

*Annu. Rev. Publ. Health. 1991. 12:41-65*  
Copyright © 1991 by Annual Reviews Inc. All rights reserved

## WORDS WITHOUT ACTION? THE PRODUCTION, DISSEMINATION, AND IMPACT OF CONSENSUS RECOMMENDATIONS

*Jonathan Lomas*

Centre for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Biostatistics, Health Sciences Centre, McMaster University, Hamilton, Ontario, L8N 3Z5, Canada

**KEY WORDS:** practice guidelines, physician behavior, group judgment, technology assessment, evaluation studies

---

"Consensus means that lots of people say collectively what nobody believes individually"

Abba Eban

### INTRODUCTION

The above quote from Israeli statesman Abba Eban voices a general concern about the increasing use of consensus processes as an imprimatur for certain practice patterns in medicine. Some critics fear that the implications of the consensus process will discourage physician autonomy or innovation. They point to the need to "protect the individual choices of each physician from the potential tyrannical domination of consensus and allow the process of development of new knowledge to continue" (61, p. 1077). Other critics claim that the methods used may overinterpret the available data and lead to conclusions based on "faith or zeal or alarm" (2, p. 1086). Finally, there are critics who, if correct, can assuage these fears, because they do not believe that consensus processes effect behavior change at all, but rather provide "primarily a dialogue among researchers . . . not a guide to action" (34, p. 2740).

Nevertheless, consensus processes as a means of information transfer are here to stay and will become even more popular. A recent *Directory on Technology Assessment* listed nearly 60 organizations in the United States with formal programs that are establishing hundreds of recommendations for practice. Nearly all of these organizations use some form of consensus in their deliberations (32). Formal consensus development programs now exist in Canada, Britain, Sweden, Norway, Finland, Denmark, Holland, and France (32). The programs use terms that differ across a spectrum of "consensus development conferences," "task force reports," "appropriateness ratings," "practice parameters or guidelines," or "technology assessment reports." However, they generally use varying degrees of formality to produce accessible and readily understandable consensus recommendations that summarize the implications of existing research evidence.

The need for such summaries is manifold: the burgeoning biomedical and social science literature, which increases awareness of the inadequacy of traditional journal articles as sources for direct adoption decisions (35, 93); evidence of uncertainty from medical practice variations on appropriate clinical policies (3); awareness that a significant proportion of care is inappropriately provided (10, 88); a shift from providing those services with expected benefit to only those with actual benefit (21); demands by third-party funding and quality monitoring agents for guidance on and succinct recommendations about appropriate practice (81); and pressure from the public for increased input to medical and technological decision-making (4).

The variety of catalysts produces variety among the specific aims, conduct, and output from each of the consensus processes. The target audience may be as narrow as a particular kind of practitioner or researcher (9, 39); as eclectic as politicians, administrators, clinicians, and planners (13, 86); or, as in Denmark, as diffuse as "public participation . . . for ensuring the democratic influence on decisions about medical technology" (4, p. 308).

The most common target is the practicing clinician, and the most common aim is to influence the clinician to improve the quality of care. Sometimes, the influence is indirect; it comes through funding or regulatory bodies. In this review, I focus largely on this specific purpose and target of consensus processes.

In the first section, I review some of the numerous methods for consensus production and the controversies surrounding them. I observe the lack of any recognized standards by which to judge the validity of the various approaches. One intent of the review is to begin assembling such a set of standards. This task can only be started here, however, because research comparing the effectiveness of different group judgment processes on "how best to put medical technologies to use, is not commensurate with the [research] effort and care devoted to developing these technologies" (37).

In the second section, I review the dissemination and impact of consensus by focusing on the apparently false assumption that dissemination is necessary and sufficient for behavior change. I review the methods used for dissemination and appraise the 19 studies identified in the literature that evaluate the impact of such dissemination on behavior. This appraisal yields a pessimistic conclusion—in most cases the words do not translate into action. I found some hope, however, in reviews of the impact of consensus on cognitive, rather than behavioral, outcomes, and the potential for combining the output of consensus with more active strategies for implementing changes in clinical practice.

I conclude with suggestions from recent work on how to improve the impact of the output. I also suggest a tentative set of standards by which to judge the validity of different consensus production methods.

I derived the materials for the review from a combination of sources:

1. Computerized searches of the US National Library of Medicine data base (MEDLINE) and the Educational Resources Information Clearinghouse (ERIC) for 1980 to the present, which use the search terms "consensus development," "guidelines," "standards," "official policy," "technology assessment," "evaluation studies," "epidemiologic methods," and "research";
2. Bibliographies on consensus methodology (26, 32a) and practice guidelines (75);
3. Citations in the articles retrieved;
4. Citations provided by colleagues; and
5. Personal files accumulated over the past eight years.

## THE PRODUCTION OF CONSENSUS

### *A Framework for Evaluating Consensus Production*

Ironically, the recognition standards routinely required of research studies are rarely found in the consensus reports that make the synthesis of such studies popularly available. The systematic and explicit description of the methods used is often partially or wholly missing. Thus, many consensus reports would fail the tests of replicability and defensibility: Based on the information given, could the methods in the exercise be replicated? Is the rationale for methodological choices provided?

The method and/or rationale for at least the following choices are desirable:

1. The topic selected;
2. The membership of the consensus group;
3. The nature and extent of background preparation;

4. The inclusion/exclusion criteria for information inputs;
5. The type of group process and definition of consensus;
6. The criteria for qualification as a recommendation; and
7. The preparation process and format of the report.

By making these choices and rationales explicit, the consensus group moves from "simply applying their intuitions and stating their beliefs, to reasoning through a problem step by step, and justifying the conclusions" (21).

There are some existing consensus programs that come close to meeting these rigorous standards of explication and justification, although some groups argue with the particular choices made by the programs (9, 14, 91). There are many other consensus exercises, however, that fail to make their choices explicit. These are often the "one-off" variety conducted by a disease-specific association or specialty group. Perhaps the most notable of these programs is the Consensus Development Conference (CDC) approach of the National Institutes of Health (NIH). This program has come in for much scrutiny and criticism, even from those involved in its organization (39).

At the same time, the NIH CDC program has been subject to more systematic evaluations, and consequent modification, than any of its competitors. Thus, it has spawned much valuable research information on the validity of different approaches to producing consensus (41, 46, 59, 99). Similarly, derivative CDC programs have also been evaluated in other countries, such as Sweden (13, 43), Holland (15, 89), and Canada (54, 55). There have even been cross-cultural comparisons of the processes and their outcomes (11, 79).

In the following sections, I use the results of these and other evaluations to review the explicit methodological and "political" choices needed for a consensus exercise to translate biomedical research into practical advice for clinical audiences.

### *Selecting a Topic*

The topic can be either disease-based, e.g. prevention and treatment of breast cancer, or procedure-based, e.g. mammography (75). In the former, the focus is largely on the appropriate use of all alternatives for alleviating a particular burden of ill health. In the latter, the focus is on the particular indications for which an intervention should be used: it is generally termed technology assessment. Each focus brings particular advantages and disadvantages.

With disease-based topics, it is generally easier to maintain the focus of the exercise on the objective of improved patient or population health. Nevertheless, the task is more complex. It necessitates evaluation of different technologies for the same disease and questions which criteria are most appropriate for such comparisons. For instance, when attempting to reduce the morbidity and mortality of breast cancer, how do you weigh cost, con-

venience, effectiveness, labeling, and preference (patients' and providers') in comparing breast self-examination with mammography and dietary and reproductive behavior changes with lumpectomy?

With procedure-based topics, the task may be simplified. However, considerations of technical capabilities may override the criterion of actual health benefits. For example, should the threshold for routine screening for breast cancer be lowered to an age group at lower risk just because mammography now uses reduced levels of radiation? Furthermore, timing is an acute problem in technology assessments. If a new technology is appraised too soon there will not be an adequate base of scientific knowledge; if appraised too late, the technology will already have diffused across the system (25). Many persons argued that the 1984 NIH CDC on lowering blood cholesterol to prevent heart disease was undertaken too early; subsequent research made their recommendations highly suspect (2, 49, 60). In contrast, the 1979 CDC on surgery for primary breast cancer was clearly too late; the principal recommended change had already taken place (46).

Insofar as the target audience is the practicing clinician, diagnostic and surgical specialties may find procedure-based topics most relevant. Medical and public health professionals may have greater interest in disease- (or at least health problem-) based topics.

Even after the topic has been decided upon and justified, the topic area has to be specified. Many programs outline general criteria to assist them in setting priorities across competing topics. These usually include at least the following three general principles: importance (in terms of frequency, burden of morbidity/mortality, or resource-consumption grounds); reasonable base of existing scientific knowledge on effectiveness of intervention or alternatives; and resolvable on the basis of more than personal opinion and values (8, 69, 91). Although these are not unreasonable as criteria for topic selection, their focus is almost exclusively on the state of science in the potential topic area, not on the state of practice. This omission has been noted in the context of the potential objective to alter practice patterns (4, 46, 49a, 52, 74). An appraisal of existing practice provides information on both the extent (or even existence) of a problem and the nature of any changes that might be indicated. Such information is important to any decision on the need for a consensus on a particular topic—either because variations in practice demonstrate uncertainty regarding what is appropriate or because practice is uniformly not congruent with the message from current research evidence (52).

### *Consensus Group Membership*

There are at least three distinct types of consensus group membership. Coronary artery bypass surgery is an interesting example because it has been subjected to each type. First, the RAND Corporation approach (and a

Canadian counterpart) used a panel of clinical experts on the area under consideration—cardiovascular surgeons and related specialists (9, 67). Second, a NIH CDC used a panel of scientific experts in both the clinical and related aspects of the topic—clinicians plus lawyers, epidemiologists, economists, and “expert consumers” (76). Third, a nonexpert or independent panel in the UK King’s Fund conference included administrators, laymen, and providers, from a variety of specialty and disciplinary backgrounds (86).

These three panel types—clinical experts, scientific experts, and nonexperts—reflect the chosen focus and target audience of the consensus exercise. With clinical expert panels, the focus is almost entirely on the safety and effectiveness issues and on the target audience of the specialty physicians actually engaged in providing the care. Scientific expert panels have a broader mandate. They consider such additional issues as ethics, economics, or future research needs and they target an expanded audience of clinicians, scientists, and administrators. The nonexpert panels are less concerned with resolution of conflict and more with “broadening the debate among a wide range of professionals in health care and with the public about medical technologies” (86, p. 713).

The panel type is also influenced by the sponsor of the exercise. Medical specialty societies tend toward the clinical expert panel. Research council and university-based organizers favor the scientific expert panel. Public foundations or directly government-sponsored panels are more likely to be nonexpert.

Given an objective of influencing practitioners in the area under study, one concern in panel membership is that it be credible to this target audience. The difficulty of satisfying all potential audiences is illustrated by the results of one survey performed in conjunction with the above-mentioned King’s Fund conference on coronary artery surgery. When questioned whether the panel and the proceedings were biased too strongly by the cardiac medical specialties, only 24% of the cardiac specialists surveyed replied yes, compared with 88% of community medicine specialists. This difference may have reflected the fact that those with a public health perspective are more favorably disposed to disease-based rather than procedure-based topics. Indeed, the community medicine specialists generally felt that the topic “was not presented in the whole context of the prevention and treatment of coronary heart disease” (86, p. 714). Nevertheless, the target audience for the actual performance of coronary surgery—cardiac specialists—was appropriately convinced by the credibility of the panel. Again, the attitude of the intended audience may be the most appropriate criterion to judge credibility. Had the topic been the prevention and treatment of coronary heart disease, then a panel and process less focused on cardiac specialists and more on public

health personnel would likely have been more appropriate and credible to the intended audience.

Many consensus panels concentrate membership among academics, rather than community-based practitioners. Although this approach satisfies the apparent role of the consensus process in synthesizing research information, it can conflict with the credibility to a community practitioner audience. In other work, researchers found significant resistance among community practitioners to "ivory tower" medicine as espoused by academic consensus panels (35, 56). There is no obvious balance between these potentially competing demands on consensus group membership. However, if one purpose is for community practitioners to identify with and find credible the recommendations, it is advisable to at least ensure visible representation of their viewpoint.

Consensus group membership inherently has tension between the panelists' appeal to the intended audience and the necessary skills to adequately consider relevant viewpoints and appraise scientific evidence. However, the need to include at least an epidemiologist to bring both public health perspective and methodological skills to the process is increasingly recognized (46). Sometimes, an economist is also needed to evaluate opportunity costs and resource allocation matters (13, 86).

An ongoing debate surrounding the NIH CDC program concerns the panelists' ability to meet the "science court" and "judicial jury" requirement of not having strong, preexisting views on the topic under consideration (65). This is more of a concern for the expert than the nonexpert panel, given the likelihood that both expertise and opinions flow from involvement. The social psychology literature warns of the potential bias (57). Other evidence suggests that when methodologic quality is stressed as the criterion for decision-making, such preexisting views can be appropriately altered by the consensus process (56). Some groups have argued that at least the panel chairperson should be neutral toward the topic (42).

Overall, prior consensus exercises have not been explicit about the criteria used to select topics and/or members of the consensus group. Any existing criteria often have not been carefully related to the objectives of the exercise and the target audience. Wortman and his colleagues, after evaluating one consensus program, pointed out that the absence of clear criteria and procedures leads to "the potential for selection bias in the choice of conference topics, questions, and participants [which] poses a set of related problems that can undermine the credibility of the CD [consensus development] process" (99, p. 490). They point out that formalization of these selection processes can reduce the problems. They also suggest using the Delphi method (17) with relevant medical schools, researchers, and associations to "rapidly pro-

duce a list of questions, panelists, speakers, and relevant research literature in two or three rounds of mailed questionnaires" (99, p. 491). Whether such a formal process is used or not, it is advisable to describe the criteria and procedures used in selecting topics and participants, if only to clarify that they relate directly to the objectives of the consensus exercise.

### *Background Preparation*

The broad areas of potential preparation for the panel are the state of science for the topic, the state of practice in the area, and the ground rules for operation of the group process.

Some consensus processes have provided none of this background; they prefer to rely on the existing knowledge of a clinical expert panel (e.g. 22, 27, 44). Nevertheless, the majority provide at least a bibliography of relevant literature for the state of science, if not an actual synthesis or summary. The most comprehensive exercises favor organization of a synthesis around the methodological quality of the various studies of interest (e.g. 9, 14, 71, 91). This approach has the advantage of orienting the panel away from clinical opinion and toward methodologically sound evidence, when it is available, as the adjudicator of controversy.

One former director of the NIH CDCs described the importance of the state of the science background preparation by bemoaning its sporadic availability for some 30 previous topics covered by the program:

On the occasions when such a synthesis was prepared by the staff and accepted by the panel, evidence was well integrated into both the deliberations and the consensus statement . . . . When a data synthesis was unavailable or was not used . . . the difficulty of coping was exacerbated. Probably as a result, some consensus statements show evidence of influence by panelists' assertions of common sense or knowledge of acceptable practices, without having been explicitly stated" (40, p. 3039).

I have already described the infrequent use but high value of background input on the state of practice. In the context of panel preparation, the value of this information is its ability to correct any "imbalance between the nature of the panel's task and the information it has available for accomplishing the task. . . . The consensus panel is supposed to translate biomedical research findings into clinically meaningful recommendations. To do its job well, the panel should be well-informed about both the current state of science and the current state of practice" (46, p. 242). If such background preparation had been included in many previous consensus processes, it may have saved time and effort by preventing mere "codification" recommendations that reflected practice patterns already diffused and in place (e.g. 36, 49a).

In countries such as Canada, where accessible data bases on practice patterns are routinely collected, background preparation can be completed easily (12). In countries such as the United States, where access to these data



is more problematic, researchers can perform specific surveys, or even establish a routine collection, using a panel of "Neilsen hospitals," similar to the monitoring of families for television viewing patterns (46).

### *Information Inputs*

In addition to background reviews before deliberations, there is the question of the source and nature of information inputs during deliberations. The source and, to some extent, the nature interact with how cloistered or how public is the conduct of the consensus process. At one end of the continuum are the most cloistered of the clinical expert panels; they operate behind closed doors and consider little besides the published literature and their own views on safety and effectiveness (e.g. 9, 14). At the other end are the highly public forums, common in Scandinavian countries, that solicit both information and views from various sources—the general public, administrators, politicians, patients, researchers, and care-givers—on numerous aspects of the topic (4, 90).

These differences reflect, in part, the extent to which some processes restrict deliberations to the results of experimental research, whereas others wish to synthesize values and other "normative" components with the "objective research." Some researchers argue that the former is not a valid approach because even the research information is not value-free: The choice of a more public process that takes "part of its format from a societal instrument, the jury/court, which deals with moral values, means the format provides a strong impetus to evaluate health technology from societal viewpoints rather than from that of scientific evidence" (90, p. 67).<sup>1</sup>

The choice of narrow inclusion criteria for information inputs by the cloistered exercises can be seen as the equivalent of the researcher's *ceteris paribus*—by excluding the overtly normative issues (ethics, economics, patient preferences) and their sponsoring sources (social scientists, economists, the public), control is better maintained over the outcome of the safety and effectiveness variables of interest. The exercise is more easily perceived as an objective one, even if it is less relevant to the world in which the decisions are actually being made.

The degree of relevance to public, or at least nonclinician, viewpoints is nevertheless central to many of the current consensus programs. Not coincidentally, the most publicly oriented of all consensus programs, the Danish process, chose early detection of breast cancer in 1983 for its first exercise because "this was primarily a public-interest issue: Potentially it had a preventive impact, it certainly dealt with health care costs, but there was little

<sup>1</sup>Some of the difficulties encountered by the NIH CDC program might be attributable to ambiguity on this issue. They restrict consideration to matters of safety and efficacy, but also use an open format that allows input from numerous sources, including the general public (99).

basic scientific information available to answer the questions it raised, since controlled clinical trials had not yet been finalized at the time of the meeting" (90, p. 71).

For such public health topics, with ethical issues around screening and allocation, the importance of obtaining information inputs on values to integrate with clinical science may be greater than in some of the more restricted clinical areas where the demonstrations of benefit and cost-benefit may be relatively black and white. In particular, the central nature of patient preferences in determining many public health treatment or screening decisions is not easily considered by methods that ignore the importance of values (21). A few consensus exercises have, largely unsuccessfully, tried to incorporate this element using formal medical decision-making models (48, 73). Similar attempts to incorporate economic evaluations have met with somewhat more success (92).

Many researchers have expressed the "need for a more general technology to allow for the assimilation of a much wider variety of evidentiary material, including expert opinions, biological theories and supporting laboratory data, and evaluations of component pieces of the overall practice under review" (50). Some methodologies aimed at an objective assimilation of disparate evidence are being developed. There are, for instance, variations on meta-analysis (83), Bayesian meta-modeling (50), and even a science of designing practice policies (21). With the increasing popularity of evidence synthesis, we can expect more innovation.

The actual format of the information inputs is not crucial. However, many surveys of panelists, speakers, and audiences suggest that the use of "witnesses" giving oral presentations and engaging in debate of "testimony," is preferred to impersonal written submissions, at least from the perspective of the actual participants (13, 43, 86). One survey found that those groups exposed to such presentations and debate were the most likely to have taken actions based on the consensus report (43).

### *Type of Group Judgment Process*

As described above, many formal methodologies are being developed for the kind of synthesis and integration demanded in consensus processes. These methodologies, however, are largely related to marshaling the appropriate information inputs into a manageable format for consideration by the consensus group. Setting the ground rules for interaction among participants and for definition of a consensus requires a separate process. Many excellent reviews of group judgment processes already exist (e.g. 18, 26, 30, 37, 64). There also has been much written about the science court approach (87) and the NIH CDC variant, which involves elements of judicial process, scientific conference, and the town hall meeting (65). Most, if not all, of the literature

addresses three components: how to generate a common focus in a group, what criteria to use to resolve controversy, and how to define consensus.

The provision of comprehensive background materials and the opportunity for group members to have input can help generate and define a common focus. The two primary methods, however, have been the formulation of a few specific questions, whose answers form the actual consensus (70), and the development of a comprehensive set of scenarios for rating and/or discussing appropriate intervention strategies by the group (10, 67, 72).

Formulating questions is most amenable to processes that have face-to-face contact among group members. Although the practice has not been validated, no more than four to six questions should generally be posed to a group, and the questions should lead to concrete and unambiguous answers (45).

Constructing scenarios for rating is most valuable when a procedure-based topic is under consideration and when the main concerns are safety and effectiveness. Disease-based topics make it much more difficult to comprehensively describe all potential clinical situations. It is even more difficult to integrate the economic, ethical, or other social factors into the scenarios. Some researchers construct representative, rather than comprehensive, scenarios to overcome this difficulty. This approach, however, counteracts the potential advantage of being done by mail—in a manner similar to Delphi techniques (17)—because the group must come together to “fill the gaps” left by the representative scenarios. This hybrid method can, nevertheless, be valuable. The scenarios should generate a common focus and identify the areas of agreement and disagreement between panelists before the meeting. The focus can then be on initial disagreements in a face-to-face meeting that uses time more efficiently to address contentious areas in the form of preformulated questions (56).

Defining the criteria for resolution of conflict is perhaps the most difficult part of any group process. Most exercises have not made the criteria for resolution explicit, although many have pointed to the responsibility of the chairperson to ensure smooth operation of this aspect of group process (37, 42, 99). Many panels have an implicit hierarchy of criteria that place controlled trial research above other, less methodologically stringent research evidence, which in turn is valued more highly than clinical experience or personal opinion. Evidence and opinion criteria are less in competition and more complementary: When evidence is available, it is the preeminent determinant; when it is not, the credibility and experience of the various proponents of differing opinions will determine resolution. The ability of research evidence, rather than personal opinion, to better resolve disagreements among group members has been demonstrated in the literature (56, 63). In one exercise, when the above hierarchy of criteria was made explicit, disagreements among panelists before deliberations were resolved by the

consensus process for 71% of situations when research evidence existed, but for only 24% when no research was available (56).

Related to the criteria for resolving disagreement is the mechanism used to define consensus, which may be quantitative or qualitative. With the quantitative mechanism, there is less concern over being explicit about the resolving criteria (and less discretion is left in the hands of the group's chairperson). However, the basis for the consensus may be left unclear. Many investigators have explored the properties of different quantitative definitions of consensus in the context of the ratings done on the scenarios (9, 67). No clear guidelines independent of the exercise's purpose have emerged, other than the relatively obvious conclusion that the stricter the criteria the more difficult it is to arrive at consensus. One author has arbitrarily proposed that "if agreement from at least two thirds of the participants can be reached . . . consensus is established" (26).

Although there has been much written to describe alternative group judgment methods, surprisingly little research has been performed to relate chosen options to the outcome. Hence, there is little information to assist prospective convenors of consensus processes in choosing among the alternative approaches. The field needs more systematic assessments of the various methods when they are used under different consensus circumstances. At present, "although we do quite well at assembling experts, we often provide them with inadequate, largely untested means for drawing upon their expertise and for organizing and weighing the evidence" (37).

#### *Criteria for Qualification as a Recommendation*

After an exhaustive evaluation of the NIH CDC program, Kanouse et al concluded that "the purposes of the program are better served if the panel approaches its task by asking, 'What meaningful guidance can we give to clinicians based on the current scientific evidence?' rather than 'What definitive recommendations will the biomedical literature support?'" (46). Herein lies probably the single most contentious issue in the debates surrounding choice of consensus approach: Is the purpose of recommendations from consensus processes to establish the best possible guidance for clinical care despite imperfect or incomplete evidence, or is it to promulgate science based only on watertight conclusions derived from methodologically incontestable studies?

The latter approach places great reliance on the randomized controlled trial (RCT). The RCT has been favored by both the Canadian and the American Task Forces on the Periodic Health Examination, and many epidemiologically based consensus conferences restricted to the safety and effectiveness of defined clinical interventions (e.g. 51, 66, 88). When the purpose of the

exercise is entirely science-related, e.g. establishing future research requirements, the stringency of such a criterion is appropriate.

For the development of practice guidance, however, there are at least three problems with strict reliance on randomized controlled trial evidence as the only justification for a recommendation. First, such reliance significantly limits the areas of clinical practice in which consensus recommendations can be made. For instance, studies of causation in occupational health would be unethical when using a RCT. For many preventive medicine and public health issues, the length of time between intervention and potential outcome is so long as to make RCTs infeasible (8).

Second, RCTs alone are largely unable to consider economic, ethical, or other social considerations. Thus, a highly effective, but extremely expensive, drug may not warrant recommendation given the opportunity cost of its use. This is precisely the issue in the current debate of tissue plasminogen activator versus streptokinase for the treatment of thrombolysis.

Third, negative recommendations are sometimes appropriate because management options or technologies are being diffused too rapidly or too widely. For example, "the use of cesarean section is *not* indicated for women with an uncomplicated previous cesarean section" (71). Indeed, a major impetus for NIH CDC program was Congress' perception that there was widespread use of many technologies "without sufficient information about their health benefits, clinical risks, cost effectiveness, and societal side effects" (78). In this case, failure to provide a negative recommendation because of the absence of RCT evidence is placing the onus of proof on those trying to prevent unproven interventions from diffusing into practice.

Nevertheless, for strictly clinical effectiveness issues, reliance on methodologies other than the RCT, where it is feasible, can be severely misleading. Recalled experience of clinicians is a notoriously unreliable source of accurate effectiveness estimates. Such estimates tend to be overly optimistic for reasons such as recall bias, regression toward the mean, and placebo effects (20, 82). Even if the purpose of the consensus exercise is to provide guidance for practice, and not the strict promulgation of science, accuracy still requires a distinction between those recommendations supported by RCTs, and those supported by evidence of less certainty.

Many systems have been proposed for such grading of recommendations (14, 21, 82). One of the simplest systems reserves the term "recommendation" for a high level of certainty because of support from methodologically sound studies. The system uses some lesser term, such as "guideline" or "suggestion," for situations supported by less certain forms of evidence (54). The particular grading system is less important than the explicit recognition that different grades of recommendation do exist, based on the methodologic

quality of the supporting evidence (40). Thus, consensus groups need not be prevented from producing conclusions that rely largely upon their experience or their interpretation of ethical and social considerations. They should, however, be clearly differentiated as "informed opinions," rather than given the imprimatur of "proven science."

### *Report Preparation and Format*

Who prepares a report can vary from a single professional writer, through planning secretariat, to panel chairperson, and on to joint efforts of the entire consensus group. Many researchers have stressed the importance of a readily understandable and accessible report. They recommend that a professional writer should at least be involved with the final draft (39, 46, 99).

The time taken over report preparation has been the subject of much greater debate. Some panels are put under time pressure to increase their motivations for consensus, use the limited time of experts most efficiently, and capitalize on the attention generated by the group process of evidence consideration (if it has been public). This pressure has resulted in one strategy that has the consensus group drafting and finalizing the report over a period of 24 to 48 hours (39). The approach has been highly criticized because it leads to "lowest common denominator" recommendations on particularly difficult and controversial issues (99), or "is bound to lead to hurried conclusions" (68, p. 1088).

In contrast to this approach, lengthy iterative processes have been used that allow careful consideration of complex issues. It has been argued, however, that procedurally this is cumbersome, time-consuming, and expensive and fails to make maximum and efficient use of all the expert skills convened at one time for the consensus group.

These approaches are not, however, mutually exclusive. They can be combined to obtain the best of both and maximize the amount of input to the final product. After preparation of an initial draft under time pressure, circulation and feedback are undertaken; final drafting follows some weeks or months later (52, 68, 74).

The format of high quality and/or particularly influential reports has been the subject of two evaluations (45, 98). Conclusions from both are similar and, not surprisingly, suggest formats for consensus statements that "1. recommend concrete specific actions; 2. differentiate patients into subclasses when appropriate; and 3. offer didactic advice to the clinician on precise techniques that should be used" (46, p. 26).

These suggestions presume that the objective of the consensus exercise is to give guidance to clinical practice. Thus, they reflect the finding from other surveys that physicians desire "easy-to-read, short, authoritative articles giving the best medical judgment on the value and limitations of new scientific

works" (16). In one recent survey of physicians' information preferences, only about one third wished to receive information in "complete form (with evidence)"; the others preferred it in "summary form (with references)." In the same survey, almost 100% had a preference for clinically rather than research-oriented information. Nevertheless, about 60% believed that reports in professional medical journals were very important when first hearing about or deciding to use a new procedure (46).

A credible and potentially influential format clearly demonstrates that a scholarly process has been carefully followed, but presents the guidance in a nonscholarly and easy-to-read manner. It should also provide references (74) and estimates of expected outcomes (21, 37). Finally, the future validation and development of consensus methodology can best be advanced by including explicit descriptions of the procedures and the choices made for each of the seven methodologic areas described in this section on the production of consensus. Without such descriptions, it is difficult to accurately judge the scholarly credibility of the consensus.

## THE DISSEMINATION AND IMPACT OF CONSENSUS

### *Models of Diffusion*

Words, whether credible or not, rarely flow automatically into action. Recommendations must be disseminated in ways that provide incentives for such action. Or, those to whom the words are directed must be remarkably receptive to, and already prepared to act on, the message.

Unfortunately, traditional diffusion models, which appear to have been the guide for the dissemination strategies of most previous consensus exercises, "have perhaps placed too much faith in the model of the rational, information-seeking, and probabilistic practitioner, expecting the mere availability of new information to lead to changes in his or her clinical policies" (56, p. 90). This model of the practitioner has been called into question by many recent reviews that point out that research information (synthesized or otherwise) is only one of many determinants of the policies adopted by practitioners.

This more recent work stresses the interaction between characteristics of the receiver, the source, the message and, the channel of the information. This work implies that publication without regard for such interactions is a very weak form of dissemination (5, 30). Furthermore, the process of behavior change requires a set of stages starting with predisposing or priming activities to trigger consideration of change, followed by enabling strategies to motivate and facilitate change, and concluding with reinforcing activities to sustain the change (28, 29, 33, 35, 47). Thus, dissemination should be directed not only at increasing awareness but also at influencing attitudes, knowledge, and, finally, behavior.

### *Dissemination Strategies*

This more recent conception of practitioner behavior change has not, however, been reflected well in the dissemination strategies of most consensus exercises. The overwhelming strategy has been mere publication, sometimes distributed in booklets, sometimes in specialty journals, but most often in general medical journals. Some of the more high-profile programs also rely on the media for short-term dissemination via press conferences. The quality of this reporting, when assessed, has been judged as largely factual and balanced (94).

Multiple sources are identified by practitioners for where they became aware of a consensus statement. The three most frequently cited potential sources were professional medical journals (50%), printed materials such as booklets (30%), and the popular press (25%) (46, 55).

However, awareness of consensus statements among the entire relevant population of practitioners varies considerably. It has been as low as 20% with cardiac surgeons (46) and as high as 90–95% with obstetricians (55) and Swedish physicians (43). Usually, awareness is in the 30–60% range (1, 36, 46). For any specific statement, specialists are more likely to be aware of the recommendations than are general or family practitioners (43, 46).

Direct mailing of the statement to the relevant practitioner population does increase awareness, but even then awareness does not seem to exceed 40% (41, 55). This level of awareness can, however, be significantly increased by making the materials visually attractive and/or “staging” their delivery by dividing them into bite-sized chunks of information (6, 24). However, even this increased awareness may not be reflected in a consequent change in behavior.

### *The Impact of Dissemination*

Methodologically, it is not easy to definitively evaluate the impact of consensus exercises. They are widely disseminated, which makes control groups impossible. It is difficult to insulate an experimentally defined portion of the relevant practitioners from exposure to the consensus. One possibility is to evaluate impact separately from those practitioners aware and those not aware of a particular consensus. Such evaluations have produced conflicting results. Sometimes those aware of the consensus were more likely to have made recent changes that conform to the recommendations (43, 46), but sometimes not (36).

The difficulty with this approach, however, is that the assessment is not representative of the target practitioner population. This is because the likelihood of awareness is correlated with other practitioner variables, such as degree of participation in continuing medical education or journal reading



habits (46), which might equally well explain their greater propensity to change. Furthermore, the low levels of awareness for some consensus recommendations makes such an analysis of little relevance.

Thus, evaluations of impact in the entire relevant population require the use of representative chart reviews, analysis of administrative data, or surveys of self-reported behavior. The strongest conclusions about impact can be drawn from those studies that use actual practice data, rather than self-report, and take measurements both before and after the consensus (preferably with a time series to account for preexisting trends in behavior). Weaker, cross-sectional designs can provide impact information only if they require self-reported recall of prior behavior—an unreliable source of measurement. Such cross-sectional designs can, however, provide an estimate of the conformity of practice with the consensus recommendations at a point in time. If the point in time is subsequent to dissemination of a consensus, then we can measure how far the recommendations are falling short in achieving their goal, even if the degree of conformity is unrelated to impacts from the consensus.

Table 1 presents evaluations since 1980 that provide information on either the impact of consensus recommendations on practice behavior or the percent conformity with recommendations. Studies were identified from the sources described earlier. They were included if they measured impact on physician behavior, defined the consensus exercise from which recommendations were drawn, and provided enough description to adequately define the methods used. Nineteen studies met these criteria; they are divided according to whether they used actual practice data or self-reports of behavior.<sup>2</sup>

In the ten instances where impact was measured using actual practice data, six found no impact, two found minor impact, and two found major impact. Interestingly, three of the four studies showing any impact were from Europe [Fowkes & Roberts: UK (28); van Everdingen et al: Holland (88, 89)]. All six of the studies finding no impact were from North America. In the eight instances where an estimate of percent conformity with recommendations was possible, it was less than two thirds of potential for all but one. The one instance was, again, one of the Dutch studies. It was performed in a highly circumscribed practice area (reporting of Breslow thickness for the diagnosis of cutaneous melanoma) and was at a preconsensus conformity level of 83%.

The self-report studies are a less reliable indicator of impact. In the one study where both types of data were available, the percent conformity from the actual practice measure was less than half that of the self-reported estimate (27% versus 63%) (62). Nevertheless, even among these studies where impact is likely overestimated, only one of a possible four shows a major

<sup>2</sup>Two studies, McPhee et al (62) and Lomas et al (55), are reported twice because they used both actual practice data and physician self-report.

**Table 1** Studies evaluating physicians' practices for impact of or conformity with consensus recommendations

Author	Topic	Method	Impact <sup>a</sup>	Outcome	
				Impact <sup>a</sup>	Percent conformity to recommendations <sup>b</sup>
<i>Using Actual Practice Data</i>					
Romm et al 1981	Cancer screening	Chart review, x-sectional	N/A		59
Dietrich & Goldberg 1984	Cancer screening	Chart review, x-sectional	N/A		49
Woo et al 1985	Periodic health exam tests	Chart review, x-sectional	N/A		<100 <sup>c</sup>
McPhee et al 1986	Cancer screening	Chart review, x-sectional	N/A		27 <sup>d</sup>
Laurie et al 1987	Preventive maneuvers	Claims data, x-sectional	N/A		47 <sup>e</sup>
Reichlin et al 1985	Coronary surgery	Chart review, x-sectional	N/A		<100 <sup>c</sup>
Gleicher 1984	Cesarean section	Hospital discharge data, time series	0		N/A
Fowkes & Roberts 1984	Chest x-rays	Hospital records, time series	+		N/A
Van Everdingen et al 1988	Blood transfusion	Chart review, time series	(+)		N/A
Lomas et al 1989	Cesarean section	Hospital discharge data, time series	0		N/A
Ford et al 1987	Breast cancer	Chart review, before-after	0		33
	Lung cancer	Chart review, before-after	0		67

Kosecoff et al 1987	Coronary surgery	Chart review, time series	0	57 <sup>f</sup>
	Breast cancer	Chart review, time series	0	
	Cesarean section	Chart review, time series	(+)	
	Cancer pathology	Chart review, before-after	+	97
<i>Using Physician Self Report</i>				
Battista 1983	Periodic health exam	Survey, x-sectional	N/A	61 <sup>h</sup>
Winkler et al 1989	8 NIH consensus topics <sup>b</sup>	Survey, x-sectional	N/A	51 <sup>i</sup>
Abelson & Lomas 1990	Periodic health exam	Survey, x-sectional	N/A	24-85
McPhee et al 1986	Cancer screening	Survey, x-sectional	N/A	63 <sup>d</sup>
Casparie et al 1987	Bed sores	Survey, x-sectional	+	N/A
Lomas et al 1989	Cesarean section	Survey, before-after	(+)	N/A
Hill et al 1988	Hypertension	Survey, before-after	0	73 <sup>k</sup>
Johnsson 1988	4 Swedish topics <sup>m</sup>	Survey, x-sectional	0	61-83

<sup>a</sup> 0 = no impact, (+) = minor impact, + = major impact, N/A = not applicable.

<sup>b</sup> percents may represent physicians or cases; N/A = not applicable.

<sup>c</sup> results not reported in a way amenable to calculating mean percent conformity.

<sup>d</sup> mean across 5 screening tests (2 excluded because their expected compliance would be <100); range 13-39 for actual practice and 56-81 for self report.

<sup>e</sup> mean across 9 maneuvers (2 excluded because their expected compliance would be <100); range 1-93.

<sup>f</sup> median value across all 11 consensus recommendations for post recommendation period; range 16-97.

<sup>g</sup> mean across 8 screening tests; range 8-99.

<sup>h</sup> the 8 topics were: coronary surgery, thrombolysis, estrogen use in post menopausal women, pap smear, breast cancer (2), cesarean section, antenatal diagnosis.

<sup>i</sup> median value across all 49 consensus recommendations; range 7.4-98.8.

<sup>j</sup> mean value across 10 recommendations; range 31-96.

<sup>k</sup> the four topics were: hip joint replacement, myocardial infarction, depressive disorders, sight improving surgery.

impact of the consensus on practice (another Dutch study). Percent conformity in the five studies with this measure is still well below 100% in all cases.

These evaluations suggest that, in North America at least, most consensus recommendations have little impact on the behavior of the practitioners at which they are targeted, and leave actual practice far short of what is recommended. The topic areas vary from preventive and public health issues through medical diagnosis and therapy to surgery, but no major differences are discernible. Given the relatively passive dissemination strategies used by most of these consensus exercises, perhaps the results are not surprising.

In light of my earlier discussion of newer models of information diffusion and behavior change, the most that we might expect is that the consensus recommendations would predispose physicians toward change, even if the recommendations fail to motivate or enable the actual change to occur. There is some evidence to support this view of consensus recommendations as "catalysts for consideration of change" from assessments of attitudes toward them and of the changes in attitude that they bring about. In one survey, over 90% of respondents considered consensus recommendations to be usually or sometimes "realistic for clinical practice" (46). In another, nearly 90% of the relevant specialists (obstetricians) fully agreed with the recommendations of a consensus on cesarean birth; one third claimed to have changed practice, even though validating data showed that they were not translating it into action (55).

The future value of consensus exercises may well be in "softening up" practitioners to implement action based on the recommendations. On the basis of reviews elsewhere, the most successful of such behavior change strategies operate at a more local level, and with more careful targeting, than is feasible with a national or regional consensus exercise (23, 56, 84). There is, however, some indication that this "symbiotic" relationship with active strategies that enable and reinforce behavior change may be a potentially fruitful role for future consensus exercises and the recommendations they produce (28, 53).

## SUMMARY

When existing evaluations find little or no evidence of consensus recommendations leading to action, one can justifiably ask why so much of this review was dedicated to analyzing alternative ways of producing such "words without action." There are, however, at least two reasons why consensus recommendations should be produced with care and attention to validity.

First, recommendations do sometimes have an impact on behavior as a consequence of mere dissemination activity—the Dutch program, for in-

stance, was more successful than most. This success may occur when the target audience is already particularly receptive to change and the message is timely and delivered by a credible source in a clinically relevant way. Thus, although "such a conjunction of favorable conditions is probably the exception rather than the rule for consensus topics" (46, 240) it does happen.

Second, the output from consensus processes is increasingly a potential input to other processes. Consensus recommendations can be used as the criteria for evaluation and appraisal aimed at changing practice behavior, making administrative decisions on resource allocation, or defining research protocols. For instance, quality assurance activities, such as peer assessment, practitioner certification, or utilization review, are actively seeking criteria with which to make judgments and elicit changes in practice to improve the quality of care. Funding agencies are looking for information to help make reimbursement, capital expenditure, or fee-for-service decisions on cessation of insurance for particular procedures or approaches. These uses of the consensus criteria are potentially major and controversial.

Therefore, even if dissemination rarely leads to action, consensus processes should still be done carefully and with valid techniques. The use of their recommendations embedded within other activities may well lead to (forced) changes in behavior. On ethical grounds alone, we should be as sure as possible that the behavior changes being implied and encouraged are indeed advisable.

For these reasons, the review describes the decision points in the production process for consensus recommendations as a start on the development of a set of recognized standards. The review offers a critical appraisal of the various methodological choices available at each decision point. The seven decision points are selecting a topic, picking the consensus group, providing background preparation, identifying information inputs, choosing a group judgment process, defining the criteria for recommendations, and choosing a report preparation procedure and format.

At least two important points emerged from this review. First, the research is often not well enough developed to give clear indications for many of the choices on what is the "best" alternative. Second, there is often not a single and definitive best, because ultimately choices are most importantly determined by the chosen objective of the exercise. For instance, a scientific expert panel is likely most appropriate for defining future research needs, but a nonexpert panel is preferable when the aim is public participation in technology decisions. It is hoped, however, that the current popularity of consensus processes, the increasing use of their outputs, and the expanding body of research on their conduct will make more definitive conclusions about appropriate alternatives and valid methods possible in the future.

## ACKNOWLEDGMENTS

The author receives personnel support as a Career Scientist from the Ontario Ministry of Health. Support for the preparation of this review, and much of the personal research reported in it, was provided by a grant from the National Health Research and Development Programme, Health and Welfare, Canada. For support, comments, and assistance with Table 1, I am grateful to B. J. Porter.

## Literature Cited

1. Abelson, J., Lomas, J. 1990. Do health service organizations and community health centres have higher disease prevention and health promotion levels than fee-for-service practices? *Can. Med. Assoc. J.* 142:575-81
2. Ahrens, E. H. 1985. The diet-heart question in 1985: Has it really been settled? *Lancet* 1:1085-87
3. Andersen, T. F., Mooney, G., eds. 1990. *The Challenge of Medical Practice Variations*. London: Macmillan
4. Andreassen, P. B. 1988. Consensus conferences in different countries. Aims and perspectives. *Int. J. Technol. Assess. Health Care* 4:305-8
5. Asch, S. M., Lowe, C. U. 1984. The consensus development program: Theory, process, and critique. *Knowl. Creation Diffus. Util.* 5:369-85
6. Avorn, J., Soumerai, S. B. 1983. Improving drug-therapy decisions through educational outreach. A randomized trial of academically based "detailing." *N. Engl. J. Med.* 308:1457-63
7. Battista, R. N. 1983. Adult cancer prevention in primary care: Patterns of practice in Quebec. *Am. J. Public Health* 73:1036-39
8. Battista, R. N., Fletcher, S. W. 1988. Making recommendations on preventive practices: Methodological issues. *Am. J. Prev. Med.* 4(Suppl.):53-67
9. Brook, R. H., Chassin, M. R., Fink, A., Solomon, D. H., Kosecoff, J., et al. 1986. A method for the detailed assessment of the appropriateness of medical technologies. *Int. J. Technol. Assess. Health Care* 2:53-64
10. Brook, R. H., Lohr, K. 1985. Efficacy, effectiveness, variations, and quality: Boundary-crossing research. *Med. Care* 23:710-22
11. Brook, R. H., Park, R. E., Winslow, C. M., Kosecoff, J. B., Chassin, M. R., et al. 1988. Diagnosis and treatment of coronary disease: Comparison of doctors' attitudes in the USA and the UK. *Lancet* 1:750-53
12. Bunker, J. P., Roos, L. L., Fowles, J., Roos, N. P. 1985. Information systems and routine monitoring in the United States and Canada—with examples from surgical practice. In *Oxford Textbook of Public Health*, ed. L. Holland. Toronto: Oxford Univ. Press. pp. 77-86
13. Caltorp, J. 1988. Consensus development conferences in Sweden. Effects on health policy and administration. *Int. J. Technol. Assess. Health Care* 4:75-88
14. Canadian Task Force on the Periodic Health Examination. 1979. The periodic health examination. *Can. Med. Assoc. J.* 121:1194-1254
15. Casparie, A. F., Klazinga, N. S., van Everdingen, J. J., Touw, P. P. 1987. Health-care providers resolve clinical controversies: The Dutch consensus approach. *Aust. Clin. Rev.* 7:43-47
16. Comroe, J. H. 1978. The road from research to new diagnosis and therapy. *Science* 200:931-37
17. Dalkey, N. 1969. An experimental study of group opinion: the Delphi Method. *Futures* (September), pp. 408-26
18. Delbecq, A. L., Van de Ven, A., Gustafson, D. H. 1975. *Group Techniques for Program Planning*. Glenview, Ill: Scott, Foresman
19. Dietrich, A. J., Goldberg, H. 1984. Preventive content of adult primary care: Do generalists and subspecialists differ? *Am. J. Public Health* 74:223-27
20. Eddy, D. M. 1982. Clinical policies and the quality of clinical practice. *N. Engl. J. Med.* 307:343-47
21. Eddy, D. M. 1989. *A Manual for Assessing Health Practices and Designing Practice Policies (Draft, May 31)*. Washington, DC: Counc. Med. Specialty Soc. Task Force Pract. Policies
22. Eichhorn, A. J., Cooper, J. B., Cullen, D. J., Maier, W. R., Philip, J. H., et al. 1986. Standards for patient monitoring

- during anesthesia at Harvard Medical School. *J. Am. Med. Assoc.* 256:1017-20
23. Eisenberg, J. M. 1986. *Doctors' Decisions and the Costs of Medical Care*. Ann Arbor, Mich: Health Adm. Press
  24. Evans, C. E., Haynes, R. B., Birkett, N. J. 1986. Does a mailed continuing education program improve physician performance? Results of a randomized trial in antihypertensive care. *J. Am. Med. Assoc.* 255:501-4
  25. Feeny, D., Guyatt, G., Tugwell, P., eds. 1986. *Health Care Technology: Effectiveness, Efficiency, and Public Policy*. Montreal: Inst. Res. Public Policy
  26. Fink, A., Kosecoff, J., Chassin, M. R., Brook, R. H. 1984. Consensus methods: Characteristics and guidelines for use. *Am. J. Public Health* 74:979-83
  27. Ford, L. G., Hunter, C. P., Diehr, P., Frelick, R. W., Yates, J. 1987. Effects of patient management guidelines on physician practice patterns: The community hospital oncology program experience. *J. Clin. Oncol.* 5:504-11
  28. Fowkes, F. G., Roberts, C. J. 1984. Introducing guidelines into clinical practice. *Effect. Health Care* 1:313-23
  29. Geertsma, R. H., Parker, R. C., Whitbourne, S. K. 1982. How physicians view the process of change in their practice behavior. *J. Med. Educ.* 57:752-61
  30. Glaser, E. M. 1980. Using behavioral science strategies for defining the state-of-the-art. *J. Appl. Behav. Sci.* 16:79-92
  31. Gleicher, N. 1984. Cesarean section rates in the United States: The short-term failure of the National Consensus Development Conference in 1980. *J. Am. Med. Assoc.* 252:3273-76
  32. Goodman, C., ed. 1988. *Medical Technology Assessment Directory: A Pilot Reference to Organizations, Assessments, and Information Sources*. Washington, DC: Natl. Acad. Press
  - 32a. Goodman, C., Baratz, S. R., eds. 1990. *Improving Consensus Development for Health Technology Assessment: An International Perspective*. Council Health Care Tech., Inst. Med., Washington, DC: Nat. Acad. Press
  33. Green, L. W., Eriksen, M. P. 1988. Behavioral determinants of preventive practices by physicians. *Am. J. Prev. Med.* 4(Suppl.):101-7
  34. Greer, A. L. 1987. The two cultures of biomedicine: Can there be consensus? *J. Am. Med. Assoc.* 258:2739-40
  35. Greer, A. L. 1988. The state of the art versus the state of the science: The diffusion of new medical technologies into practice. *Int. J. Technol. Assess. Health Care* 4:5-26
  36. Hill, M. N., Levine, D. M., Whelton, P. K. 1988. Awareness, use, and impact of the 1984 Joint National Committee consensus report on high blood pressure. *Am. J. Public Health* 78:1190-94
  37. Inst. Medicine. 1985. *Assessing Medical Technologies*. Washington, DC: Natl. Acad. Press
  38. Deleted in proof
  39. Jacoby, I. 1985. The consensus development program of the National Institutes of Health. *Int. J. Technol. Assess. Health Care* 2:420-32
  40. Jacoby, I. 1988. Evidence and consensus. *J. Am. Med. Assoc.* 259:3039
  41. Jacoby, I., Clark, S. 1986. Direct mailing as a means of disseminating NIH consensus statements. *J. Am. Med. Assoc.* 255:1328-30
  42. Jennett, B. 1985. First consensus development conference in United Kingdom: On coronary artery bypass grafting. II. Commentary by chairman of conference. *Br. Med. J.* 291:716-18
  43. Johnsson, M. 1988. Evaluation of the consensus conference program in Sweden. Its impact on physicians. *Int. J. Technol. Assess. Health Care* 4:89-94
  44. Joint Natl. Comm. Detection, Evaluation, Treatment of High Blood Pressure. 1977. Report of the JNC on Detection, Evaluation, and Treatment of High Blood Pressure. *J. Am. Med. Assoc.* 237:255-61
  45. Kahan, J. P., Kanouse, D. E., Winkler, J. D. 1988. Stylistic variations in Natl. Inst. Health consensus statements, 1979-1983. *Int. J. Technol. Assess. Health Care* 4:289-304
  46. Kanouse, D. E., Brook, R. H., Winkler, J. D., Kosecoff, J., Berry, S. H., et al. 1989. *Changing Medical Practice Through Technology Assessment. An Evaluation of the NIH Consensus Development Program*. Ann Arbor, Mich: Health Adm. Press
  47. Kanouse, D. E., Jacoby, I. 1988. When does information change practitioners' behavior? *Int. J. Technol. Assess. Health Care* 4:27-33
  48. Klazinga, N. S., Casparie, A. F., van Everdingen, J. J. 1987. Contribution of medical decision-making to consensus development conferences. *Health Policy* 8:339-46
  49. Kolata, G. 1985. Heart panel's conclusions questioned. *Science* 227:40-41
  - 49a. Kosecoff, J., Kanouse, D. E., Rogers, W. H., McCloskey, L., Winslow, C., Brook, R. H. 1987. Effects of the

- National Institutes of Health Consensus Development Program on physician practice. *J. Am. Med. Assoc.* 258:2708-13
50. Lane, D. A. 1988. Making recommendations on preventive practices: Methodological issues. *Am. J. Prev. Med.* 4(Suppl.):68-72
  51. Logan, A. G. 1984. Report of the Canadian Hypertension Society's consensus conference on the management of mild hypertension. *Can. Med. Assoc. J.* 131:1053-56
  52. Lomas, J. 1986. The consensus process and evidence dissemination. *Can. Med. Assoc. J.* 134:1340-41
  53. Lomas, J. 1989. *The role of a consensus statement in changing physicians' awareness, attitudes, knowledge, and behaviour.* Presented at Ann. Meet. Int. Soc. Technol. Assess. Health Care, 5th, London
  54. Lomas, J., Anderson, G. M., Enkin, M., Vayda, E., Roberts, R., et al. 1988. The role of evidence in the consensus process. Results from a Canadian consensus exercise. *J. Am. Med. Assoc.* 259:3001-5
  55. Lomas, J., Anderson, G. M., Pierre, K. D., Vayda, E., Enkin, M. W., et al. 1989. Do practice guidelines guide practice? The effect of a consensus statement on the practice of physicians. *N. Engl. J. Med.* 321:1306-11
  56. Lomas, J., Haynes, R. B. 1988. A taxonomy and critical review of tested strategies for the application of clinical practice recommendations; from "official" to "individual" clinical policy. *Am. J. Prev. Med.* 4(Suppl.):77-94
  57. Lord, C., Ross, L., Lepper, M. 1979. Biased assimilation and attitude polarization: The effect of prior theories on subsequently considered evidence. *J. Pers. Soc. Psychol.* 37:2098-109
  58. Lurie, N., Manning, W. G., Peterson, C., Goldberg, G. A., Phelps, C. A., et al. 1987. Preventive care: Do we practice what we preach? *Am. J. Public Health* 77:801-4
  59. Markle, G. E., Chubin, D. E. 1987. Consensus development in biomedicine: The liver transplant controversy. *Milbank Mem. Fund Q.* 65:1-24
  60. Marmot, M. G. 1986. Epidemiology and the art of the soluble. *Lancet* 1:897-900
  61. May, W. E. 1985. Consensus or coercion. *J. Am. Med. Assoc.* 254:1077
  62. McPhee, S. J., Richard, R. J., Solkowitz, S. N. 1986. Performance of cancer screening in a university general internal medicine practice: Comparison with the American Cancer Society guidelines. *J. Gen. Intern. Med.* 1:275-81
  63. Merrick, N. J., Fink, A., Park, R. E. 1987. Derivation of clinical indications for carotid endarterectomy by an expert panel. *Am. J. Public Health* 77:187-90
  64. Moore, C. 1986. *Group Decision-Making Techniques.* Beverly Hills, Calif: Sage
  65. Mullan, F., Jacoby, I. 1985. The town meeting for technology: The maturation of consensus conferences. *J. Am. Med. Assoc.* 254:1068-72
  66. Natl. Heart Lung Blood Inst., 1982. Management of patient compliance in the treatment of hypertension. Report of a Working Group. *Hypertension* 4:415-23
  67. Naylor, C. D., Basinski, A., Baigre, R. S., Goldman, B. S., Lomas, J. 1990. Placing patients in the queue for coronary revascularization: Evidence for practice variations from an expert panel process. *Am. J. Public Health.* In press
  68. Oliver, M. F. 1985. Consensus or non-consensus conferences on coronary heart disease. *Lancet* 1:1087-89
  69. Off. Med. Appl. Res. n.d. *Participants Guide to Consensus Development Conferences.* Bethesda, Md: Natl. Inst. Health
  70. Off. Med. Appl. Res. 1983. *Guidelines for the Selection and Management of Consensus Development Conferences.* Bethesda, Md: Natl. Inst. Health
  71. Panel Natl. Consensus Conf. Aspects Cesarean Birth. 1986. Indications for cesarean section: Final statement of the panel of the Natl. Consensus Conf. Aspects Cesarean Birth. *Can. Med. Assoc. J.* 134:48-52
  72. Park, R. E., Fink, A., Brook, R. H., Chassin, M. R., Kahn, K. L., et al. 1986. Physician ratings of appropriate indications for six medical and surgical procedures. *Am. J. Public Health* 76:766-72
  73. Pauker, S. 1986. Decision analysis as a synthetic tool for achieving consensus in technology assessment. *Int. J. Technol. Assess. Health Care* 2:83-91
  74. Perry, S. 1987. The NIH consensus development program: A decade later. *N. Engl. J. Med.* 317:485-88
  75. Physician Paym. Rev. Comm. 1989. *Annu. Rep. Congress,* pp. 219-30. Washington, DC: Physician Paym. Rev. Comm.
  76. Rahimtoola, S. H. 1981. A consensus on coronary bypass surgery. *Ann. Intern. Med.* 94:272-73
  77. Retchin, S. M., Fletcher, R. H., Buescher, P. C. 1985. The application of



- official policy: Prophylaxis recommendations for patients with mitral valve prolapse. *Med. Care* 23:1156-62
78. Richmond, J. B. 1978. Statement before the subcommittee on domestic and international scientific planning, analysis, and cooperation. *Congr. Rec.* Oct. 6
  79. Rogers, E. M., Larsen, J. K., Lowe, C. U. 1982. The consensus development process for medical technologies: A cross-cultural comparison of Sweden and the United States. *J. Am. Med. Assoc.* 248:1880-82
  80. Romm, F. J., Fletcher, S. W., Hulka, B. S. 1981. The periodic health examination: comparison of recommendations and internists' performance. *South. Med. J.* 74:265-70
  81. Roper, W., Winkenwerder, W., Hackbarth, G., Krakauer, H. 1988. Effectiveness in health care: An initiative to evaluate and improve medical practice. *N. Engl. J. Med.* 319:1197-1202
  82. Sackett, D. L. 1986. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest* 89 (Suppl.):2S-3S
  83. Sacks, H. S., Berrier, J., Reitman, D. 1987. Meta-analysis of randomized clinical trials. *N. Engl. J. Med.* 316:450-55
  84. Schroeder, S. A. 1987. Strategies for reducing medical costs by changing physicians' behavior: Efficacy and impact on quality of care. *Int. J. Technol. Asses. Health Care* 3:39-50
  85. Siu, A., Sonnenberg, F., Manning, W. H. 1986. Inappropriate use of hospitals in a randomized trial of health insurance plans. *N. Engl. J. Med.* 315:1259-66
  86. Stocking, B. 1985. First consensus development conference in the United Kingdom: On coronary artery bypass grafting. I Views of audience, panel and speakers. *Br. Med. J.* 291:713-16
  87. Task Force Pres. Advis. Group Anticipated Adv. Sci. Technol. 1976. The science court experiment: an interim report. *Science* 193:653-56
  88. van Everdingen, J. J., Klazinga, N. S., Casparie, A. F. 1988. Blood transfusion policy in Dutch hospitals. *Int. J. Health Care Qual. Assur.* 1:16-19
  89. van Everdingen, J. J., Rampen, F. H., Ruiters, D. J., Casparie, A. F. 1989. Evaluatie consensus melanoom van de hui op grond van pathologisch-anatomische verslagen. *Ned. Tijdschr. Geneesk.* 133:2285-88
  90. Vang, J. 1986. The consensus development conference and the European experience. *Int. J. Technol. Assess. Health Care* 2:65-76
  91. White, L. J., Ball, J. R. 1985. The Clinical Efficacy Assessment Project of the American College of Physicians. *Int. J. Technol. Assess. Health Care* 1:169-74
  92. Williams, A. H. 1985. Economics of coronary artery bypass grafting. *Br. Med. J.* 291:326-29
  93. Williamson, J. W., German, P. S., Weiss, R., Skinner, E. A., Bowes, F. 1989. Health science information management in continuing education of physicians: A survey of US primary care practitioners and their opinion leaders. *Ann. Intern. Med.* 110:151-60
  94. Winkler, J. D., Kanouse, D. E., Brodsky, L., Brook, R. H. 1986. Popular press coverage of eight Natl. Inst. Health consensus development topics. *J. Am. Med. Assoc.* 255:1323-27
  95. Winkler, J. D., Kanouse, D. E., Berry, S. H., Brook, R. H. 1989. Physicians' conformity to consensus recommendations. In *Changing Medical Practice Through Technology Assessment*, ed. D. E. Kanouse, pp. 102-26. Ann Arbor, Mich: Health Adm. Press
  96. Winkler, J. D., Lohr, K., Brook, R. H. 1985. Persuasive communication and medical technology assessment. *Arch. Intern. Med.* 145:314-17
  97. Woo, B., Woo, B., Cook, E. F., Weisberg, M., Goldman, L. 1985. Screening procedures in the asymptomatic adult: Comparison of physicians' recommendations, patients' desires, published guidelines, and actual practice. *J. Am. Med. Assoc.* 254:1480-84
  98. Wortman, P. M., Vinokur, A., Sechrest, L. 1982. *Evaluation of NIH Consensus Development Program, Phase I: Final Report*. Ann Arbor, Mich: Inst. Soc. Res., Univ. Mich.
  99. Wortman, P. M., Vinokur, A., Sechrest, L. 1988. Do consensus conferences work? A process evaluation of the NIH consensus development program. *J. Health Polit. Policy Law* 13:469-98

