's content but disagrees with it; he believes that the disease is frequently not systemic. This theoretical disagreement leads directly to disagreement on whether axillary dissection is of therapeutic value.

SECONDARY LITERATURE

We found 20 articles citing the primary literature on the conference. They are located primarily in specialty or general interest journals. We analyze the content of the citations in two groups of roughly equal size, dealing first with articles citing the CDC findings and then with articles citing the minority reports by panelists Moxley et al., Urban, and Fisher.
Citations to the Consensus Statement

Nine articles cite the Consensus Statement or its reprints and summaries.

The CDC's view that total mastectomy has superseded radical mastectomy as the treatment standard appears to reflect current professional opinion. Three articles reporting separate studies repeat the consensus panel's statement that "the current treatment standard" is no longer the Halsted radical as part of the background for their work in evaluating treatments that result in even greater tissue conservation (Goldenberg, Prosnitz, and Peters, 1980; Ray et al., 1983; Weissberg and Prosnitz, 1982). A clinical review of lumpectomy endorses the standard statement for "most cases of early breast cancer" (Montague, 1982). Leis (1980b) repeats the "current treatment standard" statement without comment.

The CDC's recommendation of a two-step procedure received a mixed response. A letter in the Canadian Journal of Surgery that cites the Consensus Statement disputes in detail the recommendation for a two-step procedure. It argues that the patient is incapable of rational decisionmaking, that delay causes the patient to suffer anxiety, and that two operations increase risk from the anesthetic and inefficient use of resources (Devitt, 1982). However, beginning on the same page, the journal editor provides a similarly detailed case in favor of the procedure. He argues that the biopsy can be done in the office with trivial risk and cost and that the anxiety about whether one will wake up without a breast is worse than the delay (Mueller, 1982). Fentiman, Millis, and Hayward (1980) also approve of the recommendation for a two-step procedure.

Two citing articles stand apart. One review article objects to the composition of the CDC panel and the lack of specificity of the recommendations (Leis, 1980a). Levitt (1980) cites CDC statements concerning the need for further study of radiation therapy to buttress the argument that radiation therapy is not known to be harmful.
Citations to the Minority Reports

Sixteen articles contain a total of 18 citations to the "minority report" articles written by Fisher (1979), Moxley et al. (1980), and Urban (1980).

Kinne (1982) objects to the two-step procedure recommended by the panel. The article, in CA: A Cancer Journal for Clinicians, reviews the evidence relating delay between biopsy and surgery to outcome and finds that selected patients face an increased risk. Kinne cites Moxley as a reference to the CDC recommendations and cites Urban's contrary opinion to endorse it. He states that the CDC recommendation was "primarily an attempt to encourage patients to enter more clinical trials."

Two other articles cite Moxley et al. (1980). Liberati et al. (1982) refer to Moxley and ten other articles (including Fisher, 1979) to demonstrate that there are successful treatment modalities for breast cancer. Gefter, Friedman, and Goodman (1982) cite Moxley et al. to suggest that these CDC panelists are among "a number of authors (who) have proposed that breast cancer be treated by removal of palpable tumor followed by radiation therapy, rather than mastectomy." This appears to us to be a major distortion of the cited authors' position.

All the remaining articles cite Fisher (14 total), most in reference to his systemic theory. Two articles express disagreement with the theory (Leis, 1980b; Urban, 1980b). Eight articles apparently agree with the theory. Five of these, including two written by Fisher and his colleagues, deal with conservative treatments of breast cancer (Deffebach, Goodman, and Miller, 1982; Fisher et al., 1981a, 1981b; Montague, 1982; Weissberg and Prosnitz, 1982). Of the other three, one deals with breast cancer screening (Slater et al., 1980) and two with other cancers (Bizer, 1980; Edis, Kiernan, and Taylor, 1980).

Vilcog et al. (1981) cite Fisher's advocacy of further clinical trials (which was also the CDC position). Gefter, Friedman, and Goodman (1982) mention Fisher's advocacy of conservative treatment. Liberati et al. (1982), who cite both Fisher and Moxley, have been mentioned. The remaining reference (Levitt, 1980) is too vague to convey meaning.
CONCLUSIONS

Our findings suggest the following conclusions:

1. For at least one issue, the conference documented a prevailing, perhaps unspoken, consensus—that the Halsted radical was no longer seen as the treatment standard. No one expressed any explicit objection to the "current standard" statement in subsequent published discussion. It was used by several authors to place their own more conservative treatment regimens in perspective.

2. The consensus conference on treatment of primary breast cancer provided an occasion for public recognition of other controversial issues. At least some controversy surrounded two conference recommendations: (1) the two-step procedure and (2) the endorsement of the NSABP protocol. Although these controversies predated the conference, the consensus statement provided the occasion for the publication of a continuum of opinions.

3. Publication of the CDC findings provided an opportunity for some authors to advance their views. Fisher used the occasion of publication of the Consensus Statement to write a concise editorial presenting his view of tumor biology. Although the article ended with the need for further study of breast conserving operations (including a plea for surgeons to participate in the NSABP), several of the citations to the editorial imply acceptance of both the theory and its implication that conservative surgery is an appropriate treatment.

BIBLIOGRAPHY

Primary Articles


Secondary Articles


V. STEROID RECEPTORS IN BREAST CANCER

Breast cancer is a major cause of mortality in the United States. In the early 1970s it was noted in laboratory experiments that cells from some breast cancers contained high levels of steroid receptors. The clinical importance of this laboratory finding came from the possible relationship between high levels of steroid receptors and response to hormone therapies.

By the late 1970s further research had been done on the link between steroid receptors and response to hormone therapy. The preeminent question at that time was how this laboratory research could be applied in the clinical setting.

A Consensus Development Conference on Steroid Receptors in Breast Cancer was held June 27-29, 1979. It was sponsored by the National Cancer Institute's Division of Cancer Biology and Diagnosis.

The conference addressed the following question:

- What is the value of steroid receptors assays to the clinical management of breast cancer?

The consensus development panel's conclusions and recommendations fall into categories defined by two central issues:

1. Clinical correlates: the relationship between receptor assays and clinical outcomes in breast cancer, and
2. Analytical methods: the choice of appropriate analytical techniques and quality control measures for the measurement of receptor levels.

Regarding the clinical correlates of receptor assays, the panel found that the presence of estrogen receptors in breast cancer is clearly related to the response to endocrine therapies and concluded that all primary breast cancer tumors should be assayed for estrogen receptors. Regarding analytical methods, the panel concluded that
biochemical techniques currently provide the most valid and reliable assays but called for the development of new histochemical techniques and quality control for receptor assays.

Like other CDCs, this conference had separate messages for two different audiences: conclusions concerning clinical practice and recommendations for research. It differed from other CDCs in how it presented its findings to those audiences. As usual, the consensus development panel issued a Consensus Statement, "Steroid Receptors in Breast Cancer," which summarizes its findings. Unlike the other CDCs, however, in this case the Consensus Statement itself was not published in any journal. Instead, a special report on the conference was published in the New England Journal of Medicine. The research studies that were presented at the conference were published in a supplement to a major oncology journal (Cancer, Vol. 46, No. 12, pp. 2759-2964, December 15, 1980). Possibly as a result of this two-part dissemination strategy, the findings of this CDC have had a much greater effect on the medical and scientific literature than have those of any other conference we studied.

Because of the distinctive way in which this CDC disseminated its findings, the organization of this section is also distinctive. The first subsection reviews the special report on the conference published in the NEJM and analyzes the citations to it. The second reviews the 1980 special supplement to Cancer devoted to the conference papers and analyzes the citations to them. The third offers our conclusions and their implications for future CDCs. Because of the very large number (338) of articles citing articles in the supplement, the annotated bibliography is limited to the primary literature and citations to the NEJM special report.

SPECIAL REPORT IN THE NEW ENGLAND JOURNAL OF MEDICINE

A Special Report published in the NEJM (DeSombre et al., 1979) discusses the CDC and presents its recommendations, quoting extensively from the Consensus Statement. It differs from the Statement in that it provides background on both the consensus process and the history of receptor research.
This Special Report was cited in 36 articles from its publication in November of 1979 through the end of 1983. These articles fall into two groups: (1) Articles that refer to the Special Report as a source of information on clinical correlates and (2) articles that discuss analytic techniques.

**CITATIONS CONCERNING CLINICAL CORRELATES**

The Special Report's conclusions regarding the clinical implications of receptor assays were widely cited in the literature. Of the 17 references to the clinical correlates, nine appear in basic science journals, four in cancer journals, and four in general medical journals.

All nine references in the basic science literature refer to the report as providing confirmation of the link between estrogen receptors and response to hormone therapy. Five of these are original research articles, in four of which the reference is found in the introduction to the paper and the citation is mentioned in describing the clinical justification for the basic research. In one research article the reference appears as part of the discussion. The four remaining references in the basic science literature appear in review articles and cite the report's conclusion on the link between receptors and response to hormonal therapy.

All four citations in cancer journals appear in articles that present original research. Only one cites the report as engendering the research; the other three refer to it as providing confirmation of facts.

The four references in general medical journals appear in journals with limited readership or in journals not clinically related to breast cancer. Two references in urology journals use the relationship of hormone receptor to therapy and survival in breast cancer as a model for potential relationships in prostate cancer. An article by Nobbs (1982) is the only article that clearly states the recommendation regarding the need to assay all primary breast tumors. The article by Loven et al. (1981) explicitly uses the report as a framework for interpreting research.
Citations Concerning Analytical Methods

The report's conclusions on analytical methods were also widely cited in the literature. Although most of the references to the clinical conclusions invoke the report as providing confirmation of background facts, references to its conclusions regarding analytical methods are more directly related to the research reported in the papers.

The 20 citations to the Special Report that concern analytical methods can be divided into three groups:

1. Eight papers cite the need for the development of quality control measures in measuring receptor levels.
2. Six papers cite the report as providing justification for the use of a particular biochemical receptor assay method.
3. Six papers cite the report as a source of information on the poor reliability and validity of histochemical methods for determining receptor levels.

Ten of the 20 references appear in basic science journals. Of these, four deal with histochemical methods and refer to the report as indicating the need for the validation of such techniques. Three refer to the report as justifying the choice of an analytical method. The final three cite the need for quality control and for studying the effects of various factors on assay results.

Seven references to the report appear in the cancer literature. Two of these are reviews that note the need for quality control in assay measurement. Two original research papers deal with factors that underlie the variability of assay results. Three studies of new assay methods refer to the report as defining appropriate assay techniques.

As in the case of the citations concerning clinical correlates, the citations to the analytical methods recommendations found in the general medical literature appear in journals with limited readerships. Two of these references refer to the report as defining a valid assay procedure and the last cites the report on the need for quality control.
THE CONFERENCE SUPPLEMENT IN CANCER

The papers that were presented at the CDC were subsequently published in the journal Cancer. Of these 42 papers, 27 deal primarily with the clinical correlates of receptor levels and 15 with analytical methods.

There were 338 citations to the articles in the Cancer supplement from its publication in 1980 through the end of 1983. On average, each article was cited a little over eight times, but the number of citations ranged from one to 30. The number of citations increases steadily over the years after the publication of the supplement.

Citations Concerning Clinical Correlates

The 27 papers dealing with the clinical correlates of receptor levels can be divided into two groups: (1) 12 articles that discuss the relationship of receptor levels to response to hormone therapy and (2) 15 articles related to other clinical correlates.

Table 5.1 lists the 12 articles in the Cancer supplement that deal primarily with the relationship between receptor levels and the response to hormone therapy. The table also indicates the number of citations each of those articles received from 1981 through 1983. The most cited article (Osborne et al., 1980) reviews a large number of patients and draws the following conclusions: (1) estrogen receptor negative patients have a shorter survival, (2) quantitative measures of estrogen receptors improves the prediction of response to hormone therapy, and (3) progesterone receptor levels improve the prediction of response to hormone therapy. The article by DeSombre and Jensen (1980), the second most widely cited, points out that estrogen receptor assays of the primary tumor are useful in predicting the response to hormone therapy and that quantitative estrogen receptor levels improve prediction of response. Table 5.2 lists citation totals by year for the 15 articles that deal primarily with the other three clinical correlates discussed at the CDC:
Table 5.1
CITATIONS TO CANCER SUPPLEMENT ARTICLES ON THE
RELATIONSHIP BETWEEN RECEPTOR LEVELS AND
RESPONSE TO HORMONE THERAPY, 1981 THROUGH 1983

<table>
<thead>
<tr>
<th>Primary Article</th>
<th>Number of Citations by Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brooks et al.</td>
<td>1 3 1 5</td>
</tr>
<tr>
<td>Dao and Nemoto</td>
<td>0 2 5 7</td>
</tr>
<tr>
<td>DeSombre and Jensen</td>
<td>3 6 7 16</td>
</tr>
<tr>
<td>Degenshein, Bloom, and Tobin</td>
<td>0 4 0 4</td>
</tr>
<tr>
<td>Lippman and Allegra</td>
<td>3 2 5 10</td>
</tr>
<tr>
<td>Maass et al.</td>
<td>2 1 3 6</td>
</tr>
<tr>
<td>Manni, Arafah, and Pearson</td>
<td>0 1 2 3</td>
</tr>
<tr>
<td>McCarty et al. (1980b)</td>
<td>0 0 7 7</td>
</tr>
<tr>
<td>Nomura et al.</td>
<td>0 2 4 6</td>
</tr>
<tr>
<td>Osborne et al.</td>
<td>5 10 14 29</td>
</tr>
<tr>
<td>Paridaens et al.</td>
<td>2 3 7 12</td>
</tr>
<tr>
<td>Skinner, Barnes, and Ribeiro</td>
<td>1 1 3 5</td>
</tr>
<tr>
<td>Total</td>
<td>17 35 58 110</td>
</tr>
</tbody>
</table>

1. The relationship between histopathological features and receptor levels,
2. The relationship between receptor status and response to chemotherapy, and
3. The relationship between estrogen receptor levels and survival.

The paper by McCarty et al. (1980c) confirms previous findings that there is no consistent relationship between the histological type of the tumor and estrogen receptor levels. However, both McCarty et al. (1980c) and King (1980a) suggest that the receptor levels are associated with the histological grade of the tumor.

The paper most commonly cited on the relationship between receptors and response to chemotherapy (Hilf et al., 1980) carefully reviews a series of patients and concludes that the prognostic value of estrogen receptor data in nonhormonal therapy settings remains unproven. Young, Ehrlich, and Einhorn (1980), also cited widely, review 54 patients with
Table 5.2
CITATIONS TO CANCER SUPPLEMENT ARTICLES ON OTHER CLINICAL CORRELATES, 1981 THROUGH 1983

<table>
<thead>
<tr>
<th>Primary Article by Topic</th>
<th>Number of Citations by Year</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1981</td>
</tr>
<tr>
<td>Histopathology</td>
<td></td>
</tr>
<tr>
<td>King (1980b)</td>
<td>2</td>
</tr>
<tr>
<td>McCarty et al. (1980c)</td>
<td>4</td>
</tr>
<tr>
<td>Millis</td>
<td>2</td>
</tr>
<tr>
<td>Chemotherapy Response</td>
<td></td>
</tr>
<tr>
<td>Hilf et al.</td>
<td>2</td>
</tr>
<tr>
<td>Jonat et al.</td>
<td>1</td>
</tr>
<tr>
<td>Kiang et al.</td>
<td>1</td>
</tr>
<tr>
<td>Rosenbaum et al.</td>
<td>1</td>
</tr>
<tr>
<td>Rubens and Hayward</td>
<td>1</td>
</tr>
<tr>
<td>Samal et al.</td>
<td>1</td>
</tr>
<tr>
<td>Sears and Olson</td>
<td>0</td>
</tr>
<tr>
<td>Young, Ehrlich, and Einhorn</td>
<td>0</td>
</tr>
<tr>
<td>Survival</td>
<td></td>
</tr>
<tr>
<td>Blamey et al.</td>
<td>5</td>
</tr>
<tr>
<td>Furmanski et al.</td>
<td>3</td>
</tr>
<tr>
<td>Hubay et al.</td>
<td>2</td>
</tr>
<tr>
<td>Singhakowinta et al.</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
</tr>
</tbody>
</table>

metastatic disease and argue that more elaborate and better controlled studies are needed before a definitive conclusion can be drawn on the relationship between estrogen receptors and response to chemotherapy.

Among the group of papers on survival, those by Blamey et al. (1980) and Furmanski et al. (1980) conclude that positive receptor status is related to both longer survival and longer disease-free intervals.

Table 5.3 lists the 15 articles on analytical methods and the number of references to these articles from the beginning of 1981 through the end of 1983. Again, the number of references generally increases from year to year. The average number of references per paper is a little over six, but the number ranges widely, from one to 16. The most commonly cited paper (McCarty et al., 1980a) examines the
Table 5.3
CITATIONS TO CANCER SUPPLEMENT ARTICLES ON ANALYTICAL
METHODS 1981 THROUGH 1983

<table>
<thead>
<tr>
<th>Primary Article by Topic</th>
<th>1981</th>
<th>1982</th>
<th>1983</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytical Techniques</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benraad and Koenders</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Bojar et al.</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Hoffman et al.</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Lee</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>McCarty et al. (1980a)</td>
<td>2</td>
<td>5</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>Mercer et al.</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Morrow et al.</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Pertschuk et al.</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Rao et al.</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Rodbard, Munson, and Thakur</td>
<td>0</td>
<td>1</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Segaloff</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Quality Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>King (1980a)</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Smith</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Wagner and Jungblut</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Wittliff</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>7</td>
<td>37</td>
<td>54</td>
<td>98</td>
</tr>
</tbody>
</table>

relationship between biochemical and histochemical assay techniques and concludes that none of the histochemical techniques is suitable for clinical evaluation of patients. The article by Mercer et al. (1980) also deals with histochemical techniques and concludes that they are not specific for estrogen receptors. The article by Rodbard, Munson, and Thakur (1980), cited 12 times, outlines the correct method for statistical analysis of receptor concentration.

CONCLUSIONS

Our findings lead us to the following conclusions:
1. The CDC on Steroid Receptors provided a venue in which researchers could review the available evidence and come to agreement that a particular hypotheses had been confirmed. As a result, most articles citing the CDC's clinical conclusions do so in the context of confirming facts.

2. The CDC panel's central recommendation, regarding the need to assay all primary tumors, has not been widely cited in the scientific literature. There were, however, a large number of references to recommendations and findings related to clinical correlates and analytic methods. The difference in effect of the recommendations on the literature may reflect the fact that the professional literature is primarily directed at research rather than clinical policy.

3. The distinctive dissemination strategies of this CDC appear to have been quite successful. Even though it lacks a bibliography or references, the Special Report in the New England Journal of Medicine offers more background information than the usual brief Consensus Statement and--whether coincidentally or not--has been heavily cited. Moreover, publication of the conference presentations in a premier specialty journal seems to have provided a valuable resource to those involved in research. These papers too have been frequently cited.

BIBLIOGRAPHY

Primary Literature

A. Statement


B. Special Report

C. Cancer Supplement


Secondary Literature (Articles Citing Special Report)


Refers to preference for biochemical assay.


Refers to preference for biochemical assay.


Refers to preference for biochemical assay.


Refers to need for quality control.


Refers to the need for histochemical assays.


A review of ovarian steroids; refers to relationship between receptors and hormone response.


Review of histochemical assay methods.

Review article noting need for quality control.


Review article noting need for quality control.


Research study citing the relationship between receptors and response to hormone therapy.


Refers to relationship between receptors and response to hormone therapy.


Refers to relationship between receptors and response to hormone therapy.


Refers to relationship between receptors and response to hormone therapy.

Refers to use of biochemical assay.


Refers to need for quality control.


A review article by a group of well-known researchers; refers to the relationship between receptors and response to therapy.


A review article that refers to the relationship between receptors and response to therapy.


Refers to relationship between receptors and response to therapy.


Original research article; specifically mentions the recommendation to assay all primary tumors.


Refers to relationship between receptors and response to therapy.

Refers to relationship between receptors and response to chemotherapy.


Refers to need for quality control.


Refers to need for quality control.


Review of adjuvant endocrine therapy that mentions the need for quality control of receptor assays in future clinical trials; refers to need for quality control.


Refers to development of histochemical assay.


Refers to use of biochemical assay.


Refers to development of histochemical assay.


Refers to need for quality control.

Review of receptors in prostatic cancer that refers to relationship between receptors and response to hormone therapy.


Refers to development of histochemical assay.


Refers to relationship between receptors and survival.


Refers to development of histochemical assay.


Refers to relationship between receptors and response to hormone therapy.


Cites the special report twice, once in relation to clinical correlates and once with respect to analytical methods.


Refers to relationship between receptors and response to hormone therapy.
VI. ESTROGEN USE AND POSTMENOPAUSAL WOMEN

Beginning in the early 1960s, estrogen was vigorously promoted to alleviate the undesirable symptoms and signs that accompany and follow menopause. By the early 1970s, between 18 and 21 million prescriptions were written annually.¹

In 1975, the first of several retrospective case control studies linked estrogen use to the development of endometrial cancer. By 1977, the number of prescriptions had dropped to 7 million. However, the late 1970s also saw the emergence of evidence that estrogen could be of significant long-term value in preventing osteoporosis—a condition characterized by loss of bone mass that frequently contributes to spine, hip, and other fractures. Several articles also suggest the possibility that estrogen administration may be linked to breast cancer or might prevent heart attacks.

The Consensus Development Conference on Estrogen Use and Postmenopausal Women was held September 13-14, 1979. The conference addressed the following subjects:

- Benefits of estrogen use, including treatment for menopausal symptoms and possible prevention of osteoporosis;
- Hazards of estrogen use;
- Relative risks and benefits of various types of estrogen therapy;
- Economic issues;
- Indications for and contraindications to estrogen use.

The conference concluded that estrogen replacement therapy (ERT) is effective for many menopausal symptoms, but not for treatment of psychological problems. Although the CDC panel found that ERT can retard bone loss, it judged that further data were needed concerning whether ERT prevents fractures. The CDC summary also describes an increase in the incidence of endometrial cancer due to estrogen therapy.

¹See Gastel and Brody, 1978.
The conference summary was reprinted in virtually complete form in one specialty journal and two regional journals, and a second specialty journal reported on the conference findings. The conference also generated four articles, three based on speaker presentations. One was published in the NEJM and the others in specialty journals.

Although few references have appeared to the bulk of this primary literature, the three published presentations have been cited frequently. Many of the citing articles also cite scientific developments immediately following the conference that cast doubt on the continued validity of some of the CDC findings. In particular, additional evidence accumulated that long term estrogen therapy has important benefits in reducing the incidence of many kinds of fractures and that its major proven risk--endometrial cancer--can be reduced through administration of a combination therapy.\(^2\) These scientific developments were widely and rapidly publicized by authors of clinical reviews, who urged greater use of combination estrogen therapy as a prophylaxis for osteoporosis. It may be that the CDC contributed to this dissemination by showing that the clinical decision hinges on the truth of these two scientific propositions, believed by the CDC members to be uncertain at the time of the CDC.

**PRIMARY LITERATURE**

The CDC on Estrogen Use and Postmenopausal Women was reported in several professional journals. Two journals reprinted the entire summary statement with only slight editing (*The Annals of Internal Medicine*, 1979, and *Connecticut Medicine*, 1980). A third article condensed the summary by omitting some of the statements that supported recommendations and a few of the statements about needed research (Bouis, 1980).

A short but informative news article provided information about the professions of panel members, speakers, and the audience. It briefly touched on every one of the conclusions reached by the panel and also

\(^2\)There were also developments concerning the risk of cardiovascular disease, but these are not addressed here because the discussion never directly cited CDC primary literature.
reported on "panel guidelines given in a question and answer session" but not repeated in the summary statement. It reported that the panel reached its conclusions at 5 a.m. after "intense debate" and that audience reaction to the report was "mixed," with some audience physicians believing that estrogen use is indicated in more situations than the panel recognized (Eichelberger, 1979).

The four remaining primary articles provided much more information about the scientific background for the CDC. Three of these were written versions of the formal presentations made to the panel (Hulka, 1980; Schiff and Ryan, 1980; and Weinstein, 1980). Although there is some overlap in the content of these three articles, Hulka concentrated on the risks of estrogen use, Schiff and Ryan concentrated on the benefits of estrogen use, and Weinstein balanced risks and benefits in a cost effectiveness analysis. Drawing heavily on these three presentations, the remaining article directly related each of the conclusions reached by the CDC panel to the scientific data base. It identified each of the pieces of research referred to in the summary statement. This article repeated one or more paragraphs from the CDC summary and then added commentary, covering the entire substantive section of the CDC in this fashion. The authors were from the conference sponsor, National Institute of Aging (Gastel, Corno, Huntley and Brody, 1980).

SECONDARY LITERATURE

The speaker presentations were cited most---Weinstein's article was cited 25 times, Hulka's article nine times, and Schiff and Ryan's six times. The other articles were cited much less frequently: Gastel, Corno, Huntley, and Brody were cited only once and the CDC summary only five times. Three citing articles are in German. All other articles were retrieved and read.

Most of the citations concern one of three major research areas:

1. prevention of osteoporosis,
2. endometrial cancer, and
3. the economic evaluation of health care procedures.

Below we summarize the content of the CDC and the speaker articles in each of these areas and indicate how these articles have been cited in the literature. We then characterize the citing articles that do not fall into any of the three major categories.

Prevention of Osteoporosis

The CDC panel decided against recommending estrogens for the sole purpose of preventing osteoporosis. Weinstein's cost effectiveness analysis showed that benefits in addition to osteoporosis prevention are necessary to justify estrogen replacement therapy at the assumed disease incidence rates. While acknowledging that estrogens retard bone loss, the Consensus panel stated that "more data are definitely needed before the efficacy of estrogens in preventing fractures can be established." The CDC further recommended research on the identification of patients at increased risk for osteoporosis because it is likely that this group will present valid indications for estrogen therapy.

In addition to the Hutchinson, Polansky, and Finstein (1979) study reviewed by the CDC while in press, at least four studies published following the conference have presented statistics suggesting that estrogens do prevent fractures. Only one of these (Jensen, Christiansen, and Transbol, 1982b) cites any of the conference papers. One paper (Lindsay et al., 1980) is a prospective controlled trial of estrogens in oophorectomized women. The others are retrospective studies (Weiss et al., 1980, and Paganini-Hill et al., 1981).

Eleven citing articles (accounting for nine references to Weinstein, three to Hulka, three to the reprinted Statement, and four to Schiff) deal extensively with osteoporosis prevention. Nine of these are clinical reviews rather than reports of original research.

The emerging evidence in favor of wider use of ERT led the majority of citing authors to differ from the CDC position on the issue. The Council on Scientific Affairs of the American Medical Association began its January 1983 report on estrogen replacement in the menopause by noting the controversy surrounding estrogen use; by citing the CDC
summary, its three speakers' presentations, and various other statements on the issue; and by declaring that its purpose was to place these cited statements "in perspective for the practicing physician." Although many of its findings are similar to the CDC findings, it did find that "well controlled studies indicate that estrogen therapy reduces the incidence of both vertebral and long bone fractures."

Other citing authors appear to agree. Brenner (1982) states that "the benefits of long term cyclic estrogen-gestagen therapy in preventing osteoporosis far outweigh the risks." Coope (1983) straightforwardly lists "prevention of osteoporosis" as an indication for hormone therapy. Mallette (1982) reviews the basic findings of the Weinstein article, then notes that if the cost effectiveness ratio is neutral for the average patient, it must be positive for those with a high risk of osteoporosis--such as thin people, smokers, and those with low calcium diets. Similarly, Raisz (1982), citing Weinstein, notes that the costs and complications of estrogen replacement therapy are relatively high, but concludes that ERT should nevertheless be given to patients with a high risk of osteoporosis. Jensen, Christiansen, and Transbol (1982b) state that events since the CDC and Weinstein's analysis make it a "sound principle to use (estrogens for) prophylaxis in high risk patients."

Three of the remaining citing articles were closer to the CDC position. Korenman (1982) calls a clinical state with a propensity for osteoporosis a "controversial indication" for hormone therapy. In a letter, Mosher and Whelan (1981) cite Weinstein and Schiff to rebut another letter by Gordon and Vaughn (1981), who argue in favor of greater use of estrogen therapy. Ryan (1982) states that more information is necessary before recommending routine use of estrogens to prevent osteoporosis.

Two citing papers consider alternatives to estrogen replacement. Sandler and La Porte (1983) hypothesize that an increased amount of physical activity may lower the need for calcium intake and be a more cost-beneficial prophylaxis for osteoporosis than is ERT. As part of their argument that a better prophylaxis is needed, they note that Weinstein found that the benefits of estrogen do not clearly outweigh the risks of the regimen. Cooke (1983) wrote a letter to the British
in which he reviews the problems involved in long term hormone therapy. He argues in favor of a clinical trial that would compare vitamin D and calcium dietary supplements to estrogen replacement therapy. He cites Schiff and Ryan (1980) for fracture and mortality statistics.

Risk of Endometrial Cancer

The CDC summary states that the incidence rate for endometrial cancer is increased by the use of conjugated estrogens. Hulka's CDC presentation (as reported in its written version) provided a detailed review of the evidence on this issue from her own work as well as from eight other recent case control studies. Weinstein counted the increased costs of tests and treatment for this disease as a significant part of the treatment costs of estrogen replacement therapy.

One of the CDC panel's research recommendations concerns this area of risk. The CDC found:

that the use of progestins for several days of each estrogen treatment cycle has been demonstrated to decrease the occurrence of endometrial hyperplasia and might also reduce the associated risk of developing cancer of the endometrium. Before the application of combined therapy becomes established, risks of the various progestins must be adequately evaluated.

Collins (1981) wrote an editorial arguing that a greater understanding of the pharmacology of estrogens will result in the elimination of the increased risk of endometrial cancer. He refers his readers to Weinstein to find references for the current existence of an increased risk.

However, four other citing articles juxtapose a reference to either the Hulka or Weinstein articles with a reference to other data suggesting that the use of progesterone in combination with estrogen will eliminate the risk now (Johansson, Mesinis, and Nillius, 1981; Center for Disease Control, 1983; Jensen, Christiansen, and Transbol, 1982a; Christensen et al., 1982). The citing articles base the value of progesterone on either: (1) the biological role of progesterone as an inhibitor of the growth-stimulating effect of estrogen or (2) case
control studies (almost all published after the CDC) showing a protective effect, including one published by Hulka and colleagues (1982). Several of the clinical reviews discussed in the osteoporosis section also urge the use of progestin with estrogen, but without direct reference to any of the CDC primary articles.

The remaining secondary articles concerning endometrial cancer cite the CDC primary literature in varying contexts:

1. Kauppila, Gronroos, and Nieminen (1982) report the clinical outcome of endometrial cancer in 1,113 patients. They cite the Gastel, Cornoni-Huntley, and Brody statement that estrogen use is correlated with a decline in the mortality of endometrial cancer patients but find no correlation between progesterone use and survival in their population.

2. Chu, Schweid, and Weiss (1982) compare the endometrial cancer survival rates of estrogen users and nonusers. They cite Weinstein and other articles as showing that the increased risk of cancer does not preclude use of estrogen therapy, because there are important benefits. Chu, Schweid, and Weiss find longer survival in the estrogen users. The CDC also found that estrogen use is most strongly associated with the mildest cases of this cancer, but Chu and colleagues did not mention that. Although several of the articles discussed elsewhere contain statements agreeing with this position, only Kauppila, Gronroos, and Nieminen (1982) cite any of the CDC primary articles to make this point.

3. Lauver (1983) gives advice to nurses who care for women with irregular bleeding, including those who are premenopausal. She refers her readers to Hulka (1980) for further information about the incidence of endometrial cancer.

4. Persson et al. (1983) describe how estrogen therapy for the menopause is practiced in Sweden. They cite the Hulka article for its review of articles demonstrating an increased risk of cancer.
Economic Evaluation Literature

Six articles cite Weinstein in connection with the economic evaluation of health care. The thrust of the references varies. Although only one critic objected to the validity of cost benefit or cost effectiveness analyses, several discussed ways in which the methodology of the field could be improved. The citing articles were as follows:

1. Orient (1981) argues that cost benefit and cost effectiveness ratios are not meaningful concepts—in part because costs are not measurable in a public care system.

2. Boyle et al. (1983) provide a cost effectiveness analysis of neonatal intensive care of very low birth weight infants. They cite the Weinstein article as a predecessor that used "cost per quality-adjusted life year" as one of its economic evaluation measures, and they suggest the need for even greater standardization of the methodology for economic evaluation of health care.

3. Mitchell (1981) reviews the state of obstetrics and gynecology in this address. A major part of the discussion concerns the ways that decision analysis is valuable in both clinical and academic medicine.

4. Bush et al. (1982) discuss several technical controversies concerning a health-related quality of life scale that is an important input to economic policy analyses. One of their points is that policy conclusions should not be sensitive to small errors in the measurement of preferences. The statement: "Indeed, useful and influential analyses have been produced using arbitrary preference assignments" concludes with a reference to Weinstein.

5. Weinstein (1980) is a primer on cost effectiveness and cost benefit analysis presented during a French symposium. Data from the estrogen study are cited and used in a teaching example.
6. In a two-part *JAMA* article, Evans (1983) discusses the implications of health care technology. The Weinstein article is one of many articles that are cited as a group to indicate that health care resources are limited and that cost effectiveness is a well-known term.

Two additional articles appear to have been published with the primary intent of providing better data to future cost benefit analyses. Investigators from the Mayo Clinic published an article on the incidence rates for Colles' fracture (Owen et al., 1982) and on the epidemiology of humeral fractures (Rose et al., 1982). In each case they stress the value of their data for cost benefit analyses of osteoporosis prevention, citing Weinstein. In the case of Colles' fracture, they suggest that their data are a significant improvement over the data available to Weinstein. The paper on humeral fractures points out that Weinstein did not include these fractures and argues that future calculations should include them. In both cases, a recalculation with other things equal would show that prevention has higher benefits than Weinstein calculated.

**Miscellaneous**

The remaining secondary articles cite the CDC primary literature in a variety of contexts:

1. Moore, Moore, and Moore (1983) provide an extensive review of research on all factors hypothesized or demonstrated to be etiologically related to breast cancer. They note that Hulka reported no association between breast carcinoma and the use of estrogens in the treatment of postmenopausal women.
2. An editorial by Kauppila (1982) considers the costs and benefits of performing oophorectomy as a routine part of hysterectomy. He repeatedly cites Schiff and Ryan as the source of clinical information on the outcome of removal of the ovaries.

4. An annotated bibliography of literature on endocrinology and metabolism (Freinkel et al., 1982) includes Weinstein's article in its section devoted to the reproductive system in women.

5. Elkik et al. (1982) compare the level of various plasma constituents in postmenopausal women receiving estradiol (a particular form of estrogen) percutaneously with levels in similar women receiving conjugated estrogens orally. They cite both Weinstein and the Annals of Internal Medicine version of the CDC summary in a very general manner in the introduction to their paper. Although it is not mentioned in the paper, their findings appear to address one of the areas that the CDC singled out as needing research when they noted that the biological consequences of the absorption of estrogen-containing creams into the bloodstream "are undetermined and require study."

6. Taggart et al. (1982) cite Weinstein's article in the course of reporting on changes in lipids in ten postmenopausal women who received an anabolic steroid.

7. A letter from Huth, the editor of the Annals of Internal Medicine, to the editor of the New England Journal informs journal readers that the Annals publishes many of the CDC summaries.

8. Ryan (1982) repeatedly cites his and Schiff's CDC paper as a source of information on vasomotor flushes.³

³This paper was previously mentioned in the subsection on osteoporosis.
CONCLUSIONS

The CDC discussion appears to have been viewed as very relevant to clinical practice. Presentations given at the conference or the CDC summary statement itself were cited in 11 review papers aimed at medical practitioners. However, most of these reviews also cite scientific developments in the years immediately following the conference that cast doubt on the continued validity of a few of the CDC findings. In particular, additional evidence accumulated that long-term estrogen therapy has important benefits in the reduction of the incidence of many kinds of fractures and that its major proven risk--endometrial cancer--can be reduced through administration of a combination therapy. These developments, which address needs identified in the CDC research recommendations, led several of the writers of these clinical reviews to recommend long-term ERT to prevent osteoporosis, which is at variance with the CDC position. It is possible that the CDC crystallized the argument so that the importance of these studies was perceived more quickly or more widely than it would otherwise have been.

The conference seems to have had a considerable effect on the research community. The conference's research recommendations appear to have prompted, guided, or drawn attention to subsequent research. The three published speakers' presentations seem to have been especially helpful to the research community, perhaps because they discuss the scientific evidence on which the CDC based its findings and recommendations. Weinstein's research sparked an active discussion of the value of cost effectiveness analysis and may have inspired the generation of data needed to perform better analyses in the future. The Hulka, and Schiff and Ryan reviews were cited in later research in much the same ways as any state of the art review.

BIBLIOGRAPHY

Primary Articles


**Secondary Articles**


**Related Articles**


VII. THROMBOLYTIC THERAPY

Thrombolytic therapy utilizes chemical agents (typically urokinase or streptokinase) to lyse, or dissolve, an acute vascular blockage such as a deep-vein thrombus or pulmonary embolus. Such conditions have been typically treated by anticoagulants such as heparin. Anticoagulation has little effect on existing thrombi or emboli, however; it primarily prevents their further formation. Thus, thrombolytic therapy represents a more ideal form of management, but because it dissolves fibrin, there is additional risk of bleeding associated with its use. This risk is particularly acute where bleeding has occurred recently (e.g., sites of invasive procedures), and such recent bleeding constitutes its most important contraindication.

In the late 1960s and early 1970s, the NHLBI sponsored a large clinical trial on the use of urokinase and streptokinase in patients with pulmonary embolism. Streptokinase and urokinase were subsequently licensed by the FDA; but as a condition of licensure, manufacturers were required to monitor experience with use and side effects. Several years after licensure, NHLBI and FDA suggested that a consensus development conference would be timely, to review clinical experience with the agents and to educate clinicians in basic concepts in fibrinolysis as they relate to thrombolytic therapy. The NIH Consensus Conference on Thrombolytic Therapy was held April 10-12, 1980.

The conference discussion addressed the following questions:

1. Are there proven benefits for the use of thrombolytic therapy in the more severe forms of acute deep vein thrombosis and pulmonary embolism?
2. What are the risks associated with the use of thrombolytic agents?
3. What are the guidelines for patient selection?
4. Why are these drugs not used more frequently?
5. How should patients using these drugs be managed and monitored?
6. What are the economic aspects of thrombolytic therapy?
7. What future clinical research is needed?

The consensus development panel issued a statement, "Thrombolytic Therapy in Thrombosis," which summarizes its findings concerning these seven questions. Very briefly, the conference concluded that thrombolytic therapy has succeeded anticoagulation as the ideal therapy for the management of a large number of cases of acute deep vein thrombosis and pulmonary embolism. Unlike anticoagulation, thrombolytic agents can lyse thrombi and emboli, restoring circulation to normal.

The NIH Consensus Statement on thrombolytic therapy received fairly wide distribution. It was reprinted by two highly regarded and widely read journals in the United States (Annals of Internal Medicine) and Great Britain (British Medical Journal), as well as by four regional and local medical association journals in the United States. However, this Statement did not occasion any "news" articles in the professional literature, and individuals associated with the conference did not subsequently communicate conference findings or materials through published articles. This may have minimized the influence of the conference somewhat or limited its recognition, relative to some of the other topics.

Conference findings were cited to a moderate degree, but the citations indicate that the recommendations made by the Consensus Panel were by no means universally accepted. The most prestigious publication outlets received the vast majority of the citations.

PRIMARY LITERATURE

Our search of Index Medicus discovered six primary articles discussing the results of this conference. All of the articles appeared in the professional literature within seven months following the conference. Five of these were reprints of the complete Consensus Development Conference Statement, with minor or no editorial changes. One condensed the statement into a short summary. We found no articles written by conference participants or NIH officials that described or elaborated conference findings.
The summary of the conference findings, which appeared in the *Southern Medical Journal*, tailors the Consensus Statement for an audience of clinicians. It briefly covers the major conclusions of the Consensus Statement regarding benefits, risks and contraindications, and "unreasonable" rationales contravening widespread use of thrombolytic therapy. It refers to the Consensus Statement regarding guidelines for patient management and monitoring. It does not mention the Statement's discussions of economic aspects involved in thrombolytic therapy or of suggested research directions.

SECONDARY LITERATURE

Our search of the Institute for Scientific Information (ISI) biomedicine citation index found only 11 citations to the six primary articles described above. Nine of the eleven citing articles cited the *Annals of Internal Medicine* reprint of the Consensus Statement. The *British Medical Journal* reprint received one citation, and the summary of the Consensus Statement printed in the *Southern Medical Journal* received one citation. None of the reprints in the state medical journals received any citations.

Six of the 11 citing articles were reviews, three were research articles (including one case report), one was a commentary, and the final article was a letter. These are discussed below in turn.

Review Articles

Review articles that cited the primary articles were published largely in specialty and general interest publications. In the first of these, Murray, Garnic, and Bettmann (1982) cite the *British Medical Journal* reprint of the Consensus Statement in a review article on the pharmacology of angioplasty and intravascular thrombosis. They cite the Statement regarding the use of thrombin times and levels of fibrinogen breakdown products as parameters of patient management during thrombolytic therapy. Although they express some uncertainty over whether these parameters correlate significantly with hemorrhagic complications, they term the patient management recommendations made by the panel as "prudent."
Another review article found some basis for disagreement. Rosenow, Osmundson, and Brown (1981) reviewed the general subject of pulmonary embolism, including methods of treatment, for the Mayo Clinic Proceedings. They state that heparin therapy is the preferred initial treatment for nearly all cases of deep venous thrombosis and pulmonary embolism. They argue (p. 174) that the major indication for using fibrinolytic therapy is to improve a "deteriorating hemodynamic situation as rapidly as possible and principally within the first 24 hours of massive pulmonary embolism." They acknowledge the Annals reprint of Consensus Statement, however, in stating that some experts believe that fibrinolytic therapy is underutilized. They also cite the Annals article while explaining the pharmacological basis of fibrinolytic therapy.

Hull and Hirsh (1981) chronicled "advances" and "controversies" in the diagnosis, prevention, and treatment of venous thromboembolism for a review article in Progress in Hematology. Their discussion of the pros and cons of thrombolytic therapy is fairly balanced. Although they state that heparin is the standard initial treatment of an acute thrombotic condition, they acknowledge various problems with heparin therapy that led some (such as the NIH Consensus Panel) to advocate initial use of thrombolytic therapy for selected patients with pulmonary embolism or acute venous thrombosis. Citing the Annals reprint of the Consensus Statement, the authors review many of the advantages of thrombolytic therapy advanced by the Panel and report the Panel's guidelines for patient selection and management.

Two other reviews include more disagreement with some of the points made in the Consensus Statement. Stambaugh and Alexander (1981), as part of the American Journal of Hospital Pharmacy's continuing education series, review studies comparing the effectiveness of thrombolytic agents against standard anticoagulants for both pulmonary and deep vein thrombosis. They conclude that various disadvantages militate against the routine use of thrombolytic therapy for treatment of these conditions, and they cite the NIH Consensus Statement (reprinted in the Annals) as calling for more research to improve the efficacy and reduce the risk of bleeding associated with the use of thrombolytic therapy with pulmonary emboli.
In a task force report prepared for the *American Journal of Cardiology*, Gunnar et al. (1982) recommend pharmacologic therapies for several cardiac-related disorders. In discussing patient management during treatment for pulmonary embolism, they cite the Consensus Statement's guidelines for when thrombolytic therapy should be terminated and heparin therapy should be initiated. They state in boldface, however, that the treatment of choice for pulmonary embolism is intravenous administration of heparin, and they term "controversial" the use of thrombolytic therapy for patients with massive pulmonary embolism and deteriorating hemodynamic status.

The final review article is Lo's (1982) clinical guide in the *Western Journal of Medicine* for helping primary physicians diagnose cases of pulmonary embolus. It cites the *Annals* reprint of the NIH Consensus Statement as a general background reference.

**Research Articles**

The Consensus Statement appeared to occasion publication of some basic as well as clinical research. Smith et al. (1982) present an experimental study in *Thrombosis and Haemostasis* of the kinetics of fibrinolysis comparing free and acylated enzymes. They interpret their results as failing to support suggestions made in the research recommendations portion of the Consensus Statement (reprinted in the *Annals*) regarding the possible therapeutic value of certain new thrombolytic agents.

In another article, Totty et al. (1982a) cite the *Annals* reprint of the Consensus Statement as demonstrating that systemic fibrinolytic therapy is "an established method" for treating selected cases of massive pulmonary emboli and acute deep vein thrombosis. This statement provides the background justification for their report, published in *Radiology*, of the results of therapy in 22 patients. Totty et al. (1982b) also published a case report in the *American Journal of Roentgenology* documenting an incident of serum sickness following the local low-dose administration of streptokinase onto thrombi. As background for this report, they refer to the *Annals* reprint of the NIH Consensus Statement to demonstrate that the clinical efficacy of
thrombolytic therapy using streptokinase and urokinase is well-established.

Remaining Articles

The final citations to the primary articles on thrombolytic therapy consist of one commentary and one letter.

Varkey (1982) cites the Annals reprint of the Consensus Statement in an introduction to a three-article symposium on pulmonary disorders in Pulmonary Medicine. While concluding that the recommendations made by the NIH Consensus Panel were "reasonable and acceptable," he alludes to controversy in the literature regarding the indications for thrombolytic therapy between those who advocate thrombolytic therapy for all patients with proximal deep vein thrombosis or acute pulmonary embolism, and those who would restrict its use to a minority of patients.

Lawrence (1982) published a letter in the Western Journal of Medicine encouraging the use of thrombolytic therapy for deep venous thrombosis or pulmonary embolism associated with spinal cord injury. Citing the summary of the consensus conference that appeared in the Southern Medical Journal, she briefly summarizes the Panel's guidelines for the use of thrombolytic therapy pertaining to patient selection, contraindications, and management techniques.

CONCLUSIONS

Our findings suggest the following conclusions:

1. The NIH Consensus Statement on thrombolytic therapy received fairly wide coverage. Six journals in the space of seven months chose to reprint the Consensus Statement in complete or summary form. Authors of the secondary literature, however, overwhelmingly depended upon the reprint of the Consensus Statement that appeared in Annals of Internal Medicine.

2. The Consensus Panel's hope that thrombolytic therapy would become the initial treatment for all forms of acute deep vein thrombosis and pulmonary embolism appears to have encountered considerable resistance. Controversy still exists over the
acceptability of thrombolytic therapy relative to heparin therapy. Four of six review articles on thrombosis published subsequent to the conference still regard heparin therapy as the treatment of choice for most cases of acute deep vein thrombosis and pulmonary embolism, and the application of thrombolytic therapy is encouraged for a handful of cases at best. Only one review article clearly favors more widespread use of the technique. Rationales that the Panel regarded as "unreasonable" appear to persist. These include the belief that response to heparin should be first evaluated and the fear of bleeding.

3. Citations clearly indicate, however, that if thrombolytic therapy is to be undertaken, the Consensus Statement is a definitive source of information regarding patient selection and management. The Consensus Statement was frequently cited for its delineation of contraindications and for its detailed presentation of how to administer the therapy. It also received some credit for its discussion of the pharmacology of fibrinolysis.

4. Although the Consensus Statement may not have succeeded in supplanting anticoagulation with thrombolytic therapy in physicians' preferences, it seemed to engender some necessary research and clinical interest. At least one basic research report and two clinical studies were occasioned by the Consensus Statement. Its more widespread clinical use was also advocated in published columns directed at clinicians. Thus, the Consensus Statement may have placed a "stamp of approval" on the use of thrombolytic therapy for some, perhaps encouraging wider if not widespread use.

BIBLIOGRAPHY

Primary Literature


**Secondary Literature**


VIII. CERVICAL CANCER SCREENING: THE PAP SMEAR

Dr. George N. Papanicolaou developed the test for the presence of cervical cancer, called Pap smear in his honor. The test is effective in detecting cervical cancer in many cases and has no adverse effects. The test was recommended and given annually to the majority of women in this country for many years.

Early in 1980 the American Cancer Society (ACS) revised its guidelines for early cancer detection, recommending Pap smears "at least every three years" rather than annually. The rationale for the change was that invasive cancer is preceded by a period of five to 30 years during which it is in situ and can be detected by the Pap smear and easily treated. The American College of Obstetrics and Gynecology formed a task force that criticized the new ACS guidelines. They argued that the duration of the carcinoma in situ is uncertain and that the Pap smear has a large false negative rate that can best be countered by providing repeated tests.

A Consensus Development Conference on "Cervical Cancer Screening: The Pap Smear" was held at NIH on July 23-25, 1980. It was sponsored by the National Cancer Institute, the National Institute on Aging, and the National Institute of Child Health and Human Development in conjunction with the National Center for Health Care Technology and with the assistance of the Office of Medical Applications of Research, NIH.

The conference met to address the following questions:

1. Does screening with a Pap smear affect the mortality from cervical cancer?
2. Is the Pap smear safe as a screening procedure?
3. Should the Pap smear be used as a routine screening procedure?
4. What critical factors are needed to ensure that the procedure is reliable?
5. What are the responsibilities of those doing the screening for followup, confirmation of findings, and initiation of screening?
The panel issued a Consensus Statement, "Cervical Cancer Screening: The Pap Smear," which summarizes its findings concerning the questions listed above and its recommendations for future research.

Despite the wide agenda, the discussion in the literature focused on a single aspect of the third question--namely, the optimal frequency for Pap smears. This had been a source of contention between the ACS and the American College of Obstetrics and Gynecology (ACOG). The central conclusion of the consensus panel was that after two negative tests, most healthy women should receive Pap smears every "one-to-three years"; in short, the panel failed to resolve this controversy between the ACS and ACOG. The panel noted that the optimal screening frequency could not be determined until research addressed the critical question of "whether carcinoma in situ develops and progresses at different rates in women with different assessed levels of risk."

The Consensus Statement also includes a "minority opinion" written by an economist on the panel. Calling for the increased use of computer models in clinical decisionmaking, he claims that several computer-based, cost effectiveness decision models indicate that yearly screening is justified only for very high risk groups, and that for other groups "screening intervals greater than 3 years could be appropriate."

Perhaps because it failed to shed any new light on the controversy or to reach a consensus, this CDC received little discussion in the medical and scientific literature compared with the other CDCs we have studied. The Consensus Statement appeared in three journals, and a summary version appeared in another. There were also two news articles in major journals. However, we found only five articles citing this primary literature.

PRIMARY LITERATURE

"JAMA" and Science published news articles on the conference (Gunby, 1980, and Marx, 1980). Both Gunby and Marx report that the CDC panel members were divided on the issue of the optimal frequency for Pap smears. The wording of the summary statement—the tests should be repeated "at regular intervals of one to three years"—shows no preference for either view. According to Gunby, the wording was arrived at after an "all night discussion."

SECONDARY LITERATURE

We found six citations to the Consensus Statement, its reprints, and the two news articles. These citations appeared in five articles: a letter, two research reports, an editorial, and a clinical review. None of the citing articles refer to the primary literature published in regional journals.

The source and context for each citation follow:

1. An editorial by Romm, Fletcher, and Hulka (1981) describes the "problems and challenges" in implementing prevention programs in clinical practice. The authors illustrate one of these problems, disagreement among experts, with the controversy over the frequency of Pap smears. After listing the ACS and ACOG positions (and others), they conclude by saying that the CDC "hedged at one to three years," while citing the Marx report in Science.

2. A letter by Parkin and Clayden (1980) cites the British Medical Journal reprint of the Consensus Statement and was clearly occasioned by that publication. It begins by noting the degree of uncertainty about the appropriate screening interval and quoting the "one to three year" recommendation. The authors "support" the minority opinion of Dr. Neuhauser about the value of using computer-based policy models to understand the dynamics of screening programs and the sensitivity of outcomes to assumptions. They further note that one of the values of "consensus statements of the type made by NIH" is that they can provide data or validation of such models.
3. Douglass (1981) writes to inform readers of the *Mayo Clinic Proceedings* of his views concerning examining healthy patients. Among the issues he treats is that of the new ACS guidelines, which he notes are controversial in a variety of ways. He cites the Gunby article to illustrate one of these controversies, the frequency of Pap smears. He continues this section with the "not unanimous" opinions of members of the Mayo Clinic's Division of Preventive Medicine concerning cancer screening. These opinions include the desirability of annual Pap smears.

4. Kizilbash and Mettlin (1981) report on the costs and effectiveness of a program for selective Pap smear screening of outpatients. The program began in April 1979. A Pap smear history is taken from each female patient of age 18 or more. If she "has not had one in the previous 12 months, she is offered one. If she has not had one taken in at least three years, a current Pap smear is recommended to her. This approach is consistent with the one to three year screening intervals recommended by the NIH Consensus Development Conference on Cervical Cancer Screening and the American Cancer Society." Gunby is cited here. Kizilbash and Mettlin recommend their program to community hospitals as an inexpensive way of reaching the underscreened.

5. Beilby et al. (1982) report on the effectiveness of two types of cervical spatulas and on the reduction in false negatives produced by taking two samples at a time. Citing Marx, they note the controversy over appropriate Pap smear frequency. They suggest that increased frequency is one way to reduce false negatives, but acknowledge that it has some drawbacks. They note that several groups recommend that an initial negative smear be repeated one year later. Here the authors cite the *British Medical Journal* reprint of the Consensus Statement. Beilby et al. recommend the alternative of taking two smears at the same visit.
CONCLUSIONS

Our review and analysis of the literature suggests the following conclusions:

1. The CDC on Cervical Cancer Screening apparently contributed little to the substantive discussion of the issue of Pap smear frequency. It publicized a controversy rather than a consensus. This controversy was the focus of the two news articles and of most of the small number of citations.

2. A possibility remains that this CDC will contribute to the future acceptance of rational models of clinical decisionmaking. Dr. Duncan Neuhauser, author of the Minority Opinion endorsing the use of computer-based models, claimed that the conference "will be viewed as one of the historical landmarks in this transformation in clinical reasoning." The letter by Parkin and Clayden recalls the value of explicating the process of clinical decisionmaking. This is the only citation to this, in sharp contrast with the 25 citations to the Weinstein presentation on cost benefit analysis to the estrogen CDC. Perhaps this is because Dr. Neuhauser never published the details of his computer model so that others could comment on its assumptions or use some of its features for further work.

BIBLIOGRAPHY

Primary Literature


National Institutes of Health Consensus Development Panel Summary,

National Institutes of Health, "Cervical Cancer Screening: The Pap
Smear," *Journal of the Tennessee Medical Association*, Vol. 74, No. 5,

Secondary Literature

Beilby, J. O. W., R. Born, J. Builleboud, and S. T. Steele, "Paired
Cervical Smears: A Method of Reducing the False Negative Rate in

Douglass, B. E., "Examining Healthy Patients: How and How Often," *Mayo

Kizilbash, M., and C. Mettlin, "Pap Smears: Access to Hospital
Outpatients of Nongynecological Services," *Journal of Surgical

Parkin, D. M., and A. D. Clayden, "Cervical Cancer Screening," *British

Romm, F. J., S. W. Fletcher, and B. S. Hulka, "Prevention in Clinical
Practice," *Archives of Internal Medicine*, Vol. 141, No. 11, 1981,
p. 1418.
IX. CHILDBIRTH BY CESAREAN DELIVERY

Between 1970 and 1978, the cesarean birth rate in the United States increased from 5.5 percent to 15.2 percent. This major shift in obstetrical care delivery became a major concern to both the medical profession and the public.

The American College of Obstetricians and Gynecologists, prompted by their own concerns and the interest of federal agencies, consumer groups, patient advocates, and physicians, argued the need for an in-depth unbiased study of the use of cesarean section. The College suggested that the National Institute of Child Health and Human Development consider sponsoring a consensus development conference on the medical, social, economic, and ethical effect of the rising cesarean section rate.

A Consensus Development Conference on Childbirth by Cesarean Delivery was held at NIH on September 22-24, 1980. It was sponsored by the National Institute of Child Health and Human Development (NICHD) in conjunction with the National Center for Health Care Technology (NCHCT) and with the assistance of the Office for Medical Applications of Research, NIH.

The conference addressed the following questions:

1. Why and how have cesarean delivery rates changed in the United States and elsewhere, and how have these changes affected pregnancy outcomes?
2. What is the evidence that cesarean delivery improves the outcome of various complications of pregnancy?
3. What are the medical and psychological effects of cesarean delivery on the mother, child, and family?
4. What economic factors are related to the rising cesarean rate?
5. What legal and ethical considerations are involved in decisions on cesarean delivery?
The short (two and a half days) format of the CDC did not permit it to address all aspects of this wide agenda fully, and the bulk of the discussion focused on clinical issues.

The consensus panel issued a fairly long (10 page) Consensus Statement, "Cesarean Childbirth," which summarizes its conclusions and recommendations regarding the questions listed above. NICHD's Office of Research Reporting also made available a full task force report that includes a description and analysis of the literature and the data on which the Consensus Statement is based. Along with the Consensus Statement and the task force report, a report from a review panel that evaluated the consensus process was also published as part of the conference summary.

The overall conclusion of the consensus panel was that "this trend of rising cesarean birth rates may be stopped and perhaps reversed, while continuing to make improvements in maternal and fetal outcomes." The panel also noted that longitudinal studies, although expensive and difficult, are needed to resolve most pivotal issues concerning cesarean delivery.

Despite its length, the Consensus Statement was widely reprinted: it appeared in a national journal, four specialty journals, and three state journals.

**PRIMARY LITERATURE**

Eleven primary references were located. Eight of these are reprints of the Consensus Statement. Three are commentaries.

One reprint of the Consensus Statement appeared in a major clinical journal (*British Medical Journal*) and three appeared in state medical journals (*Journal of the Tennessee Medical Association, Connecticut Medicine, Maryland State Medical Journal*). Reprints also appeared in two major obstetric journals (*Obstetrics and Gynecology, American Journal of Obstetrics and Gynecology*) and two other specialty journals (*Clinical Pediatrics* and *Journal of Reproductive Medicine*). Three of these journals (*Obstetrics and Gynecology, American Journal of Obstetrical Gynecology, British Medical Journal*) have large readerships.
The remaining three primary articles are commentaries that include varying degrees of criticism of the conference.

A report on the conference in *Science* (Kolata, 1980) focuses on the recommendations regarding the indications for cesarean section. It points out that the consensus panel was hampered by the lack of information on the effects of cesarean birth.

Panel member Jeanne Guillemin, in an article discussing who should control the decision regarding cesarean delivery (Guillemin, 1981), asserts that the summary statement "focuses almost exclusively on the clinician's rationale." Her article examines the complex interplay of four types of influence on the rate of cesarean delivery: professional authority, hospital resources, reimbursement policy, and public opinion.

Summey and Hurst (1981) also touch on the theme that the CDC represents the views of physicians to the exclusion of other groups. They are critical of both the panel membership and the consensus process.

SECONDARY LITERATURE

We found 17 secondary articles. Nine of these present original research, four are reviews, and four are editorials on the subject.

These articles refer to the primary literature as a source of information on the trends in and determinants of cesarean section, or to specific recommendations regarding the indications for cesarean delivery. CDC findings that deal with the health effects of cesarean section and with ethical and economic considerations have apparently had little effect on the subsequent literature.¹

Although the Consensus Statement contains many recommendations that bear on research priorities and policies, we find little evidence that these recommendations have affected the subsequent research literature. Instead, research articles cite the Consensus Statement as a source of epidemiological evidence and for its recommendations regarding specific interventions. However, given the long time that typically elapses

¹The primary articles by Summey and Hurst (1981) and by Guillemin (1981) briefly touch on the ethical issues raised by the Task Force.
between the development of research proposals and the publication of results, one might not expect to see CDC-generated research results until several years after the conference.

No one source in the primary literature dominates citations to the conference; the research literature cites both the reprints of the Consensus Statement that appear in major medical journals and the Consensus Statement itself, as well as the task force report.

The sources and contexts for each of the citations in the research articles are:

1. Benedetti, Platt, and Druzin's (1982) study of vaginal delivery after previous cesarean section (VDAC) cites the CDC summary published in the American Journal of Obstetrics and Gynecology as showing the need to reevaluate the use of repeat cesarean section.

2. Polk et al. (1982) cite the Science editorial (Kolata, 1980) as a source of data on the increasing cesarean section rate.

3. Discussant Jacques F. Roux questions Barton, Garbaciak, and Ryan's (1982) use of the cesarean delivery rate as a yardstick for measuring the usefulness of x-ray pelvimetry in evaluating cephalopelvic disproportion. In support of his argument, Roux cites the Statement summary that appeared in the Journal of Reproductive Medicine to the effect that 60 percent of all cesarean deliveries are performed for reasons that have nothing to do with cephalopelvic disproportion.

4. Phillips, Thornton, and Gleicher (1982) cite both the NIH summary and the task force report as background and motivation for their own research on tests for possible physician biases on the decision to perform cesarean deliveries.

5. Rosenberg, Hepburn, and McIlwaine (1982), reporting results from an audit of cesarean section in a large maternity hospital in Scotland, refer to the British Medical Journal's (1981) publication of the CDC summary as a source of epidemiological data and as justifying research into the effects of cesarean birth.
6. Gellman et al.'s (1983) study of VDAC outlined in citation 1 above cites the NIH summary both as a source of epidemiological data and for its recommendation regarding VDAC, which the authors believe is supported by their own data.

7. Williams and Chen (1983), in an article that suggests the publication of hospital-specific cesarean rates as a method of stemming the rapid rise in cesarean birth rates, cite the task force report to document the explosive growth in the cesarean delivery rate during the 1970s.

8. In an earlier article on perinatal mortality in California, Williams and Chen (1982) cite the NIH summary statement to document the existence of concern and controversy over the increasing cesarean delivery rate. Their findings indicate that during the 1970s, mortality rates in California declined more rapidly for cesarean deliveries than for vaginal births, a fact that they view as at least moderately reassuring. They also note sharp differences between their own results and those based on New York City data that were reported at the consensus conference. For births involving breech presentation, they find neonatal mortality rates to be four times lower for cesarean births than for vaginal births, even among low birth weight infants, whereas in the New York City data there were no consistent differences for low birth weight infants. Differences in the relative rates of decline in neonatal mortality among different birthweight groups are also noted.

9. Robert P. Pulliam, discussing an article by Amirikia, Zarewych, and Evans (1981), cites the draft version of the task force report and reviews its major recommendations in some detail. He contrasts it favorably with the Amirikia and colleagues article regarding the report's recognition that the cesarean section rate must be viewed in the context of maternal and fetal outcomes.
Four review articles also cite the primary literature. A review of pelvimetry, published in a radiology journal (Bean and Rodan, 1982), cites the American Journal of Obstetrics and Gynecology (1981) summary of the CDC for its epidemiological data and discussion and recommendation regarding failure to progress and breech presentation. An article in Medical Hypotheses on medical practitioners' resistance to new ideas (Forman, 1981) cites Kolata's Science editorial with respect to the relationship between electronic fetal monitoring and cesarean section rates. Two newsletters from the American College of Obstetrics and Gynecologists cite the task force report on cesarean birth. The first cites the conclusion that vaginal delivery after cesarean childbirth is an option appropriate for consideration by obstetricians. The second newsletter cites the report as providing data on the rapidly increasing cesarean section rate and the role of dystocia in the overall rate.

Two editorial reactions to the primary conference literature (Pearse, 1981; Garlinghouse, 1981) appeared in a single issue of the Nebraska Medical Journal. Both address the increasing cesarean delivery rate and both recommend the CDC summary to other clinicians. The articles focus specifically on the indications for cesarean birth. A third editorial, in the Rhode Island Medical Journal (1982), outlines some of the recommendations of the CDC, especially with respect to indications. The final editorial (Pearse, 1983) suggests that the conference was a response to the rapidly increasing cesarean section rate during the 1970s, and suggests that since the year of the conference (1980), published articles have been mostly concerned with relating cesarean deliveries to perinatal mortality, examining hospital characteristics associated with different cesarean rates, and implementing one of the recommendations to emerge from the conference—a trial of vaginal delivery following a previous cesarean delivery.
CONCLUSIONS

Our findings lead us to the following conclusions:

1. The CDC on Cesarean Childbirth has been recognized in the literature primarily as a source of information on the increasing cesarean birth rate. The Task Force brought together a large amount of data, which it summarized concisely and effectively. For that reason, the conference probably deserves much of the credit for the widespread recognition that cesarean delivery rates had been increasing quite rapidly during the 1970s and that this was a phenomenon to be concerned about. By pointing out that the increase in the cesarean birth rate is directly related to the handling of four specific indications (previous cesarean section, fetal distress, breech presentation, and dystocia), the Consensus Statement engendered both discussion and research on the current obstetrical approach to those problems. Through the period covered by our review, however, it failed to prompt additional published research on the outcomes of obstetrical care for either mother or child, despite the fact that both the CDC report and the Science article (Kolata, 1980) explicitly noted the need for such research. However, our review of published literature may not reflect ongoing but as yet unpublished research prompted by the CDC.

2. Although no summary of the task force report was published, the full task force report served as a source of information to the profession.

3. The Consensus Statement was widely reprinted. Two major obstetric journals carried it, but the only major journal aimed at general practitioners that reprinted the Statement was the British Medical Journal, whose audience in this country is limited. Several state medical journals also reprinted the Statement.
4. Commentaries by members of the medical profession on the cesarean delivery CDC focus mainly on the epidemiological aspects of the report and its specific recommendations regarding indications for cesarean section. Few citing authors who are medically oriented appear critical of the consensus process or the composition of the panel. However, cesarean childbirth has in recent years become a social and women's issue, as well as a medical issue. In that larger context, there has been some criticism of the CDC for not including more nonphysicians in the consensus process.

BIBLIOGRAPHY

Primary Literature


**Secondary Literature**


X. CORONARY ARTERY BYPASS SURGERY

Coronary artery disease is a major cause of mortality and morbidity. In 1980 there were two general approaches to the management of this disease: (1) coronary artery bypass surgery (CABS) and (2) medical therapy, which may include long-term use of medications, alterations in life style, and various measures to control risk factors for atherosclerosis. Bypass surgery was developed in the 1960s and refined in the 1970s. The use of CABS increased rapidly, with an estimated 100,000 procedures performed in 1979 at a cost of approximately $2 billion. Meanwhile, significant advances were being made in medical therapy during the 1970s. By the end of the decade, the stage had been set for a controversy over the relative advantages of medical versus surgical therapies for the management of various types of coronary artery disease.

A Consensus Development Conference on Coronary Artery Bypass Surgery was held at NIH on December 3-5, 1980. It was sponsored by the National Heart, Lung, and Blood Institute in conjunction with the National Center for Health Care Technology and conducted with the assistance of the Office of Medical Applications of Research, NIH.

The conference panel addressed the following questions:

1. What is the overall management of coronary artery disease; that is, in what context should coronary artery surgery be considered?
2. What constitutes a reasonable diagnostic workup before recommending medical or surgical therapy?
3. What is known about long-term survival after CABS in specific patient groups?
4. What is known about long-term quality of life after CABS?
5. What is the range of success rates for the procedure in various settings, and what factors may be important in influencing this outcome?
The panel issued a Consensus Statement, "Coronary Artery Bypass Surgery: Scientific and Clinical Aspects," which summarizes its findings. Very briefly, the consensus panel concluded that CABS "represents a major advance in the treatment of patients with coronary artery disease." At least in selected groups of patients, CABS has improved the quality of life, decreased myocardial ischemia, and increased survival.

The Consensus Statement contains few specific clinical recommendations regarding the use of coronary artery bypass surgery. The main thrust of the statement is that recommendations for medical or surgical therapy are based on two fundamental questions. The first concerns which therapy provides the greatest protection from disabling myocardial infarction or death. The second question concerns which therapy provides the best quality of life for the patient. The panel notes that the literature dealing with the first question is complex and often contradictory. It also suggests that there is very little evidence regarding the effects of these therapies on quality of life.

Based on the available data, the panel concludes that CABS is appropriate for patients with severe disease of the left main coronary artery who are at high risk for infarction and sudden death, and for patients with unacceptable symptoms who are receiving adequate medical therapy. The panel indicates that surgery is inappropriate for patients with noncritical stenosis and controlled symptoms. Recommendations for patients between those extremes are hedged. Often, the panel cites inadequate evidence and the need for further investigation of the role of patient characteristics.

Although it contained no surprising conclusions, the Consensus Statement was widely reprinted. It appeared in complete form in two specialty journals and three regional journals, and a national journal published a slightly revised version. Five commentaries on the conference findings also appeared, including four in national medical journals and one in a premier general science journal.

In addition, the journal Circulation published a supplement devoted to the CABS conference. Unlike the supplement to Cancer, which was devoted to the CDC on steroid receptors, this supplement to Circulation
does not reprint the papers presented at the conference before the panel. Rather, it reprints the Consensus Statement from the conference together with 21 new papers, including two by conference panelists.

Because of the distinctive way in which this CDC affected the professional literature, this section is also organized distinctively. The next subsection reviews the primary literature related to the Conference, with the exception of material appearing in the Circulation supplement. We then analyze the citations to this primary literature and the citations to the articles in the supplement.

PRIMARY LITERATURE

We found 11 primary articles. Six are reprints of the Consensus Statement. One appeared in NEJM, two in specialty journals (Circulation and the Journal of Cardiovascular Medicine), and three in regional journals (Connecticut Medicine, Journal of the Florida Medical Association, and North Carolina Medical Journal). The reprint published in Circulation was part of the special supplement on the consensus conference.

The other five primary articles are commentaries on the conference and the Consensus Statement. All of these appeared in prominent publications: one in NEJM (Rennie, 1981), one in Annals of Internal Medicine (Rahimtoola, 1981), one in Science (Kolata, 1980), and two in JAMA. Summaries of the commentaries follow:

- In his NEJM editorial, Drummond Rennie writes that the conference was valuable as "an event" marked by vigorous debate on worthy problems. Although the conclusion that CABS represents "a major advance" is not "startling," the facts are complex and the Statement poses as many questions as it answers. The Statement, printed by NEJM in a slightly revised form, provides practitioners with a "benchmark" that can guide them in the field. Although Rennie's editorial praises the CDC on CABS, it condemns some other CDCs and outlines standards for effective and useful conferences and reports. He explains that the NEJM does not publish all CDC Consensus Statements for reasons that range from the "feeling that our readership has
heard it all before to the impression that the panel has come up with nothing useful." He disapproves of bland generalities "so well established that surely everyone must already know them." Reports instead should stimulate debate and avoid being too prescriptive. More specifically, Rennie wishes "fewer topics were chosen and ways would be found to generate new knowledge, for example, by making fuller recommendations about future research projects in their areas." At their best, the Consensus Statements represent "an attempt by a panel of honest experts to reach an agreement on what we think we know and what we do not know about important medical topics."

- G. B. Kolata's (1981) article in Science emphasizes the conference as an arena for discussion more than as an attempt to make specific recommendations. Her article focuses on the issues relating to the medical management of patients. She is especially interested in the animated discussions on the psychological factors surrounding surgery, and prints off-the-record remarks by five panelists and statements from five eminent nonpanelist scholars. Reviewing the discussions, Kolata shows that surgery may be chosen for quite subjective reasons by both patient and physician. Quality of life is difficult to measure, and patients who have had surgery may not be completely honest with their physicians about their symptoms. That CABS does not appear to increase a patient's chances of resuming gainful employment may reflect a failure of surgery to relieve symptoms as much as commonly believed, or it may reflect other factors unrelated to the patient's postsurgical symptoms (e.g., a belief that the stress of work helped cause the heart disease in the first place). After casting doubt on the capacity of objective criteria to provide adequate indicators and evaluation of CABS, Kolata closes by noting that a second conference on CABS, scheduled for April 1981, "will emphasize the economic, social, and ethical issues arising from bypass surgery and will address the question left unanswered by Rapaport's statement: How much of an improvement in quality of life, and at what costs?"
• Like Kolata, Johnson (1981) emphasizes the conference as a forum for discussion. He finds the conference's conclusions valid but considers its recommendations conservative. His report includes technical data omitted from the Consensus Statement, and, like Kolata, he captures the quality of debate at the conference by quoting from the panelists Frye, Meier, William Sheldon, and W. Gerald Austen. He barely touches on medical treatment, but focuses instead on details of cardiac disease and graft technique in relation to successful outcomes. Like Kolata, he indicates that surgery is chosen not so much to extend survival as to relieve pain and improve quality of life. And he also refers, as a postscript, to the second conference.

• The article by Rahimtoola (1981) provides a summary of the Conference Statement. He notes that the results for surgery have improved during the 1970s and that the use of surgery should be reassessed as more data become available.

• The report from the NCHCT provides a summary of an assessment forum it ran on coronary bypass surgery. This forum drew on the consensus conference but dealt more with the economic and clinical aspects of the surgery than with scientific and clinical aspects.

SECONDARY LITERATURE

We located 21 articles citing the primary literature. Three of these articles cite more than one of the primary articles. Nine of the secondary articles cite the Consensus Statement in some form. Seven of these cite the slightly edited version published in NEJM, one cites the Consensus Statement itself, and another cites the reprint in the Journal of Cardiovascular Medicine.

The Consensus Statement is cited for a variety of reasons:

• Four of the citations refer to survival. Boulay, David, and Bourassa (1982) restated the panel's overall conclusion regarding survival; Freeman and Freeman (1982) note its conclusion regarding single vessel disease; Levy (1981) cites
the increased survival in left main disease; and Kornfeld et al. (1982) note the increased survival but also cite the need for further research on the quality of life;

- Blumstein (1982) refers to the importance of communication between physician and patient in the treatment of coronary artery disease;
- In a review of randomized trials in the treatment of coronary artery disease, Fisher and Kennedy (1982) note that "expert views" such as those expressed by the CDC put a great deal of weight on the results of randomized trials;
- Bourassa (1981) uses the question format of the CDC to review the status of coronary artery bypass surgery;
- In the introduction to the Circulation supplement, Frye, Frommer, and McCallum (1982) note that the Statement was widely circulated, referring to the publication of the Statement itself in NEJM and Journal of Cardiovascular Medicine and to commentaries by Rahimtoola (1981) and Johnson (1981);
- Cutter et al. (1982) refer to the incidence of repeat procedures.

Overall, the secondary literature (especially the article by Fisher and Kennedy, 1982, and the article by Freeman and Freeman, 1982) treats the statement as important. Moreover, the CDC may have influenced the way in which Bourassa (1981) structured his review of coronary bypass surgery, or helped frame the content in which Kornfeld et al. (1982) interpreted their research on quality of life.


- Frye, Frommer, and McCallum (1982) mention it as demonstrating the wide distribution of the panel's recommendations;
- Carr and Rau (1981) cite it as concluding that coronary artery bypass surgery is a major advance in the treatment of occlusive coronary artery disease;
Hecht and Rahimtoola (1982) refer both to Rahimtoola's editorial and to the Consensus Statement reprint published in the *Journal of Cardiovascular Medicine* as recommending the use of coronary angiography in all patients with unstable angina;

Gohlke et al.'s (1983) article on exercise testing cites Rahimtoola (1981) regarding improvements in effort angina noted after surgery;

The final two references to the editorial appear in review articles by Rahimtoola (1982a, 1982b) and refer to the review of CABs undertaken by the NIH.

The Rahimtoola (1981) editorial differs from the other commentaries on the conference in that it sticks quite closely to the content and structure of the statement itself rather than discussing the conference proceedings and peripheral issues that were discussed at the conference but not covered in detail in the statement. The references to the editorial cite it as a source of justification for stated facts.

Seven articles cite the commentaries in *Science* and *JAMA*. Six cite the *Science* article by Kolata (1981).

Three of these (Holley, Ponganis, and Stanski, 1982; Weinstein and Stason, 1982; Rotstein and Holley, 1982) refer to the commentary as providing an estimate of approximately 100,000 CABs being performed per year.

Jensen, Clayton, and Liddle (1982) cite the commentary as demonstrating the controversy over CABS.

Kornfeld et al. (1982) note the possibility, mentioned by Kolata, that some patients opt too quickly for surgery.

Discussing dialysis, Johnson, McCauley, and Copley (1982) noted the importance of quality of life assessments in escalating radial interventions.

Harrison mentions the Kolata article in a bibliography on bypass surgery with no specific reference.
The single reference to Johnson's (1981) commentary in JAMA notes the increased survival for patients with left main and triple vessel disease.

CIRCULATION SUPPLEMENT

In June 1982 the journal Circulation published a supplement devoted to the CDC on coronary artery bypass surgery. The supplement contains an introduction, 21 articles on coronary bypass surgery, and the Consensus Statement. The papers emphasize the five questions addressed by the conference panel. The introduction briefly outlines the consensus process and describes the relationship between the questions addressed by the panel and the articles in the supplement. Two of the articles were written by panelists and the remainder by other researchers and practitioners in the field.

Neither the introduction to the supplement nor the Consensus Statement contained in the supplement were cited by subsequent authors. The other articles in the supplement, categorized by topic, and their subsequent citations are shown in Table 10.1; 93 citations to the supplement articles appeared from June 1982 to June 1984. The majority of the articles address the questions of survival and quality of life after surgery, and these articles have generated the largest effect on the literature. Those articles that review the management of patients with coronary artery disease (Reeves, 1982; Willman, 1982; Froelicher, 1982; Friesinger, 1982) have generally had a smaller influence on the subsequent literature than have articles that present original research results. The most cited article (Takaro et al., 1982) presents some results of a large randomized trial of coronary artery bypass grafting. The article by Bourassa et al. (1982), cited 13 times, deals with graft patency and outcomes at followup.

CONCLUSIONS

Our findings lead us to the following conclusions:

1. The primary literature seems to have been valuable chiefly as a review of the state of knowledge on coronary bypass surgery:
Table 10.1
CITATIONS TO CIRCULATION SUPPLEMENT ARTICLES, 1982 THROUGH JUNE 1984

<table>
<thead>
<tr>
<th>Primary Article by Topic</th>
<th>1982</th>
<th>1983</th>
<th>1984</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall management</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reeves</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Wilman</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Diagnostic workup</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Froelicher</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Friesinger</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Long-term survival</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rosati et al.</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Ware</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cosgrove, Loop, and Sheldon</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Lawrie and Morris</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Whalen et al.</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Hammermeister, DeRouen, and Dodge</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Takaro et al.</td>
<td>2</td>
<td>8</td>
<td>7</td>
<td>17</td>
</tr>
<tr>
<td>European Coronary Surgery Study Group</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Russell, Rackley, and Kouchoukos</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>McIntosh</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Long-term quality of life</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bourassa et al.</td>
<td>0</td>
<td>7</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>Chesebro et al.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Rahimtooala</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Oberman et al.</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Smith et al.</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Institutional variations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ochsner</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Kaiser</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>46</td>
<td>33</td>
<td>93</td>
</tr>
</tbody>
</table>

- The clinical recommendations received little attention in the subsequent literature, perhaps because they were not specific;
- The Consensus Statement appears to have had little effect on subsequent research;
- The commentaries on the conference were cited more frequently than the Consensus Statement itself; most citing articles referred to the commentaries as sources of information on
survival and on the number of procedures performed in the United States.

2. Most articles citing the Consensus Statement referred to the reprinted statement in the *NEJM*. This suggests the importance of publishing CDC findings in major medical journals.

3. The articles in the *Circulation* supplement seem to have had a larger influence on the literature than did the Consensus Statement; 93 citations to the supplement articles had appeared within two years after its publication. The vast majority of these references are in research articles and reviews. The primary articles receiving the most subsequent citations are the supplement articles reporting original research rather than the review articles. There are several possible reasons for this. One is that clinical reviews are aimed at clinicians (who seldom cite previous work) more than at researchers (who often do). A second reason is that in a field such as the treatment of coronary artery disease, in which both medical and surgical therapy are evolving rapidly, clinical reviews are rather quickly dated.

BIBLIOGRAPHY

Primary Literature

A. Statement and Reprints


Of the nine subsequent citations to the Consensus Statement, seven cite this article.

**B. Commentaries**


Mentions psychological factors and medical therapy.


Discusses not only the coronary bypass conference but also the consensus processes and previous reports.

**C. Circulation Supplement**


Introduction to the supplement that mentions the conference and relates each paper to questions addressed by panel.

Report from a large followup study.


Progress report from another large randomized trial.


Overview of appropriate workup.


Overview of medical management.


Report from a large randomized trial. Most cited references.


Overview of surgical management.
Secondary Literature


Cites conference as an important recent review.


Cites NEJM and *Journal of Cardiovascular Medicine* Consensus Statements as well as commentaries by Rahimtoola and Kolata.


Cites both summary in *Journal of Cardiovascular Medicine* and Rahimtoola's editorial.


Cites both *NEJM* consensus reprint and Kolata commentary.


XI. SYNTHESIS AND CONCLUSIONS

In this section we synthesize the findings from our case studies and draw out their implications. We begin by summarizing our findings on the dissemination of the sample conferences' recommendations through what we have labeled primary publications. We next discuss the frequency of citations to each conference and the likely explanations for the large variance across conferences in citation rates. Then we categorize the context of the citations and explore how and why the conferences differed in the extent of the controversy that surrounded their recommendations. Finally, we consider how our findings relate to NIH's strategies for selecting conference topics, managing the conferences, and disseminating the findings.

OVERVIEW OF THE PRIMARY LITERATURE

As described in Sec. II, the primary literature for a CDC can be divided into two categories: (1) primary literature that is derived directly or with only minor changes from products of the consensus process, and (2) primary literature that discusses or describes the conference proceedings but that differs considerably in nature or content from direct products of the consensus process. This division separates the primary literature into a category that was produced or sponsored either by the panel or by the NIH as part of the consensus process and a component that was not produced under those auspices. It is a conceptually important division in the sense that the first category is largely controllable by those in charge of the panel process, whereas the second category is much less so.

The first category includes: (1) verbatim copies of the final statements or edited summaries of the statement, (2) task force report summaries, and (3) sponsored background papers or presentations.

Task force reports were produced for the conferences on antenatal diagnosis and cesarean childbirth. The task force report summaries were similar to the consensus statement summaries in that they included the conference's final recommendations. However, they contained supporting
material not found in conference summaries. Because they contained the
CDC's final recommendations, the task force report summaries are
included with the conference summaries in our analysis.

Rounding out the first category of primary literature are
presentations, defined as published background studies from a
conference. Presentations are similar to the task force reports in that
they provide an overview of evidence related to the topic, but they do
not include the panel recommendations. Presentations were published
from steroid receptors, estrogen use, and coronary artery bypass surgery
conferences. Because these presentations were published under the
auspices of the conference, they are considered part of the first
category of primary literature. Because they do not include the panel
recommendations, they are analyzed separately from the summaries and
task force reports.

The second category of primary literature includes news articles in
professional journals such as Science, commentaries, and discussions of
the CDC process and recommendations along with comments made by
panelists after the conference. These commentaries are divided into two
components: those written by panelists or NIH officials and those
written by others. This division recognizes that panelists and NIH
officials may be addressing audiences different from those other
commentators address.

Overall Patterns in Primary Literature

The eight CDCs reviewed in this project received different amounts
of coverage as defined by their primary literature. Differences are
apparent in both the primary literature and in the accompanying
commentaries. Table 11.1 shows the number of primary articles for each
conference. The articles are classified as direct CDC products
(summaries or presentations) or as commentaries (subdivided by whether
CDC panelists or NIH officials wrote them or nonpanelists wrote them).
The number of published summaries ranged from a high of eight for the
antenatal diagnosis and cesarean birth conferences to a low of one for
the conference on steroid receptors in breast cancer. Recall that seven
task force report summaries were published in connection with the
antenatal diagnosis conference and that these task force report
Table 11.1
CLASSIFICATION OF PRIMARY ARTICLES

<table>
<thead>
<tr>
<th>Conference</th>
<th>CDC Direct Products</th>
<th>Commentaries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Summaries</td>
<td>Presentations</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breast cancer</td>
<td>2 0</td>
<td>3</td>
</tr>
<tr>
<td>Steroid receptors</td>
<td>1 42</td>
<td>0</td>
</tr>
<tr>
<td>Estrogen use</td>
<td>3 3</td>
<td>1</td>
</tr>
<tr>
<td>Thrombolytic therapy</td>
<td>6 0</td>
<td>0</td>
</tr>
<tr>
<td>Pap smear</td>
<td>4 0</td>
<td>0</td>
</tr>
<tr>
<td>Cesarean delivery</td>
<td>8 22</td>
<td>1</td>
</tr>
<tr>
<td>CABS</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

a A single supplement of Cancer was devoted to the conference presentations and contained 42 articles.

b A single supplement of Circulation was devoted to the CDC and contained 22 articles.

summaries contained more information than found in the conference summary. Three conferences published supplementary material based on conference presentations. The conferences on steroid receptors in breast cancer and coronary artery bypass surgery each led to the publication of a journal supplement containing material presented at the conference. Three articles reported the formal presentations at the conference on estrogen use in menopausal women.

Only the conference on thrombolytic therapy did not lead to published commentaries. The CDC on antenatal diagnosis yielded seven commentaries. The CDC on the treatment of primary breast cancer resulted in six commentaries. Three of these were written by panelists with differing views and in some sense can be thought of as minority reports.

All the CDCs resulted in the journal publication of at least one summary statement, sometimes after light editing. Table 11.2 shows the types of journals publishing summaries or presentations. Many of the summaries were published in regional medical journals. In no case was
Table 11.2
TYPES OF JOURNALS PUBLISHING CONSENSUS STATEMENTS AND CONFERENCE PRESENTATIONS

<table>
<thead>
<tr>
<th>Conference</th>
<th>National</th>
<th>Regional</th>
<th>Specialty Journals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antenatal diagnosis</td>
<td>0</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Steroid receptors</td>
<td>1</td>
<td>0</td>
<td>1&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Estrogen use</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Thrombolytic therapy</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Pap smear</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Cesarean delivery</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>CABS</td>
<td>1</td>
<td>3</td>
<td>2&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Journal supplement.

<sup>b</sup>Journal supplement containing both research papers and Consensus Statement.

the summary published in more than one national medical journal. (The second article on estrogen use in menopausal women was a conference presentation, not a summary.) The most widely published conference summaries were for antenatal diagnosis and cesarean childbirth. The majority of summaries for these two conferences were published in specialty journals.

The potential audience for a conference's primary literature depends not only on the number of journals publishing summaries, but also on the characteristics of those journals. The journal's circulation is one obviously important factor, but there are others as well, including the prestige of the journal and its subject area. Thus, although the CDC summary for steroid receptors was published in only the New England Journal of Medicine, the status and readership of that journal are such that the statement was guaranteed a wide and diverse clinical audience. We expect that the size and type of the audience is related to our four-part categorization of biomedical journals (medical specialty journals, national general medical journals, regional medical
journals, or basic science journals). In particular, we expect that the publication of CDC summaries in such regional medical journals as Connecticut Medicine, Journal of the Tennessee Medical Association, and Southern Medical Journal, while making a positive contribution to the dissemination of consensus conference findings, will have considerably less influence than publication in a national clinical journal, for reasons of both circulation and prestige. We also expect that topics that are mainly of interest to specialized practitioners may be most effectively disseminated in journals aimed at those specialties. Indeed, NIH appears to have followed such a principle in its targeting strategy. For example, recommendations of the CDCs on fetal monitoring and cesarean birth were published in the American Journal of Obstetrics and Gynecology, whose readership consists mainly of physicians who provide obstetrical care.

Several CDCs produced background research that was subsequently reprinted in professional journals. The CDC on cesarean birth produced a task force report whose length may have precluded its reproduction in a journal setting. The CDC on estrogens, however, produced three reasonably sized research and review articles that were subsequently published in a major national medical journal and two specialty journals. The supplements presenting background materials for the CDCs on steroid receptors and CABS were published in major specialty journals. The task force report summaries on antenatal diagnosis, which contained research reviews, were also widely disseminated.

For two CDCs, antenatal diagnosis and estrogen use in postmenopausal women, commentaries by NIH officials appeared in major clinical journals. These commentaries were built around the consensus statements but contained considerable background information and comment.

All but one of the eight topics had at least some primary literature that was identifiable as commentary. (The single exception was the conference on thrombolytic therapy.) The conference on antenatal diagnosis yielded seven commentaries. The CDC on treatment of primary breast cancer resulted in six commentaries, three of which were "minority reports" by panel members. The journals that published the greatest number of commentaries were Science and JAMA. However, as
Table 11.3 shows, this type of indirect primary literature appeared in a wide variety of journals.

PATTERNS IN THE SECONDARY LITERATURE

As noted in Sec. II, we searched the Science Citation Index to locate published articles that cited conference reports and commentaries. The number of articles citing this literature and their characteristics help us gauge the breadth of influence of published consensus findings. Our results indicate that the pattern of citations differed widely from conference to conference. Some conferences received far more citations than others. To a varying degree, conferences attracted citations in articles devoted to research, review, or commentary. The emphasis of the journals containing secondary articles also differed in important ways: Some conferences were cited heavily in journals of general clinical interest, whereas others received more citations in specialty journals.

Table 11.3

PUBLISHED COMMENTARIES ON CONSENSUS CONFERENCES BY TYPE OF JOURNAL

<table>
<thead>
<tr>
<th>Conference</th>
<th>General Medical</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>National Regional Specialty Basic Science Not Biomedical</td>
</tr>
<tr>
<td>Antenatal diagnosis</td>
<td>4 0 3 0 0</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>4 0 1 1 0</td>
</tr>
<tr>
<td>Steroid receptors</td>
<td>0 0 0 0 0</td>
</tr>
<tr>
<td>Estrogen use</td>
<td>0 0 2 0 0</td>
</tr>
<tr>
<td>Thrombolytic therapy</td>
<td>0 0 0 0 0</td>
</tr>
<tr>
<td>Pap smear</td>
<td>0 0 1 1 0</td>
</tr>
<tr>
<td>Cesarean delivery</td>
<td>0 0 0 1 2</td>
</tr>
<tr>
<td>CABS</td>
<td>3 0 1 1 0</td>
</tr>
</tbody>
</table>
Number of Citations

We found a total of 618 citations to the 130 primary articles discussing the eight conference topics.¹ These include 471 citations to articles in the Cancer supplement on steroid receptors, the Circulation supplement on CABS, and the three conference presentations on estrogen use that were published after the conference. Clearly, the decision to publish conference proceedings dramatically affected the number of citations received by these consensus topics. The additional citations appear to have provided these conferences with considerable published visibility.

The number of citations ranged widely from topic to topic. As Table 11.4 indicates, primary articles on the topic of steroid receptors received the largest number of citations, followed by CABS. In contrast, primary articles discussing the Pap smear conference received the fewest citations. Others received modest numbers of citations: estrogen use ranked third, followed by treatment of primary breast cancer, antenatal diagnosis, childbirth by cesarean delivery, and thrombolytic therapy.

These rankings generally correspond with the amount of primary coverage received by the various topics, suggesting that a larger primary literature assures more widespread subsequent citation. At the same time, the publication of supplementary conference materials may have widened this influence, for reasons other than adding to the number of primary articles. The primary literature describing the estrogen conference, which included three supplementary publications of conference papers, received more citations than did antenatal diagnosis and cesarean section, which contained more primary articles.

Published reports of some conferences received very little attention in subsequent published literature. If the frequency of citations indicates the visibility of conference findings, the Pap smear conference had the least visibility. Both the Pap smear conference and

¹The number of citations indicates the frequency with which each primary article was cited. Because some secondary articles cite more than one primary article, the number of citing articles is somewhat smaller than the number of citations.
the conference on cesarean delivery also attracted few citations in relation to the number of primary articles.

As noted earlier, the number of citations an article is likely to receive depends on the size and activity of the scientific field it covers as well as how influential it is within that field. To assess an article's effect, it is necessary to compare the number of citations it receives with the numbers received by other articles in its field. In this study, we compared citation rates for primary articles with those for the field as a whole as well as the rate for a comparable "benchmark" review article on the same topic.

The field-specific citation rates and the number of citations found for each of the selected review articles appear in Table 11.5. By both indicators, the literatures concerning antenatal diagnosis, estrogen use, Pap smear, and cesarean delivery appear to be much smaller than those for the remaining topics. When this is taken into account, the citation rate to the estrogen conference appears conspicuously high. The opposite is true in the case of the primary articles for CABS, whose numerous citations look much less impressive when we consider the field-specific citation rate and the number of citations received by the benchmark review article for this topic. The citation rate for the conference on steroid receptors, however, remains impressive relative to both its benchmark review article and its field-specific rate.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Citations</th>
<th>Primary Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antenatal diagnosis</td>
<td>22</td>
<td>15</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>28</td>
<td>8</td>
</tr>
<tr>
<td>Steroid receptors</td>
<td>374</td>
<td>43</td>
</tr>
<tr>
<td>Estrogen use</td>
<td>46</td>
<td>8</td>
</tr>
<tr>
<td>Thrombolytic therapy</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Pap smear</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Cesarean delivery</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>CABS</td>
<td>114</td>
<td>33</td>
</tr>
</tbody>
</table>
### Table 11.5

**COMPARISON OF CITATION RATES TO CONFERENCE PUBLICATIONS WITH INDICATORS OF THE SIZE OF THE FIELD**

<table>
<thead>
<tr>
<th>Topic</th>
<th>Average Citations per Primary Article</th>
<th>Comparison Subfields</th>
<th>Subfield 4-year Citation Rate</th>
<th>&quot;Benchmark&quot; Review</th>
<th>Citations to Benchmark Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antenatal diagnosis</td>
<td>1.5</td>
<td>Pediatrics</td>
<td>6.4</td>
<td>Banta, 1979</td>
<td>7</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>3.5</td>
<td>Cancer</td>
<td>12.2</td>
<td>Henderson, 1980</td>
<td>43</td>
</tr>
<tr>
<td>Steroid receptors</td>
<td>8.7</td>
<td>Cancer</td>
<td>12.2</td>
<td>Jensen, 1981</td>
<td>14</td>
</tr>
<tr>
<td>Estrogen use</td>
<td>5.8</td>
<td>Ob/Gyn</td>
<td>6.3</td>
<td>Ryan, 1982</td>
<td>1</td>
</tr>
<tr>
<td>Thrombolytic therapy</td>
<td>1.8</td>
<td>Gen. and Int. Medical</td>
<td>9.3</td>
<td>Bell and Meek, 1979</td>
<td>20</td>
</tr>
<tr>
<td>Pap smear</td>
<td>1.0</td>
<td>Ob/Gyn</td>
<td>6.3</td>
<td>Am. Cancer Assoc., 1981</td>
<td>7</td>
</tr>
<tr>
<td>Cesarean delivery</td>
<td>1.5</td>
<td>Ob/Gyn</td>
<td>6.3</td>
<td>Harley, 1980</td>
<td>1</td>
</tr>
<tr>
<td>CABS</td>
<td>5.5</td>
<td>Cardiovascular System</td>
<td>10.9</td>
<td>Braunwald, 1978</td>
<td>38</td>
</tr>
</tbody>
</table>

*Average citation during 1974 to 1977 to papers published in 1973 (from Narin and McAllister, 1982).*

### Types of Citing Articles

Table 11.6 presents data describing characteristics of secondary articles on each of the eight topics. We classified each citing article according to whether it was devoted primarily to the presentation of original research results, review of clinical medicine or basic science, or commentary (i.e., letters or editorials). As Table 11.6 indicates, the majority of citations to primary articles occurred in research articles, followed by review articles and commentary. All topics except for thrombolytic therapy received the plurality of their citations to primary articles in this category. Nearly all conferences, therefore, were viewed by citing authors as having contributed to their research literatures. The proportion of research citations indicates this was particularly true for steroid receptors and the treatment of primary breast cancer.

---

2We used a computer-assisted search of *Index Medicus* to obtain abstracts of the large number of citations to the *Circulation* and *Cancer* supplements. Some articles were not abstracted in this database, so we were unable to determine the content of some secondary articles. We read all secondary articles in their entirety for all other conferences.
Table 11.6
NUMBER OF SECONDARY ARTICLES BY TYPE

<table>
<thead>
<tr>
<th>Consensus Topic</th>
<th>Original Research</th>
<th>Review</th>
<th>Commentary</th>
<th>Not Obtainable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antenatal diagnosis</td>
<td>10</td>
<td>10</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>13</td>
<td>5</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Steroid receptors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To statement</td>
<td>25</td>
<td>11</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>To supplement&lt;sup&gt;a&lt;/sup&gt;</td>
<td>215</td>
<td>60</td>
<td>5</td>
<td>58</td>
</tr>
<tr>
<td>Estrogen use</td>
<td>15</td>
<td>14</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Thrombolytic therapy</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Pap smear</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Cesarean delivery</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>CABS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To statement</td>
<td>7</td>
<td>11</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>To supplement&lt;sup&gt;a&lt;/sup&gt;</td>
<td>49</td>
<td>37</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>348</td>
<td>159</td>
<td>30</td>
<td>64</td>
</tr>
</tbody>
</table>

<sup>a</sup>Numbers presented are citations and contain some duplication of articles.

Citations to conference papers in review articles are of particular interest, because these articles are a good vehicle for clinicians to keep informed of advances in medicine. We found that review articles accounted for more than a third of the citations to the primary articles (excluding supplements), but only about a quarter of the citations to supplementary reports are review articles. Primary articles reporting the thrombolytic therapy conference received the largest proportion of review-based citations although, as we have seen, most reviews did not endorse the major recommendation of the consensus panel. Antenatal diagnosis also received a large proportion of review-based citations, followed by CABS and estrogen use. Cesarean delivery, steroid receptors, and breast cancer received proportionately fewer review citations than the other topics.

Published commentary is of interest because letters and editorials often highlight controversial issues. We found that commentary accounted for few secondary citations for any of the topics. Conforming
to the overall citation count, primary articles on estrogen use, CABS, and steroid receptors received the greatest number of citations in this category. However, two of the five citations to the Pap smear conference were classifiable as commentary.

As we noted, the two conferences that published their proceedings received many more citations than any of the other conferences. Table 11.6 shows that the supplementary conference materials received somewhat different types of citations from those for the other primary articles on these two topics. In particular, supplementary materials receive considerable attention in research articles relative to review articles and commentary. At the same time, Table 11.6 also makes clear that the number of citations to these conferences is greater even in the review and commentary categories, suggesting that publication of supplementary material may increase dissemination of information from the conference to clinicians, as well as to the research community.

**Types of Journals**

Articles citing the primary literature appeared in every type of journal identified in our coding scheme (Table 11.7). Specialty journals account for most citations for all topics. Next in importance are clinical journals with a national readership, followed by basic science journals. Secondary articles rarely appeared in regional medical journals, which is interesting given the frequency with which primary articles appear there. However, the primary literature on thrombolytic therapy and on antenatal diagnosis were also fairly well cited in this category.

Table 11.7 reveals patterns of variation by topic in citations by basic science and national clinical journals. For example, the conference on steroid receptors accounted for nearly all citations to primary articles in basic science journals. This is as one might expect, given that steroid receptors drew most heavily of all the conferences on findings from a basic science (in this case, biochemistry). A contrasting example is provided by the conferences on estrogen use—a topic of concern to a broad range of primary care providers as well as specialists; it received a large share of its citations in national clinical journals.
Table 11.7

NUMBER OF SECONDARY ARTICLES APPEARING IN VARIOUS TYPES OF JOURNALS

<table>
<thead>
<tr>
<th>Consensus Topic</th>
<th>Basic Science</th>
<th>Clinical National</th>
<th>Specialty</th>
<th>Regional</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antenatal diagnosis</td>
<td>0</td>
<td>7</td>
<td>14</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>0</td>
<td>4</td>
<td>15</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Steroid receptors</td>
<td>18</td>
<td>2</td>
<td>15</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>To supplement</td>
<td>47</td>
<td>20</td>
<td>270</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>64</td>
<td>22</td>
<td>285</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Estrogen use</td>
<td>3</td>
<td>13</td>
<td>16</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Thrombolytic therapy</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Pap smear</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Cesarean delivery</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>CABS</td>
<td>1</td>
<td>4</td>
<td>14</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>To supplement</td>
<td>0</td>
<td>25</td>
<td>65</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>29</td>
<td>79</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>70</td>
<td>86</td>
<td>424</td>
<td>11</td>
<td>6</td>
</tr>
</tbody>
</table>

*Numbers presented are citations and contain some duplication of articles.

As noted above, most topics were heavily cited in specialty journals. Primary literature on steroid receptors, breast cancer, CABS, antenatal diagnosis, and Pap smear each received over three-fifths of their citations in journals of this type.

Content of Citing Articles

The case studies in Secs. III through X describe how each citing article refers to the primary conference literature, as well as the degree of acceptance and controversy reflected in the subsequent literature. Here we categorize our findings so that the outcomes can be compared across conferences.

The purpose of the categorization is to summarize the types of effects revealed in the citing literature. First, we distinguish three categories of conference content that may be cited--clinical recommendations, research recommendations, and scientific findings.
Each of these categories is then subdivided according to whether the context of the citation shows an affirmation of the CDC panel's judgment or reflects continuing controversy on the issue. The rules we used are summarized in Table 11.8. Note that we categorize any repetition aimed at a clinical audience (typically, an article reviewing the latest developments in the care of particular types of patients) in the same manner as an explicit endorsement of the recommendation—unless, of course, the citing article explicitly disagrees. Citations that report

<table>
<thead>
<tr>
<th>What Was Cited</th>
<th>Context of Citation</th>
<th>Type of Effect Suggested by this Type of Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical recommendation</td>
<td>Explicit endorsement or Repetition without disagreement in article for a clinical audience</td>
<td>Clinical agreement</td>
</tr>
<tr>
<td>Clinical recommendation</td>
<td>Explicit disagreement or Mention of continuing disagreement in the clinical community</td>
<td>Clinical controversy</td>
</tr>
<tr>
<td>Research recommendation or Finding of uncertainty</td>
<td>Research article as either background or motivation</td>
<td>Research motivation</td>
</tr>
<tr>
<td>Research recommendation</td>
<td>Explicit disagreement with need for or value of research</td>
<td>Research controversy</td>
</tr>
<tr>
<td>Scientific finding</td>
<td>Either background or motivation or repetition without disagreement</td>
<td>Scientific agreement</td>
</tr>
<tr>
<td>Scientific finding</td>
<td>Explicit disagreement or Mention of continuing disagreement in the scientific community</td>
<td>Scientific controversy</td>
</tr>
</tbody>
</table>
or endorse conference recommendations indicate either that the author has been influenced by the conference or (perhaps more typically) that the author wishes to use the conference to establish credibility for the author's previously or independently held opinions. In either case, such citations indicate that the conference had at least some effect in furthering the cited clinical recommendations.

The category for positive effects of research recommendations (labeled "research motivation") includes cases where the recommendation was cited to show the relationship of the citing research to clinical problems, as well as cases where the research recommendations of a conference provided a direct motivation for a published research study. In either case, the citing author invokes the conference in order to persuade readers (or editors) concerning the importance of research in the area.

For each sample conference, Table 11.9 shows which types of effects were found in the literature. Table 11.10 provides details on why each individual category was chosen for each conference. All conferences, except for the Pap smear conference, had some discernible effect on the literature. Indeed, all conferences except the Pap smear and estrogen conferences appear to have furthered belief in at least some of their clinical recommendations, although the effects often appear to be quite small. The absence of discernible effects for the Pap smear conference may reflect the panel's inability to reach a consensus on the primary issue before it—the appropriate frequency for Pap smears. The estrogen conference appeared to be overtaken by events. Most citing authors disagreed with the panel's judgment in two major areas: the value of estrogen in osteoporosis prevention and whether conjugated estrogens pose the same risk of endometrial cancer as unconjugated estrogens. This disagreement was based at least in part on articles published after the conference.

Four of the eight conferences we studied were followed by continuing controversy over their clinical recommendations. These controversies seem to arise out of two different "states of science." The first is exemplified by the Pap smear, the recommendation of a two-step procedure in the treatment of primary breast cancer, and the
Table 11.9
TYPES OF CONFERENCE EFFECTS FOUND IN THE LITERATURE

<table>
<thead>
<tr>
<th>Conference</th>
<th>Clinical Agreement</th>
<th>Clinical Controversy</th>
<th>Research Motivation</th>
<th>Research Controversy</th>
<th>Scientific Agreement</th>
<th>Scientific Controversy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antenatal diagnosis</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Steroid receptors</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Estrogen use</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Thrombolytic therapy</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Pap smear</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cesarean delivery</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CABs</td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

breadth of the appropriate indications for thrombolytic therapy. In each case, there was very little scientific evidence for or against any particular decision, but of course clinical decisions must be made anyway. In the absence of scientific information about outcome, the debate and recommendations seemed to revolve around other issues: cost in the case of Pap smear, the relative convenience of the surgeon and patient in the timing of breast cancer surgery, and the desirability of sticking with known procedures in the case of thrombolytic therapy.

An entirely different type of controversy over clinical recommendations is exemplified by the conference on postmenopausal estrogen therapy. In this case, the clinical controversy was tied to a scientific controversy. There was a great deal of information about the topic and the state of science appeared to be in a continual flux in the
Table 11.10

SUMMARY OF CONFERENCE EFFECTS

<table>
<thead>
<tr>
<th>Conference</th>
<th>Type of Effect</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antenatal diagnosis</td>
<td>Clinical agreement</td>
<td>One article referred to the CDC for guidelines on the use of amniocentesis; another agreed with the recommendation concerning the use of laboratory indexes to measure fetal maturity.</td>
</tr>
<tr>
<td>Antenatal diagnosis</td>
<td>Research motivation</td>
<td>Six research articles cited the conference as pointing out the lack of knowledge about the relationship between electronic fetal monitoring and adverse outcomes.</td>
</tr>
<tr>
<td>Antenatal diagnosis</td>
<td>Scientific agreement</td>
<td>Six articles cited the CDC as sources of information regarding predictors of fetal distress; another article agreed with conference assessment of ultrasound.</td>
</tr>
<tr>
<td>Antenatal diagnosis</td>
<td>Scientific controversy</td>
<td>One article objected to the CDC's assessment of a particular study on the risks of amniocentesis, although two other articles clearly agreed with the CDC on these risks.</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Clinical agreement</td>
<td>The &quot;treatment standard&quot; statement was cited favorably by many authors and used to place other treatments in perspective.</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Clinical controversy</td>
<td>The &quot;two step procedure&quot; was endorsed by two articles and rejected by three others, including a CDC panelist.</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Research controversy</td>
<td>Both agreement and disagreement with the NSABP protocol &quot;endorsed&quot; by the CDC were expressed in the citing literature.</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Scientific controversy</td>
<td>Several panelists used the occasion of the conference to indicate opposing views concerning tumor biology and their implications for breast conserving operations. One of these views (Fisher's) received many endorsements in the literature.</td>
</tr>
<tr>
<td>Steroid receptors</td>
<td>Clinical agreement</td>
<td>The recommendation regarding technique for laboratory assay of steroid receptors and the need for quality control were frequently cited, although we found no citation to the recommendation for routine assay.</td>
</tr>
<tr>
<td>Conference</td>
<td>Type of Effect</td>
<td>Explanation</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Steroid receptors</td>
<td>Research motivation</td>
<td>Appeared to engender research into the possible relationship between androgen receptors and cancer of the prostate and concerning analytic methods for steroid assay.</td>
</tr>
<tr>
<td>Steroid receptors</td>
<td>Scientific agreement</td>
<td>The conference was frequently cited to support the link between receptors and hormone therapy.</td>
</tr>
<tr>
<td>Estrogen use</td>
<td>Clinical controversy</td>
<td>The majority of citing authors disagreed with the CDC recommendation against long term estrogen replacement therapy.</td>
</tr>
<tr>
<td>Estrogen use</td>
<td>Research motivation</td>
<td>Two articles cite Weinstein's cost benefit model as a motivating factor in their research.</td>
</tr>
<tr>
<td>Estrogen use</td>
<td>Scientific agreement</td>
<td>The Weinstein cost benefit model was widely (but not unanimously) recognized as the right way to think about estrogen replacement therapy. Acceptance of this model may have speeded up rejection of the major CDC recommendation. In addition, certain facts were cited as established by the panel or speakers. In particular, that unconjugated estrogens cause endometrial cancer and that estrogen relieves vasomotor flushes.</td>
</tr>
<tr>
<td>Estrogen use</td>
<td>Scientific controversy</td>
<td>The rationale for the refusal of the conference to endorse long-term estrogen replacement therapy rested in large part on lack of information concerning: (1) whether ERT prevents osteoporosis, and (2) whether estrogen and progesterone therapy significantly lessens the endometrial cancer risk. Many writers found the evidence conclusive as regards one or both of these propositions. The evidence for both propositions increased following the conference, although there is little evidence that the CDC affected this research.</td>
</tr>
<tr>
<td>Thrombolytic therapy</td>
<td>Clinical agreement</td>
<td>The patient management guidelines for use with thrombolytic therapy were cited favorably. (In other words, &quot;if you are going to use it, follow the CDC guidelines&quot;).</td>
</tr>
</tbody>
</table>
Table 11.10--continued

<table>
<thead>
<tr>
<th>Conference</th>
<th>Type of Effect</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thrombolytic therapy</td>
<td>Clinical controversy</td>
<td>Many citers, especially review articles, did not accept the broadening of indications for use recommended by the CDC. These still regard heparin as the treatment of choice except in extremely delimited conditions. However, two clinical articles were favorably disposed toward the clinical recommendations.</td>
</tr>
<tr>
<td>Thrombolytic therapy</td>
<td>Research motivation</td>
<td>The conference appeared to lead to the publication of two clinical case studies and one basic research study.</td>
</tr>
<tr>
<td>Thrombolytic therapy</td>
<td>Scientific agreement</td>
<td>The CDC was cited regarding its discussion of the biochemistry of thrombolytic therapy. Parameters to be used during patient management, discussed in the CDC, were extensively cited.</td>
</tr>
<tr>
<td>Pap smear</td>
<td>Clinical controversy</td>
<td>The conference was reported in two commentaries that discussed the existence of a controversy over the appropriate frequency of Pap smears. The conference was seen in those articles and in four of the five citing articles as merely stating the range of the disagreement.</td>
</tr>
<tr>
<td>Cesarean delivery</td>
<td>Clinical agreement</td>
<td>Two review articles and three commentaries repeated at least some of the clinical recommendations.</td>
</tr>
<tr>
<td>Cesarean delivery</td>
<td>Research motivation</td>
<td>The CDC was cited in six research articles that drew on the CDC as a source of justification for the research.</td>
</tr>
<tr>
<td>Cesarean delivery</td>
<td>Scientific agreement</td>
<td>Epidemiological data provided at the conference were widely cited.</td>
</tr>
<tr>
<td>CABS</td>
<td>Clinical agreement</td>
<td>The CDC was cited in support of CABS as a major treatment advance; the recommendations for diagnostic workups were cited once.</td>
</tr>
<tr>
<td>CABS</td>
<td>Research motivation</td>
<td>The CDC was cited in articles on surgical techniques and on psychosocial aspects of patient management.</td>
</tr>
<tr>
<td>CABS</td>
<td>Scientific agreement</td>
<td>The conference was frequently cited as authority for improved survival in selected patients and relief of angina following CABS. It was also cited in reference to the frequency and cost of CABS.</td>
</tr>
</tbody>
</table>
years immediately following the conference. The reason that the conference did not produce lasting clinical consensus was that the panel did not know (and did not correctly guess) the outcome of ongoing studies.

During this same time, the state of our knowledge about the benefits of CABS was similarly rapidly evolving. However, the recommendations of the CABS conference were not the subject of controversy, possibly because they were less explicit.

The third and last of the major scientific controversies found in the literature also has the property that it is concerned with an evolving scientific database. We found much discussion in the literature of the correct model for the spread of breast cancer. One might question why we did not include the appropriate surgical technique for primary management of breast cancer as another area where controversy over clinical recommendations continued. The answer has to do with the nature of the citations. Almost all citing authors appeared to agree with the conference's two main conclusions in this area: (1) total mastectomy with axillary dissection should be the current treatment standard, and (2) further research is needed into all alternative modalities. This neat partition of the world into the realm for current practice and the realm for research was not made in the other cases, possibly because of the absence of a satisfactory treatment modality.

All the conferences except breast cancer and Pap smear (neither of which offered many scientific judgments) were cited for some of their scientific judgments or for facts brought out at the conference. These citations were sometimes of the nature that alternative sources could have been cited to make the same point. Thus, citation of the conference in these situations probably indicates that the citing authors respect the consensus process and invoke the conference because they consider it especially credible.

In addition to this "touchstone" role, a few conferences appear to have themselves greatly contributed to the state of science. Several papers that were prepared for the conference and published in the special supplements for CABS and steroid receptors were widely cited and apparently highly useful in ongoing research.
The Weinstein paper on the cost effectiveness of estrogen therapy appears to have played a special role in shaping the scientific and clinical debate concerning that therapy. Almost all clinical reviews evaluating the updated scientific information on the subject implicitly or explicitly adopt Weinstein's model. Although not repeating Weinstein's numerical calculations, these reviews often explicitly consider how the new scientific evidence would change the results. By clearly articulating the rationale for its clinical judgments, the estrogen CDC panel made a contribution that has been more important and lasting than its specific clinical recommendations. It is ironic that in so doing, the panel probably hastened the day when its clinical judgments would be rejected.

Several conferences appear to have inspired research (in addition to the reviews that were directly commissioned for the conference). For example, several studies were performed in apparent direct response to the research recommendations of the thrombolytic therapy conference and the steroid receptor conference. Moreover, many researchers have sought to interpret their data in the light of conference findings. Because our study tracks citations for only a short time after the conference and because a long lead time is often required to carry out and publish research, we expect that a longer followup period would result in the discovery of still more effects on research.

OTHER MEASURES OF INFLUENCE
Influence of Comparable Articles

To discern the influence of a particular consensus conference on the biomedical literature, we must also examine the influence of other papers that are available concurrently. This is especially important if we are to determine whether the CDC statement performs a unique function, such as addressing a topic in a new way or making the clinical implications of scientific findings explicit for the first time. If a statement's message is largely duplicated elsewhere in the literature, then assessing its influence requires answering a somewhat different set of questions, for example whether the statement reaches a different audience from those reached by other sources, or whether it carries greater scientific credibility.
Because of the importance of knowing something about the role played by other sources, we sought in our literature review to identify published papers and review articles for each topic that influenced scientific or clinical thinking during the period covered by our study. To determine which papers were influential, we need both qualitative and quantitative criteria. We began by examining the secondary literature to determine which papers (in addition to the primary articles associated with the conference) were cited most frequently or cited on important substantive points in the ongoing scientific communication about the topic. Papers identified as possibly influential were then examined further: We read them for content and used the ISI database to determine the frequency with which they were cited in the larger biomedical literature. In the end, our judgment of which papers were influential was a qualitative one, although it used available quantitative data.

Before describing our findings, we should note a limitation of our methodology, namely the possible selectivity introduced by our use of the secondary literature as a starting point. Because this literature comprised only those articles that cited the consensus conference primary literature, papers that were never cited in this secondary literature would be unlikely to come to our attention. We do not consider this limitation very serious, however, since our goal was not to identify all relevant papers, but only those that are influential, and because in all cases we used current review articles that presumably survey the field fairly thoroughly.

**Antenatal Diagnosis**

The conference on antenatal diagnosis dealt with three topic areas: (1) hereditary disease, (2) fetal maturity, and (3) fetal distress. Our search of the literature disclosed no other information source that covered all three of these topics. We did, however, find articles that deal individually with the three topics.

Gerbie and Elias (1980), in an edition of *Clinics in Obstetrics and Gynecology*, provide a detailed overview of the use of amniocentesis for diagnosis of genetic effects. They provide a set of indications which
is congruent with, if somewhat more detailed than, the conference recommendations. They conclude that amniocentesis is a safe and highly accurate procedure. This paper was cited once in 1981 and twice in 1983.

The only study we could locate that reviewed the antenatal assessment of fetal maturity was published in 1974 (Perkins). He discusses a large number of tests but concludes the most significant of these are L/S ratio and ultrasound. The L/S method and ultrasound were still being refined at the time of publication of the article, but the author rated them as the best. The recommendations are consistent with those of the panel. This review was cited once in 1980 and twice in 1982.

Banta and Thacker (1979) published a review of the costs and benefits of electronic fetal monitoring. After an extensive review they concluded that "the literature indicates little increased benefit from EFM compared to auscultation." They go on to suggest that if EFM has a benefit over auscultation, it is only for low birth weight infants, and that this benefit has yet to be tested in a randomized controlled trial. Pointing out that there are risks and costs associated with EFM, they conclude: "Until more evidence is available on the relative merits of auscultation and EFM, informed patient choice may be the best policy." This article had a considerable influence on the subsequent literature, with three citations in 1980, four in 1981, and five in 1982.

Treatment of Primary Breast Cancer

This conference resulted in the publication of two minority reports, one of which (Fisher, 1979) was cited much more heavily than either the consensus summary itself or a more detailed article by the panel chairman that explains the reasoning behind the panel's positions.

Steroid Receptors

Elwood Jensen's research played a major role in the development of receptor assays and their application in predicting clinical response. In 1981 he published an article in the journal Cancer outlining his work (Jensen, 1981). He reviewed the research on assays and recommended, as did the conference, that all primary tumors should be assayed to provide
a guide for the selection of therapy. He also outlined the development of an immunoassay for estrogen receptors.

**Estrogen Use in Postmenopausal Women**

Soon after the conference, several studies on estrogen use in postmenopausal therapy appeared that did not support the CDC's position. In particular, studies by Lindsay et al. (1980), Weiss et al. (1980), and Paganini-Hill et al. (1981) suggested that estrogens are effective in preventing osteoporosis. Many review articles endorsed this position and also apparently accepted the position that current data adequately justified the use of progesterone in combination with estrogen. By January 1983, the American Medical Association's Council on Scientific Affairs judged the evidence to be sound that estrogens prevent osteoporosis and that combined estrogen-progesterone therapy results in a lower incidence of endometrial cancer than estrogen alone.

**Thrombolytic Therapy in Thrombosis**

A clinically oriented article by Bell and Meek (1979), published in the *New England Journal of Medicine* only four months before the conference, presented guidelines that were very similar to those later set forth by the conference panel (on which Bell served). Not surprisingly, this article in a premier clinical journal was much more heavily cited than the conference summary, which appeared in four regional journals, the *British Medical Journal* and the *Annals of Internal Medicine*.

**Cervical Cancer Screening**

As discussed earlier, the principal issue faced by the consensus panel concerned the optimal interval between screenings. In February 1980, five months before the conference, the American Cancer Society announced its recommendation of a three year interval. Shortly thereafter (and in response to this recommendation) the American College of Obstetrics and Gynecology formed a task force that argued for the continuation of annual screening. The CDC panel took a position squarely in the middle of the road, recommending an interval of one to three years. Both the ACS and ACOG recommendations have exerted much more influence, although obviously in different directions.
Cesarean Delivery

The CDC on childbirth by cesarean delivery covered a large and complex subject. This is reflected in the fact that the conference produced the largest summary and was supported by a task force report that was over 550 pages long. Given the breadth of this topic, it is probably not surprising that we located only one reference that seemed to cover the same ground as the panel. This review article was published in *Clinics in Obstetrics and Gynecology* in December 1980 (Harley). This article reviews the history of cesarean delivery, indications for the procedure, and describes various surgical techniques. The author avoids making specific recommendations regarding breech, fetal distress, and dystocia, but rather says that defining appropriate indications "is difficult for the operation is carried out only after careful consideration of all the indications at that time for a particular patient." However, he does state that a trial of labor is warranted in women with a low transverse incision in institutions that have the facilities necessary for an emergency operation.

Coronary Artery Bypass Graft Surgery

In early 1982, Rahimtoola published a review of coronary bypass surgery for chronic angina (Rahimtoola, 1982). While both the consensus statement and Rahimtoola's article reviewed the risks and benefits of bypass surgery, Rahimtoola's article provided a much more detailed set of recommendations of appropriate indications for the procedure. He agreed with the panel's conclusion regarding prolonged survival for individuals with left main and triple vessel disease. He also included a set of functional status and laboratory test results that he felt defined appropriate patients. This article was widely cited and may have affected the medical profession's use of this procedure.

Influence on Textbooks

We examined two leading textbooks of internal medicine (Wyngaarden and Smith, 1982; Petersdorf et al., 1982) to determine whether their treatment of recent consensus topics corresponds to consensus conference findings and whether the conference statements are cited. We found that
the content of discussion in both texts is generally consistent with consensus conference recommendations but that the CDC statements themselves are not cited. Wyngaarden and Smith do not mention the CDC in their discussion of coronary artery disease. Petersdorf et al. do not mention the CDC in their chapter on angiography or their chapter on the management of chronic angina, but they do refer to Rahimtoola's review on the latter topic.

In their discussion of breast cancer, Petersdorf et al. mention the shift away from radical mastectomy and note the importance of estrogen receptors but do not mention either CDC. Similarly, Wyngaarden and Smith note that radical mastectomy is obsolete and state that receptor assay should be performed on all primary tumors. They cite neither of the relevant CDCs, but they do refer to the NCI consensus statement on mammography.

Regarding thrombolytic therapy, neither text cites the CDC (although Petersdorf et al. cite Bell and Meek's clinical guidelines published a few months before the conference), and neither accepts the panel's conclusion that thrombolytic therapy is the treatment of choice in a large number of cases. Wyngaarden and Smith (1982, p. 393) state, "In general, thrombolytic therapy is indicated only for pulmonary embolism massive enough to threaten life or to leave the patient with inadequate pulmonary reserve." Petersdorf et al. (1982, p. 1556), although considerably vaguer, also differ greatly from the consensus panel: Thrombolytic agents "do not replace anti-thrombic therapy."
Their discussion cites Bell and Meek's clinical guidelines published a few months before the consensus conference.

CONCLUSIONS

The summary statements for all eight conferences were published in journals with the potential to reach large and appropriate audiences. Publication of the background material on which the conference was based was more variable, with only three of the eight conferences publishing speaker presentations in journals or supplements.

The number of citing articles appears to increase with the amount of coverage received by the conference through primary articles, suggesting that increasing the number of direct "products" that emerge
from a conference may be an effective way to increase the amount of attention the conference receives. Controlling for research field does not alter this conclusion. The majority of citing articles were research articles. Several conferences received citations in a substantial number of review articles aimed at an audience of practicing physicians; however, in many of these cases, the citing author explicitly disagreed with some of the conference recommendations.

Six of the eight conferences (all but Pap smear and estrogen use) showed some--usually small--evidence that the conference had furthered belief in at least some of its clinical recommendations. Six of the eight conferences (all except treatment of primary breast cancer and Pap smear) were cited for some of their scientific judgments or for facts brought out at the conference. The circumstances in which many of these citations occur indicate respect for the consensus conference process in the scientific community. For four of the eight conferences, citations reveal continuing controversy after the conference over at least some of the clinical recommendations. For two other conferences, scientific controversies over the subject of clinical recommendations clearly continued in the literature, although these controversies were not discussed in terms of conference recommendations.

When conferences failed to produce lasting consensus on a controversy, the reason appears to be tied to the state of science in the particular area. This makes generalization from our studies problematic, since no single factor explains many cases. Rather, the results in different cases are consistent with various standard hypotheses used to explain delays in the diffusion of innovation, including the absence of compelling scientific fact, difficulties of interpreting scientific facts, the state of competing technologies, and the absence of compelling need.

One conclusion does arise strongly from both the quantitative and qualitative analyses, however. The majority of citations refer to specific points in a cited article in order to incorporate the point in the current research either as background or as part of an argument for a particular interpretation of findings. To serve this purpose, the point for which reference is made must have scientific credibility. Thus, anything that enhances the scientific credibility of a conference
recommendation will increase its citability, hence its potential influence on the literature. One way to increase both credibility and potential influence is to provide the rationale or scientific documentation for conference recommendations. Indeed, the three sample CDCs that published speaker presentations have by far the highest citation rates. In addition to publishing supplements recording the conference's scientific presentations, it might be worthwhile to encourage panel members or the chairman to write articles reviewing the rationale for conference decisions. Another way to illuminate the panel's reasoning would be to design questions that require the panel to provide both its recommendations and the important considerations that lead to those recommendations. This latter strategy, however, has the drawback that it may result in longer consensus statements. Separate packaging of recommendations and supporting documentation is probably preferable.

When the rationale for conference decisions is clearly spelled out (as in the estrogen therapy and steroid receptor conferences), the benefits of a conference may outlast the consensus on its clinical recommendations. In such cases, the conference statement may provide a framework within which even facts discovered after the conference may be viewed.

Finally, despite the short time during which we track citations to the conference, we find evidence that consensus conferences have inspired at least some research into areas that are important for clinical decisions and that they have caused other researchers to interpret their data in the conference framework.
REFERENCES


