

Paper #90-12

WORDS WITHOUT ACTION?

THE PRODUCTION, DISSEMINATION AND IMPACT
OF CONSENSUS RECOMMENDATIONS

Jonathan Lomas

CHEPA Working Paper Series #90-12
Sept. 1990
McMaster University
Hamilton, Ontario
Canada L8N 3Z5

CHEPA WORKING PAPER SERIES

The Centre for Health Economics and Policy Analysis (CHEPA) Working Paper Series provides for the circulation on a pre-publication basis of research conducted by CHEPA faculty, staff, and internal and external associates. The Working Paper Series is intended to stimulate discussion on analytical, methodological, quantitative, and policy issues in health economics and health policy analysis. The views expressed in the papers are the views of the author(s) and do not necessarily reflect the views of the Centre or its sponsors. Readers of Working Papers are encouraged to contact the author(s) with comments, criticisms, and suggestions.

A list of papers in the Series and papers in the CHEPA Health Policy Commentary Series, together with information on how to obtain specific titles, is provided inside the back cover. Further information can be obtained by contacting the Centre for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada L8N 3Z5; (416) 525-9140, ext. 2905.

THE CENTRE FOR HEALTH ECONOMICS AND POLICY ANALYSIS

The Centre for Health Economics and Policy Analysis officially opened in May, 1988. The objectives of the Centre include research, education, and the promotion of interactions among academic researchers and policy makers. CHEPA includes researchers from a number of departments in the University including Clinical Epidemiology and Biostatistics, Economics, Management Science, Family Medicine, and Geography.

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.

This review was prepared for the "Annual Review of Public Health, Volume 12." G.S. Omenn (ed.), forthcoming 1991, Palo Alto, CA.:Annual Reviews Inc.

CONTENTS

Introduction.....	1
The Production of Consensus.....	3
A Framework for Evaluating the Production of Consensus...	3
Selecting a Topic.....	4
Consensus Group Membership.....	5
Background Preparation.....	7
Information Inputs.....	8
Type of Group Judgement Process.....	10
Criteria for Qualification as a Recommendation.....	12
Report Preparation and Format.....	14
The Dissemination and Impact of Consensus.....	15
Models of Diffusion.....	15
Dissemination Strategies.....	16
The Impact of Dissemination.....	16
Summary.....	19
Acknowledgements.....	20
Literature Cited.....	21

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

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

INTRODUCTION

The quote from Israeli statesman Abba Eban voices the concern of many about the increasing use of consensus processes to provide an imprimatur for certain practice patterns in medicine. Some critics fear the implications of the consensus process for 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" (May 1985, 1077). Other critics claim that the methods used may overinterpret the available data and lead to conclusions based on "faith or zeal or alarm" (Ahrens 1985, 1086). Finally there are those who, if correct, can assuage the fear of the former critics because they do not believe that consensus processes effect behavior change in any case, as they are "primarily a dialogue among researchers ... not a guide to action" (Greer 1987, 2740).

Nevertheless, consensus processes as a means of information transfer appear to be here to stay and destined to become even more popular. A recent Directory on Technology Assessment listed nearly 60 organizations in the United States with formal programs establishing hundreds of recommendations for practice and nearly all using some form of consensus in their deliberations (Goodman 1988). Formal consensus development programs now exist in Canada, Britain, Sweden, Norway, Finland, Denmark, Holland and France (Institute of Medicine 1990). Although the terms used vary across a spectrum of "consensus development conferences", "task force reports", "appropriateness ratings", "practice parameters or guidelines" or "technology assessment reports", the general theme is to use varying degrees of formality to produce accessible and readily understandable consensus recommendations summarizing the implications of existing research evidence.

The need for such summaries arises from the burgeoning size of the biomedical and social science literature, increasing knowledge of the inadequacy of traditional journal articles as sources for direct adoption decisions (Greer 1988, Williamson et al 1989), evidence from medical practice variations of uncertainty on appropriate clinical policies (Andersen & Mooney 1990), awareness that a significant proportion of care is inappropriately provided (Brook & Lohr 1985, Siu et al 1986), a change in onus from provision of those services with expected benefit to only those with actual benefit (Eddy 1989), demands by third party funding and quality monitoring agents for guidance on and succinct recommendations about appropriate practice (Roper et al 1988), and pressure from the public for increased input to medical and technology decision-making (Andreasen 1988).

Such a variety of catalysts inevitably produces variety among the specific aims, the conduct, and the nature of the output from each of the consensus processes. The target audience may be as narrow as a particular kind of practitioner or researcher (Jacoby 1985, Brook et al

1986); as eclectic as politicians, administrators, clinicians, and planners (Stocking 1985, Calltorp 1988); or, in Denmark, as diffuse as "public participation ... for ensuring the democratic influence on decisions about medical technology" (Andreasen 1988, 308).

Nevertheless, the most common target is the practicing clinician, and the most common aim is to influence her to improve the quality of care provided, even though the influence may sometimes be indirect and through funding or regulatory bodies. This review is largely focused on this more restricted purpose and target of consensus processes.

Even with this restriction, however, numerous alternatives are available and have been tried. In the next section some of the methods for the production of consensus, and the controversies surrounding the choice among them, are reviewed. An early observation in this section is the lack of any recognized standards by which to judge the validity of the approaches to the production of consensus, and an intent of the review is to begin the task of assembling such a set of standards. The task can only be started here because, as yet, research comparing the effectiveness of different group judgement processes on "how best to put medical technologies to use, is not commensurate with the [research] effort and care devoted to developing these technologies" (Institute of Medicine 1985, 135).

In the second section the dissemination and impact of consensus is reviewed with a focus on the apparently false assumption that dissemination is both the necessary and sufficient condition for behavior change. The methods used for dissemination are reviewed, and the 19 studies identified in the literature that evaluate the impact of such dissemination on behavior are critically appraised. This appraisal yields a pessimistic conclusion -- in most cases the words do not translate into action. Some hope is found, 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.

The review concludes with some suggestions from recent work on how to improve the impact of the output from consensus processes and a tentative set of standards by which to judge the validity of different consensus production methods.

Materials for the review were derived from a combination of sources: 1. computerized searches of the U.S. National Library of Medicine data base (MEDLINE) and the Educational Resources Information Clearinghouse (ERIC) for 1980 on, using the search terms "consensus development," "guidelines," "standards," "official policy," "technology assessment," "evaluation studies," "epidemiologic methods," and "research", 2. bibliographies on consensus methodology (Fink et al 1984, Institute of Medicine 1990) and practice guidelines (Physician Payment Review Commission 1989), 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 the Production of Consensus

It is ironic that the standards routinely required of research studies before recognition of them as valid observations, 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 for the various stages in the production of consensus is often partially or wholly missing. Thus, many consensus reports would fail the tests of replicability -- on the basis of the information given could the methods in the exercise be replicated? -- and defensibility -- 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
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" (Eddy 1989, 3).

There are some existing consensus programs that come close to meeting the rigorous standards of explication and justification described above, even if there are those who argue with the particular choices made by the programs (e.g. Canadian Task Force on the Periodic Health Examination 1979, White & Ball 1985, Brook et al 1986). 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, that has come in for much scrutiny and criticism even from those involved in its organization, is the Consensus Development Conference (CDC) approach of the National Institutes of Health (NIH) (Jacoby 1985).

At the same time, the NIH CDC program has been subject to more systematic evaluations, and consequent modification, than any of its competitors, and has thus spawned much valuable research information on the validity of different approaches to producing consensus (Jacoby & Clark 1986, Markle & Chubin 1987, Wortman et al 1988, Kanouse et al 1989). Similarly, derivative CDC programs have also been evaluated in other countries such as Sweden (Galltorp 1988, Johnsson 1988), Holland (Casparie et al 1987, van Everdingen et al 1989), or Canada (Lomas et al 1988, Lomas et al 1989). There have even been cross-cultural comparisons of the processes and their outcomes (Rogers et al 1982, Brook et al 1988).

In the following sections the results of these and other evaluations are used to review the explicit methodological and "political" choices that have to be made for a consensus exercise aimed at translating biomedical research into practical advice and recommendations to clinical audiences.

Selecting a Topic

The topic can either be disease based e.g. prevention and treatment of breast cancer, or procedure based e.g. mammography (Physician Payment Review Commission 1989). In the former the focus is largely on the appropriate use of all the alternatives for alleviating a particular burden of ill health, whereas in the latter the focus is on the particular indications for which an intervention should be used, and is what is generally captured by the term 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, necessitating evaluation of the relative worth of different technologies for the same disease and raising the question of which criteria are most appropriate for such comparisons. For instance, how do you weigh the factors of cost, convenience, effectiveness, labeling and preference (patients' and providers') in comparing breast self-examination to mammography to dietary and reproductive behavior changes to lumpectomy in reducing the morbidity and mortality from breast cancer?

Focus on a specific technology may simplify the task, but it also leads to the danger that considerations of technical capabilities may override the criterion of actual health benefits. Should the threshold for routine screening be lowered to an age group at lower risk just because mammography now uses much reduced levels of radiation? Furthermore, the problem of timing is acute 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 (Feeny et al 1986). Many argued that NIH's 1984 CDC on lowering blood cholesterol to prevent heart disease was undertaken too early, with subsequent research making their recommendations highly suspect (Ahrens 1985, Kolata 1985, Marmot 1986). In contrast, the NIH's 1979 CDC on surgery for primary breast cancer was clearly too late as the principal change it recommended had already taken place (Kanouse et al 1989).

Insofar as the target audience for the output of the consensus exercise is the practicing clinician in the area under study, then diagnostic and surgical specialties may find procedure based topics most relevant, while medical and public health professionals may have greater interest in disease (or at least health problem) based topics.

Even after the type of topic has been decided upon and justified, the topic area has to be specified. A number of programs outline general criteria used to assist them in setting priorities across competing topics. These usually include at least the following three

general principles: important (on 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 (see, for instance, Office of Medical Applications of Research n.d., White and Ball 1985, Battista and Fletcher 1988). While these are not unreasonable as part of the criteria for topic selection, their focus is almost exclusively on the state of science in the potential topic area, and not on the state of practice. This omission has been noted in the context of the potential objective to alter practice patterns (Lomas 1986, Kosecoff et al 1987, Andreassen 1988, Perry 1987, Kanouse et al 1989). 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 an important part of any decision on whether a consensus on a particular topic is actually needed -- either because variations in practice demonstrate uncertainty as to what is appropriate, or because practice is uniformly not congruent with the message from current research evidence (Lomas 1986).

Consensus Group Membership

Three distinct types of consensus group membership emerge from a review of past exercises. Interestingly coronary artery bypass surgery has been subjected to each of the types. First, by both the Rand Corporation and a Canadian variant of the Rand approach with a group composed exclusively of experts on the specific clinical area under consideration -- cardiovascular surgeons and related specialists (Brook et al 1986, Naylor et al 1990). Second, a NIH CDC panel with experts in both the clinical and related scientific aspects of the topic -- clinicians plus lawyer, epidemiologist, economist and "expert consumer" (Rahimtoola 1981). Third, a non-expert or independent panel in the U.K. King's Fund conference including administrators, and laymen alongside providers from a variety of specialty and disciplinary backgrounds (Stocking 1985).

These three panel types -- clinical experts, scientific experts, and non-experts -- reflect the chosen focus and target audience of the consensus exercise. In the case of clinical expert panels the focus is almost entirely on the safety and effectiveness issues and the target audience the specialty physicians actually engaged in providing the care. Scientific expert panels have a broader mandate in considering such additional issues as ethics, economics or future research needs, and target an expanded audience of clinicians, scientists, and administrators. Finally the non-expert 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" (Stocking 1985, 713). The panel type is also influenced by the sponsor of the exercise, with medical specialty society panels tending toward the clinical expert variety, research council and university-based organizers favouring the scientific expert panel, and

public foundations or directly government-sponsored panels being more likely to be non-expert.

Given an objective of influencing the practice of providers 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 done in conjunction with the above King's Fund conference on coronary artery surgery. In answer to the question of whether the panel and the proceedings were biased too strongly by the cardiac medical specialties only 24 percent of the cardiac specialists surveyed replied yes, compared to 88 percent of community medicine specialists. This may have reflected the fact that those with a public health perspective are more favourably disposed to disease rather than procedure based topics; indeed, the community medicine specialists in general felt that the topic "was not presented in the whole context of the prevention and treatment of coronary heart disease" (Stocking 1985, 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 attitudes of the intended audience may be the most appropriate criterion to judge whether credibility was achieved. Had the topic been the prevention and treatment of coronary heart disease, then a panel and process less focussed 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 satisfies the apparent role of the consensus process in synthesizing research information, it potentially comes into conflict with the aim of being credible to a community practitioner audience. In other survey work significant resistance was found among community practitioners to "ivory tower" medicine as espoused by academic consensus panels (Greer 1988, Lomas & Haynes 1988). There is no obvious balance between these potentially competing demands on consensus group membership, but if identification with and credibility of recommendations is desired for community practitioners it appears to be advisable to at least ensure visible representation of their viewpoint.

Indeed, consensus group membership has inherent a number of tensions between the appeal of the panelists to the intended audience and the skills that the panel requires to adequately consider relevant viewpoints and appraise scientific evidence. Increasingly, however, there is recognition of the need to include at least an epidemiologist to bring both a public health perspective and methodological skills to the process (Kanouse et al 1989), and sometimes an economist to evaluate opportunity costs and resource allocation matters (Stocking 1985, Calltorp 1988).

Finally, an ongoing debate surrounding the NIH CDC program concerns whether panelists can feasibly be expected to meet the "science court" and "judicial jury" requirement (Mullan and Jacoby 1985) of not having strongly held pre-existing views on the topic under

consideration. This is more of a concern for the expert than the non-expert type of panel, given the likelihood that both expertise and opinions flow from involvement. The social psychology literature warns of the biasing potential of strongly held pre-existing views (Lord et al 1979), but other evidence suggests that when the methodologic quality of the evidence is stressed as the criterion for decision-making such pre-existing views can be appropriately altered by the consensus process (Lomas et al 1988). Some have argued that at least the panel chairman should be neutral toward the topic (Jennett 1985).

Overall, prior consensus exercises have not been very explicit about the criteria used to select topics and/or members of the consensus group; and any criteria that do exist have often 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" (Wortman et al 1988, 490). They point out that formalization of these selection processes can reduce the problems, and suggest the use of the Delphi method (Dalkey 1969) with relevant medical schools, researchers, and associations to "rapidly produce a list of questions, panelists, speakers and relevant research literature in two or three rounds of mailed questionnaires" (Wortman et al 1988, 491). Whether such a formal process is used or not, it is clearly advisable at least to describe the criteria and procedures used in selecting topics and participants, if only to be clear oneself that they relate directly to the objectives of the consensus exercise.

Background Preparation

The three 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, preferring to rely on the existing knowledge of a clinical expert panel (e.g. Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure 1977, Eichhorn 1986, Ford et al 1987). 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 appear to favour organization of a synthesis around the methodological quality of the various studies of interest (e.g. Canadian Task Force on the Periodic Health Examination 1979, White & Ball 1985, Brook et al 1986, Panel of the National Consensus Conference on Aspects of Cesarean Birth 1986). This 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.

The importance of this state of the science background preparation was described by one former director of the NIH CDCs when

bemoaning its sporadic availability for some 30 previous topics covered by his 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" (Jacoby 1988, 3039).

The infrequent use but high value of background input on the state of practice has already been described. In the context of panel preparation the value of this type of information is in 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" (Kanouse et al 1989, 242). If such background preparation had been included in a number of previous consensus processes it may have saved time and effort by preventing mere "codification" recommendations that reflected practice patterns that were apparently already fully diffused and in place (e.g. Kosecoff et al 1987, Hill et al 1988).

In countries where accessible data bases on practice patterns are routinely collected, such as Canada (Bunker et al 1985), the requirement can be met relatively easily. In countries such as the United States, where access to such data is more problematic, it has been suggested that specific surveys can be performed, or even a routine collection can be established using a panel of "Neilsen hospitals" (similar to the monitoring of families for television viewing patterns) (Kanouse et al 1989).

Information Inputs

In addition to background reviews prior to deliberations there is the question of the source and nature of information inputs during deliberations. The source, and to some extent the nature, of these inputs appear to 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, operating behind closed doors and considering little besides the published literature and their own views on safety and effectiveness (e.g. Canadian Task Force on the Periodic Health Examination 1979, Brook et al 1986). At the other end are the highly public forums, common in Scandinavian countries (Vang 1986, Andreasen 1988), soliciting both information and views from a wide variety of sources -- the general public, administrators, politicians, patients, researchers, and care-givers -- on numerous aspects of the topic.

These differences reflect, in part, the extent to which some processes aim at restricting deliberations to the results of

experimental research and others wish to synthesize values and other "normative" components with the "objective research". Some 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" (Vang 1986, 67).¹

In this way 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 less relevant to the world in which the decisions are actually being made.

The degree of relevance to public, or at least non-clinician, viewpoints is nevertheless central to many of the consensus programs that have so far been designed. It is probably no coincidence that perhaps 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 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" (Vang 1986, 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 taken into account by methods that ignore the importance of

¹ Some of the difficulties encountered by the NIH CDC program might be attributable to their ambiguity on this issue. On the one hand they restrict consideration to matters of safety and efficacy, but on the other hand they use a very open format allowing for input from numerous sources including the general public (Wortman et al 1988).

values, or at least patient utilities (Eddy 1989). A number of consensus exercises have, largely unsuccessfully, tried to incorporate this element using formal medical decision-making models (Pauker 1986, Klazinga et al 1987). Similar attempts to incorporate economic evaluations as inputs to consensus processes have met with somewhat more success (Williams 1985).

The concern expressed by many has been 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" (Lane 1988, 69). Some methodologies aimed at a relatively objective assimilation of disparate types of evidence are, however, being developed. There are, for instance, variations on meta-analysis (Sacks et al 1987), Bayesian meta-modeling (Lane 1988), and even a science of designing practice policies (Eddy 1989). With the increasing popularity of evidence synthesis, more innovation is to be expected.

The actual format of the information inputs does not appear to be crucial, although a number of surveys of panelists, speakers and audiences do 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 (Stocking 1985, Calltorp 1988, Johnsson 1988). One survey found that those exposed to such presentations and debate were the most likely to have taken actions based on the consensus report (Johnsson 1988).

Type of Group Judgement Process

As described above, there are a number of formal methodologies under development for the kind of synthesis and integration demanded in consensus processes. These, however, are largely related to marshalling the appropriate information inputs into a manageable format for consideration by the consensus group. The ground rules for the interactions among participants, and for actually defining a consensus, require a separate process. A number of excellent reviews of group judgement processes already exist (e.g. Delbecq et al 1975, Glaser 1980, Fink et al 1984, Institute of Medicine 1985, Moore 1986). Much has also been written about the science court approach (Task Force of the Presidential Advisory Group on Anticipated Advances in Science and Technology 1976) and the NIHs CDC variant involving elements of judicial process, scientific conference and the town hall meeting (Mullan & Jacoby 1985). Most, if not all, address three components -- how to generate a common focus in a group, what criteria to use to resolve controversy, and how to define consensus.

Generating and defining a common focus can obviously be helped by the provision of comprehensive background materials, as well as an opportunity for group members to have input to the chosen focus. The two major methods, however, have been the formulation of a small number of specific questions, the answers to which form the actual consensus

(Office of Medical Applications of Research 1983), or the development of a comprehensive set of scenarios for rating and/or discussion of appropriate intervention strategies by the group (Brook et al 1985, Park et al 1986, Naylor et al 1990).

The approach of formulating questions is most amenable to processes that have face-to-face contact among group members. Although there has been no validation of the practice, it is generally accepted that no more than four to six questions should be posed to a group, and that they should be amenable to concrete and unambiguous answers (Kahan et al 1988).

Constructing scenarios for rating is most valuable when a procedure based topic is under consideration, and when the main concern is safety and effectiveness. This is because disease based topics make it much more difficult to comprehensively describe all potential clinical situations, and it is even more difficult to integrate the economic, ethical or other social factors into the scenarios. Some have taken the approach of constructing representative rather than comprehensive scenarios to overcome this difficulty, but this counteracts its potential advantage of being done by mail -- in a manner similar to Delphi techniques (Dalkey 1969) -- because the group must come together to "fill the gaps" left by the representative scenarios. This hybrid approach can, nevertheless, be valuable wherein the scenarios generate a common focus and identify ahead of time the areas of agreement and disagreement between panelists; the focus can then be on initial disagreements in a face-to-face meeting using time more efficiently to address contentious areas in the form of pre-formulated questions (Lomas et al 1988).

Defining the criteria for resolution of conflict is perhaps the most difficult part of any group process. Most exercises have not made their criteria explicit, although a number have pointed to the importance of the role of the chairman in ensuring smooth operation of this aspect of group process (Jennett 1985, Institute of Medicine 1985, Wortman et al 1988). Implicit in the operation of a number of panels has been a hierarchy of criteria that placed controlled trial research above other less methodologically stringent research evidence which, in turn, is valued more highly than clinical experience or personal opinion. In this light 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 (Merrick et al 1987, Lomas et al 1988). In one exercise, when the above hierarchy of criteria was made explicit, disagreements among panelists prior to deliberations were resolved by the consensus process for 71 percent of situations where research evidence existed but for only 24 percent when no research was available (Lomas et al 1988).

Related to the criteria for resolving disagreement is the mechanism used to define consensus. This may be quantitative or qualitative. If quantitative, then less concern has to be taken about being explicit regarding the resolving criteria (and less discretion is left in the hands of the group's chairman), but the danger is that the basis for the consensus may be left unclear. A number of investigators have explored the properties of different quantitative definitions of consensus in the context of the ratings done on the type of scenarios described above (Brook et al 1986, Naylor et al 1990). No clear guidelines independent of the purposes of the consensus exercise 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" (Fink et al 1984).

Although much has been written describing alternative group judgement methods, surprisingly little research has been done relating chosen options to the outcomes of such exercises. Hence there is little information to assist prospective convenors of consensus processes in choosing among the alternative approaches for encouraging a common focus, defining criteria for conflict resolution, and defining consensus. The major need in this area may now be for more systematic assessments of the various methods when they are used under different consensus circumstances. Presently "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" (Institute of Medicine 1985, 135).

Criteria for Qualification as a Recommendation

Following an exhaustive evaluation of the NIH's CDC program Kanouse and his colleagues 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?" (Kanouse et al 1989, 243). Herein lies probably the single most contentious issue in the debates surrounding choice of consensus approach: whether the purpose of recommendations from consensus processes is to establish the best possible guidance for clinical care even in the face of imperfect or incomplete evidence, or whether it is to promulgate science based only on watertight conclusions derived from methodologically incontestable studies.

The latter approach places great (some say total) reliance on the randomized controlled trial (RCT). It has been the approach favoured by both the Canadian and the American Task Forces on the Periodic Health Examination, and by a number of epidemiologically based consensus conferences restricted to the safety and effectiveness of defined clinical interventions (e.g. National Heart Lung and Blood Institute 1982, Logan 1984, Sackett 1986). When the purpose of the exercise is entirely science related e.g. establishing future research requirements,

the stringency of such a criterion is clearly appropriate.

For the development of practice guidance, however, there are at least three problems with relying so strictly on randomized controlled trial evidence as the only justification for a recommendation. First it significantly limits the areas of clinical practice in which consensus recommendations can be made. For instance, studies of causation in occupational health would obviously be unethical when using a RCT. For many preventive medicine and public health issues the length of time between a preventive intervention and its potential outcome is so long as to make RCTs infeasible (Battista & Fletcher 1988).

Second, RCTs alone are largely unable to take account of economic, ethical or other social considerations. Thus, for example, a highly effective but extremely expensive drug may not warrant recommendation given the opportunity cost of its use -- precisely the debate in the recent case of tissue plasminogen activator versus streptokinase for thrombolysis.

Third, negative recommendations are sometimes appropriate because of diffusion that is too fast or because of potential over-diffusion of management options or technologies e.g. "the use of cesarean section is not indicated for women with an uncomplicated previous cesarean section". Indeed, a major impetus for NIHS CDC program was the perception of Congress that there was widespread use of many technologies "without sufficient information about their health benefits, clinical risks, cost effectiveness, and societal side effects" (Richmond 1978). In this case failing 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 there is little doubt that 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 (Eddy 1982, Sackett 1986). Therefore even if the purpose of the consensus exercise is clearly to provide guidance for practice, and not the strict promulgation of science, accurate guidance still requires a distinction between those recommendations supported by RCTs, and those supported by evidence of less certainty.

A number of systems have been proposed for such grading of recommendations (Canadian Task Force on the Periodic Health Examination 1979, Sackett 1986, Eddy 1989). One of the simplest is one in which the term "recommendation" is reserved for when there is a high level of certainty because of support from methodologically sound studies, and some lesser term such as "guideline" or "suggestion" is used for situations supported by less certain forms of evidence (Lomas et al 1988). The particular grading system chosen is less important than the explicit recognition that different grades of recommendation do exist, based on the methodologic quality of the supporting evidence (Jacoby

1988). Thus consensus groups need not be prevented from producing conclusions that rely largely upon their experience or their interpretation of ethical and social considerations, although it is advisable that they 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 chairman, and on to joint efforts of the entire consensus group. The importance of making the report readily understandable and accessible has given rise to a number of recommendations that a professional writer should at least be involved in drafting the final wording of reports (Jacoby 1985, Wortman et al 1988, Kanouse et al 1989).

The time taken over report preparation has been the subject of much greater debate. On the one hand there is a desire to put panels under some 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 has resulted in one strategy that has the consensus group drafting and finalizing the report over a period of 24 to 48 hours (Jacoby 1985). The approach has come in for much criticism because it leads to "lowest common denominator" recommendations on particularly difficult and controversial issues (Wortman et al 1988), or "is bound to lead to hurried conclusions" (Oliver 1985, 1088).

On the other hand lengthy iterative processes have been used allowing for careful consideration of complex issues, but procedurally (it is argued) 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 try and obtain the best of both worlds and maximize the amount of input to the final product. After preparation of an initial draft under time pressure, circulation and feedback is undertaken, followed by final drafting some weeks or months later (Oliver 1985, Lomas 1986, Perry, 1987).

The format of reports that have been perceived as high quality and/or particularly influential has been the subject of two evaluations (Wortman et al 1982, Kahan et al 1988). Conclusions from both are similar and not surprising in suggesting 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" (Kanouse et al 1989, 26).

These suggestions obviously 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 judgement on the value and limitations of new scientific works" (Comroe 1978, quoted

in Wortman et al 1988). In one recent survey of physicians' information preferences only about one third wished to receive information in "complete form (with evidence)", whereas the other two thirds preferred it in "summary form (with references)". In the same survey almost 100 percent had a preference for clinically rather than research oriented information. Nevertheless, about 60 percent believed that reports in professional medical journals were very important in first hearing about or deciding to use a new procedure (Kanouse et al 1989).

A credible and potentially influential format is, therefore, one that clearly demonstrates that a scholarly process has been carefully followed in the development of the consensus, but presents the guidance in a non-scholarly and easy-to-read format, preferably with references appended (Perry 1987) and estimates of expected outcomes if recommendations are followed (Institute of Medicine 1985, Eddy 1989). Finally, the future validation and development of consensus methodology can best be advanced by having within the report explicit descriptions of the procedures employed and 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 produced by an exercise.

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" (Lomas & Haynes 1988, 90). This model of the practitioner has been called into question by a number of recent reviews which point out that research information (synthesized or otherwise) is only one of a number of 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, with the implication that publication without regard for such interactions is a very weak form of dissemination (Asch & Lowe 1984, Winkler et al 1985). Furthermore, the process of behavior change is viewed as requiring 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 (Geertsma et al 1982, Fowkes & Roberts 1984, Green & Eriksen 1988, Greer 1988, Kanouse & Jacoby 1988). Thus dissemination should be directed not only at

increasing awareness but also at producing impacts on 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 booklet form, sometimes in specialty journals, but most often in general medical journals. Some of the more high profile programs also rely on the popular press and media for short-term dissemination by holding press conferences. The quality of this reporting, when assessed, has been judged as largely factual and balanced (Winkler et al 1986).

Multiple sources are identified by practitioners for where they obtained their awareness of a consensus statement. The three most frequently cited potential sources were professional medical journals (50 percent), printed materials such as booklets (30 percent), and the popular press (25 percent) (Kanouse et al 1989, Lomas et al 1989).

However, awareness of consensus statements among the entire relevant population of practitioners has been very variable, sometimes as low as 20 percent (cardiac surgery: Kanouse et al 1989), other times as high as 90 - 95 percent (obstetricians: Lomas et al 1989, Swedish physicians: Johnsson 1988), but most usually in the 30 - 60 percent range (Hill et al 1988, Kanouse et al 1989, Fowler et al 1989, Abelson and Lomas 1990). For any specific statement it would appear that specialists are more likely to be aware of the recommendations than are general or family practitioners (Johnsson 1988, Kanouse et al 1989).

Direct mailing of printed versions of the statement to the relevant practitioner population do appear able to increase levels of awareness, but even then awareness attributable to this source does not seem to exceed about 40 percent (Jacoby & Clark 1986, Lomas et al 1989,). This level of awareness can, however, be significantly increased by making the mailed materials visually attractive and/or "staging" their delivery to the practitioner by dividing them into bite-sized chunks of information (Avorn & Soumerai 1983, Evans et al 1986, Fowler et al 1989). Unfortunately, as discussed below, even this high level of awareness may not be reflected in a consequent change in behavior.

The Impact of Dissemination

Methodologically it is not easy to do definitive evaluations of the impact of consensus exercises. By their nature they are widely disseminated, making control groups impossible because of the difficulty of insulating an experimentally defined portion of the relevant practitioners from exposure to the consensus. One possibility is to evaluate impact separately for those 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 (Kanouse et

al 1989, Johnsson 1988), but sometimes not (Hill 1988).

The difficulty with this approach, however, is that the assessment is not representative of the target population of relevant practitioners. 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 (Kanouse et al 1989), 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 using actual practice data, rather than self report, and taking measurements both before and after the consensus (preferably with a time series to account for pre-existing 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 degree of 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 this provides a measure of 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 a summary of 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, and 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, and are divided according to whether they used actual practice data or self reports of behavior.²

In the ten instances where impact was measured using actual practice data, six found no impact, two found a minor impact and two a major impact. Interestingly, three of the four studies showing any impact were from Europe (Fowkes & Roberts - U.K., van Everdingen et al 1988, 1989 - Holland), but 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

² Two studies, McPhee et al 1986 and Lomas et al 1989, are reported twice because they used both actual practice data and physician self-report.

Dutch studies, done 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 percent already.

The self report studies are a less reliable indicator of impact. For instance, 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 percent versus 63 percent) (McPhee et al 1986). Nevertheless, even among these studies, where impact is likely overestimated, only one of a possible four shows a major impact of the consensus on practice (another Dutch study), and percent conformity in the five studies with this measure is still well below 100 percent 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 covered vary from preventive and public health issues through medical diagnosis and therapy to surgery, but no major differences are discernible by topic area. Given the relatively passive dissemination strategies used by most of these consensus exercises the results are, perhaps, not surprising.

In the light of the earlier discussion of the newer models of information diffusion and behavior change, the most that might be expected is that the consensus recommendations act as a trigger for physicians to become predisposed toward change, even if they fail to motivate or enable the actual change to occur. There is, in fact, 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 percent of respondents considered consensus recommendations to be usually or sometimes "realistic for clinical practice" (Kanouse et al 1989); and in another nearly 90 percent of the relevant specialists (obstetricians) fully agreed with the recommendations of a consensus on cesarean birth, and one third claimed to have changed practice, even though validating data showed that they were not in fact translating it into action (Lomas et al 1989).

This being the case, the future value of consensus exercises may well be in "softening up" practitioners to be receptive to more active strategies for implementing the recommendations from the process. On the basis of reviews elsewhere, it would appear that 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 (Eisenberg 1986, Schroeder 1987, Lomas & Haynes 1988). There is, however, some indication that this "symbiotic" relationship with active strategies that enable and reinforce behavior change may be a potentially fruitful future role for consensus exercises and the recommendations they produce (Fowkes & Roberts 1984, Lomas 1989).

SUMMARY

When existing evaluations find little or no evidence of words in consensus recommendations leading to action in physicians' practices, 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 the production of consensus recommendations should be done with care and attention to validity.

First, recommendations do sometimes have an impact on behaviour as a consequence of mere dissemination activity -- the Dutch program, for instance, appears to have been more successful than most. This may be when the target audience is already particularly receptive to change, the message is timely, and it is delivered by a credible source in a clinically relevant way. Thus, although "such a conjunction of favourable conditions is probably the exception rather than the rule for consensus topics" (Kanouse et al 1989, 240), it does happen.

Second, the output from consensus processes is increasingly a potential input to other processes. Consensus recommendations can be, and are being, used as the criteria for evaluation and appraisal aimed at changing practice behaviour, 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 judgements and elicit changes in practice to improve the quality of care. Funding agencies are looking for information to help in making reimbursement, capital expenditure or fee-for-service deinsurance decisions. These are potentially major and controversial uses of the criteria implicit in consensus recommendations.

Therefore, even if dissemination rarely leads to action, consensus processes should still be done carefully and with valid techniques because the use of their recommendations embedded within other activities may well lead to (forced) changes in behaviour. Thus, on ethical grounds alone, one should be as sure as possible that the behaviour 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, and offers a critical appraisal of the various methodological choices available at each decision point. The seven decision points were: selecting a topic, picking the consensus group, providing background preparation, identifying information inputs, choosing a group judgement process, defining the criteria for recommendations, and opting for a report preparation procedure and format.

At least two important points emerged from this review. First, that in many instances the research is not yet well enough developed to give clear indications for many of the choices on what is the "best" alternative. Second, that 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 non-expert 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.

Acknowledgements: 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 BJ Porter.

Literature Cited

- 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. In press.
- Ahrens, E. H. 1985. The diet-heart question in 1985: has it really been settled? Lancet (May 11), pp. 1085-87.
- Andersen, T. F., Mooney, G., eds. 1990. The Challenge of Medical Practice Variations. London: Macmillan Publishers.
- Andreasen, P. B. 1988. Consensus conferences in different countries. Aims and perspectives. Int. J. Tech. Assess. Health Care 4:305-8.
- Asch, S. M., Lowe, C. U. 1984. The consensus development program: theory, process, and critique. Knowledge: Creation, Diffusion, Utilization 5:369-85.
- 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.
- Battista, R. N. 1983. Adult cancer prevention in primary care: patterns of practice in Quebec. Am. J. Public Health 73:1036-39.
- Battista, R. N., Fletcher, S. W. 1988. Making recommendations on preventive practices: methodological issues. Am. J. Prev. Med. 4:Suppl:53-67.
- Brook, R. H., Lohr, K. 1985. Efficacy, effectiveness, variations, and quality: boundary-crossing research. Med. Care 23:710-22.
- 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. Tech. Assess. Health Care 2:53-64.
- Brook, R. H., Park, R. E., Winslow, G., Kosecoff, J., Chassin, M. R., et al. 1988. Diagnosis and treatment of coronary disease: comparison of doctors' attitudes in the USA and the UK. Lancet (April 2), pp. 750-53.
- 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 University Press.
- Galltorp, J. 1988. Consensus development conferences in Sweden. Effects on health policy and administration. Int. J. Tech. Assess. Health Care 4:75-88.
- Canadian Task Force on the Periodic Health Examination. 1979. The periodic health examination. Can. Med. Assoc. J. 121:1194-1254.
- Casparie, A. F., Klazinga, N. S., van Everdingen, J. J., Touw, P. P. 1987. Health-care providers resolve clinical controversies: the Dutch consensus approach. Australian Clinical Review (March), pp. 43-47.
- Comroe, J. H. 1978. The road from research to new diagnosis and therapy. Science 200:931-37.
- Dalkey, N. 1969. An experimental study of group opinion: the Delphi Method. Futures (September), pp. 408-26.

- Delbecq, A. L., Van de Ven, A., Gustafson, D. H. 1975. Group Techniques for Program Planning. Glenview: Scott, Foresman.
- Dietrich, A. J., Goldberg, H. 1984. Preventive content of adult primary care: do generalists and subspecialists differ? Am. J. Public Health 74: 223-27.
- Eddy, D. M. 1982. Clinical policies and the quality of clinical practice. N. Engl. J. Med. 307:343-47.
- Eddy, D. M. 1989. A Manual for Assessing Health Practices and Designing Practice Policies (Draft, May 31). Washington D.C.: The Council of Medical Specialty Societies Task Force on Practice Policies.
- 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. JAMA 256:1017-20.
- Eisenberg, J. M. 1986. Doctors' Decisions and the Costs of Medical Care. Ann Arbor, Michigan: Health Administration Press.
- 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. JAMA 255:501-4.
- Feeny, D., Guyatt, G., Tugwell, P., eds. 1986. Health Care Technology: Effectiveness, Efficiency and Public Policy. Montreal: Institute for Research and Public Policy.
- 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.
- 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.
- Fowkes, F. G., Roberts, C. J. 1984. Introducing guidelines into clinical practice. Effective Health Care 1:313-23.
- 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.
- Glaser, E. M. 1980. Using behavioral science strategies for defining the state-of-the-art. J. Applied Behav. Sci. 16:79-92.
- Gleicher, N. 1984. Cesarean section rates in the United States: the short-term failure of the National Consensus Development Conference in 1980. JAMA 252:3273-76.
- Goodman, C. ed. 1988. Medical Technology Assessment Directory: A Pilot Reference to Organizations, Assessments, and Information Sources. Washington, D. C.: National Academy Press.
- Green, L. W., Eriksen, M. P. 1988. Behavioral determinants of preventive practices by physicians. Am. J. Prev. Med. 4:Suppl:101-7.
- Greer, A. L. 1987. The two cultures of biomedicine: can there be consensus? JAMA 258:2739-40.
- 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. Tech. Assess. Health Care 4:5-26.

- 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.
- Institute of Medicine, 1985. Assessing Medical Technologies. Washington, D. C.: National Academy Press.
- Institute of Medicine, 1990. Proceedings of a Conference on Consensus Methods. Washington, D. C.: National Academy Press.
- Jacoby, I. 1985. The consensus development program of the National Institutes of Health. Int. J. Tech. Assess. Health Care 2:420-32
- Jacoby, I. 1988. Evidence and consensus. JAMA 259:3039.
- Jacoby, I., Clark, S. 1986. Direct mailing as a means of disseminating NIH consensus statements. JAMA 255:1328-30.
- Jennett, B. 1985. First consensus development conference in United Kingdom: on coronary artery bypass grafting. II Commentary by chairman of conference. BMJ 291:716-18.
- Johnsson, M. 1988. Evaluation of the consensus conference program in Sweden. Its impact on physicians. Int. J. Tech. Assess. Health Care 4:89-94.
- Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure, 1977. Report of the JNC on Detection, Evaluation, and Treatment of High Blood Pressure. JAMA 237:255-61.
- Kahan, J. P., Kanouse, D. E., Winkler, J. D. 1988. Stylistic variations in National Institutes of Health consensus statements, 1979-1983. Int. J. Tech. Assess. Health Care 4:289-304.
- Kanouse, D. E., Jacoby, I. 1988. When does information change practitioners' behavior? Int. J. Tech. Assess. Health Care 4:27-33.
- 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, Michigan: Health Administration Press.
- 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.
- Kolata, G. 1985. Heart panel's conclusions questioned. Science 227:40-41.
- Lane, D. A. 1988. Making recommendations on preventive practices: methodological issues. Am. J. Prev. Med. 4:Suppl:68-72.
- 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.
- Lomas, J. 1986. The consensus process and evidence dissemination. Can. Med. Assoc. J. 134:1340-41.
- Lomas, J. 1989. The role of a consensus statement in changing physicians' awareness, attitudes, knowledge, and behaviour. Presented at Ann. Meet. Int. Soc. Tech. Assess. Health Care, 5th, London.
- 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.
- 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. JAMA 259:3001-5.
- 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.
- 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.
- 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.
- Markle, G. E., Chubin, D. E. 1987. Consensus development in biomedicine: the liver transplant controversy. Milbank Q. 65:1-24.
- Marmot, M. G. 1986. Epidemiology and the art of the soluble. Lancet (April 19), pp. 897-900.
- May, W. E. 1985. Consensus or coercion. JAMA 254:1077.
- 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.
- 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.
- Moore, C. 1986. Group Decision-Making Techniques. Beverly Hills, California: Sage Publications.
- Mullan, F., Jacoby, I. 1985. The town meeting for technology: the maturation of consensus conferences. JAMA 254:1068-72.
- National Heart Lung and Blood Institute, 1982. Management of patient compliance in the treatment of hypertension. Report of a Working Group. Hypertension 4: 415-23.
- 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.
- Oliver, M. F. 1985. Consensus or nonsensus conferences on coronary heart disease. Lancet (May 11), pp. 1087-89.
- Office of Medical Applications of Research, n.d. Participants Guide to Consensus Development Conferences. Bethesda, Maryland: National Institutes of Health.
- Office of Medical Applications of Research, 1983. Guidelines for the Selection and Management of Consensus Development Conferences. Bethesda, Maryland: National Institutes of Health.
- Panel of the National Consensus Conference on Aspects of Cesarean Birth, 1986. Indications for cesarean section: final statement of the panel of the National Consensus Conference on Aspects of Cesarean

- Birth. Can. Med. Assoc. J. 1348-52.
- 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.
- Pauker, S. 1986. Decision analysis as a synthetic tool for achieving consensus in technology assessment. Int. J. Tech. Assess. Health Care 2:83-91.
- Perry, S. 1987. The NIH consensus development program: a decade later. N. Engl. J. Med. 317:485-88.
- Physician Payment Review Commission, 1989. Annual Report to Congress. Washington, D. C.: Physician Payment Review Commission, pp. 219-30.
- Rahimtoola, S. H. 1981. A consensus on coronary bypass surgery. Annals Intern. Med. 94:272-73.
- 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.
- Richmond, J. B. 1978. Statement before the subcommittee on domestic and international scientific planning, analysis, and cooperation. Congressional Record. October 6.
- Rogers, E. M., Larsen, J. K., Lowe, G. U. 1982. The consensus development process for medical technologies: a cross-cultural comparison of Sweden and the United States. JAMA 248:1880-82.
- 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.
- 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. 319:1197-202.
- Sackett, D. L. 1986. Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest 89:Suppl.:2S-3S.
- Sacks, H. S., Berrier, J., Reitman, D. 1987. Meta-analysis of randomized clinical trials. N. Engl. J. Med. 316:450-55.
- Schroeder, S. A. 1987. Strategies for reducing medical costs by changing physicians' behavior: efficacy and impact on quality of care. Int. J. Tech. Assess. Health Care 3:39-50.
- 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.
- Stocking, B. 1985. First consensus development conference in the United Kingdom: on coronary artery bypass grafting. I Views of audience, panel and speakers. BMJ 291:713-16.
- Task Force of the Presidential Advisory Group on Anticipated Advances in Science and Technology. 1976. The science court experiment: an interim report. Science 193:653-56.
- 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.
- van Everdingen, J. J., Rampen, F. H., Ruiter, D. J., Casparie, A. F.

1989. Evaluatie consensus melanoom van de hui op grond van pathologisch-anatomische verslagen. Ned. Tijdschr. Geneeskd 133:2285-88.
- Vang, J. 1986. The consensus development conference and the European experience. Int. J. Tech. Assess. Health Care 2:65-76.
- White, L. J., Ball, J. R. 1985. The Clinical Efficacy Assessment Project of the American College of Physicians. Int. J. Tech. Assess Health Care 1:169-74.
- Williams, A. H. 1985. Economics of coronary artery bypass grafting. BMJ 291:326-29.
- 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 U.S. primary care practitioners and their opinion leaders. Annals of Inter. Med. 110:151-60.
- Winkler, J. D., Lohr, K., Brook, R. H. 1985. Persuasive communication and medical technology assessment. Arch. Intern. Med. 145:314-17.
- Winkler, J. D., Kanouse, D. E., Brodsky, L., Brook, R. H. 1986. Popular press coverage of eight National Institutes of Health consensus development topics. JAMA 255: 1323-27.
- 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, 102-126. Ann Arbor, Michigan: Health Administration Press.
- 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. JAMA 254:1480-84.
- Wortman, P. M., Vinokur, A., Sechrest, L. 1982. Evaluation of NIH Consensus Development Program. Phase I: Final Report. Ann Arbor, Michigan: Institute of Social Research, University of Michigan.
- Wortman, P. M., Vinokur, A., Sechrest, L. 1988. Do consensus conferences work? A process evaluation of the NIH consensus development program. J. Health Politics Policy Law 13:469-98.

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

Author	Topic	Method	Outcome	
			Impact ^a	Percent conformity to recommendations ^b
<u>Using Actual Practice Data</u>				
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 ^c
McPhee et al 1986	Cancer screening	Chart review, x-sectional	N/A	27 ^d
Lurie et al 1987	Preventive maneuvers	Claims data, x-sectional	N/A	47 ^e
Retchin et al 1985	Coronary surgery	Chart review, x-sectional	N/A	<100 ^e
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 ^f
	Breast cancer	Chart review, time series	0	
	Cesarean section	Chart review, time series	(+)	
Van Everdingen et al 1989	Cancer pathology	Chart review, before-after	+	97
<u>Using Physician Self Report</u>				
Battista 1983	Periodic health exam	Survey, x-sectional	N/A	61 ^g
Winkler et al 1989	8 NIH consensus topics ^h	Survey, x-sectional	N/A	51 ^j
Abelson & Lomas 1990	Periodic health exam	Survey, x-sectional	N/A	<100 ^c
McPhee et al 1986	Cancer screening	Survey, x-sectional	N/A	63 ^d
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 ^k
Johnsson 1988	4 Swedish topics ^m	Survey, x-sectional	0	61 - 83

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

^b percents may represent physicians or cases; N/A = not applicable.

^c results not reported in a way amenable to calculating mean percent conformity.

^d 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.

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

^f median value across all 11 consensus recommendations for post recommendation period; range 16 - 97

^g mean across 8 screening tests; range 8 - 99.

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

^j median value across all 49 consensus recommendations; range 7.4 - 98.8.

^k mean value across 10 recommendations; range 31 - 96.

^m the four topics were: hip joint replacement, myocardial infarction, depressive disorders, sight improving surgery.

CENTRE FOR HEALTH ECONOMICS AND POLICY ANALYSIS

WORKING PAPER RELEASES

<u>Number</u>	<u>Title</u>	<u>Author(s)</u>	<u>Price</u>
No. 1	Physicians' Choices of Specialty, Location and Mode: A Re-examination Within an Interdependent Decision Framework, May 1988, 63 p.	Jeremiah E. Hurley	7.50
No. 2	Accommodating Rapid Growth in Physician Supply: Lessons from Israel, Warnings for Canada, May 1988, 31 p.	Morris L. Barer Amiram Gafni Jonathan Lomas	5.00
No. 3	The Role of Evidence in the Consensus Process: Results from a Canadian Consensus Exercise, May 1988, 23 p.	Jonathan Lomas Geoffrey Anderson Murray Enkin et al.	5.00
No. 4	Toward Improved Health Technology Policy in Canada: A Proposal for the National Health Technology Assessment Council, May 1988, 27 p.	David Feeny Greg L. Stoddart	5.00
No. 5	Cost-Effectiveness Analysis of Primary Tetanus Immunization Among Elderly Canadians, May 1988, 30 p.	Brian Hutchison Greg L. Stoddart	5.00
No. 6	A Framework for Comparative Evaluation of Risk Management Programs, May 1988, 47 p.	George W. Torrance Andrew D. Oxman	7.50
No. 7	Health Technology in Ontario Report to the Ontario Health Review Panel, June 1988, 43 p.	David Feeny	7.50
No. 8	Promoting Clinical Policy Change: Using The Art To Promote The Science In Medicine, December 1988, 27 p.	Jonathan Lomas	5.00
No. 9	The Effect of Malpractice Claims On Physicians' Behavior: A Theoretical Approach, January 1989, 25 p.	Amiram Gafni Giora Moore	5.00
No. 10	Minding Our Ps and Qs: Simultaneous Control of the Price and Quantity of Physician Services in Canada, January 1989, 36 p.	Jonathan Lomas Catherine Fooks Thomas Rice Robert J. Labelle	5.00
No. 11	Simulated Effects of Incomes-Based Policies on the Distribution of Physicians, March 1989, 36 p.	Jeremiah E. Hurley	5.00
No. 12	Incorporating Utility-Based Quality-of-Life Assessment Measures in Clinical Trials: Two Examples, April 1989, 25 p.	David Feeny George W. Torrance	5.00
No. 13	Evaluating Health Related Quality of Life: An Indifference Curve Interpretation for The Time Trade-Off Technique, April 1989, 22 p.	Abraham Mehrez Amiram Gafni	5.00

WORKING PAPER RELEASES (Cont'd)

No. 14	Publicly Financed Competition in Health Care Delivery: A Canadian Simulation Model, May 1989, 46 p.	Jacqueline M. Muldoon Greg L. Stoddart	7.50
No. 15	The Relative Cost-effectiveness of Water Fluoridation Across Communities: Analysis of Variations According to Underlying Caries Levels, May 1989, 22 p.	Stephen Birch	5.00
No. 16	Students Performance vs Hospital Performance in the Labor Market for Medical Interns and Residents: Do hospitals Perform Better Due to Asymmetry in Information Available to Them Relative to Students? June 1989, 11 p.	Amiram Gafni Yufei Yuan	5.00
No. 17	Managerial Approach to Allocating Indirect Fixed Costs in Health Care Organizations, August 1989, 24 p.	Yaaqov Goldschmidt Amiram Gafni	5.00
No. 18	Do Physicians Induce Demand for Medical Services? A Response to Feldman and Sloan, August 1989, 30 p.	Thomas H. Rice Roberta J. Labelle	5.00
No. 19	Policy Commentary: Is the Canadian Health Care System Suffering from Abuse? August 1989, 25 p.	Christel A. Woodward Greg L. Stoddart	5.00
No. 20	The Compensating Variation Approach: The Case Of Lives Saved Versus Lives Taken, August 1989, 9 p.	Amiram Gafni S. Abraham Ravid	5.00
No. 21	Do Health Service Organizations and Community Health Centres Have Higher Disease Prevention and Health Promotion Activity Levels Than Fee-For-Service Practices? September 1989, 25 p.	Julia Abelson Jonathan Lomas	5.00
No. 22	Healthy Years Equivalent: How to Measure Them Using the Standard Gamble Approach, October 1989, 22 p.	Abraham Mehrez Amiram Gafni	5.00
No. 23	Universal Precautions and the Prevention of Aids in Health-Care Workers - An Economic Analysis, October 1989, 31 p.	Susan R. Stock Amiram Gafni Ralph F. Bloch	5.00
No. 24	Incentives to be Healthy: An Economic Model of Health-related Behaviour, November 1989, 32 pp.	Stephen Birch Greg L. Stoddart	5.00
No. 25	Health-Care Technology Assessment: Ontario and Beyond, November 1989, 8 p.	David Feeny Ronald W. Wall	5.00
No. 26	Elderly Patients use of Hospital-Based Emergency Services, December 1989, 21 p.	Francois Beland Anne Lemay Lise Philibert Brigitte Maheux Ginette Gravel	5.00
No. 27	Charging for Health Care: Evidence on the Utilisation of NHS Prescribed Drugs, December 1989, 25 p.	Mandy Ryan Stephen Birch	5.00

WORKING PAPER RELEASES - 1990

No. 90-1	HSO Performance: A Critical Appraisal of Current Research, January 1990, 60 p.	Stephen Birch Jonathan Lomas Michael Rachlis Julia Abelson	10.00
No. 90-2	Economic Evaluation of Alternative Programmes of Reduced-Stay Senile Cataract Surgery, January 1990, 27 p.	Ronald Wall Stephen Birch Millie McQuillin	5.00
No. 90-3	<u>Policy Commentary</u> : Quality Assurance and Effectiveness in Health Care: An Overview, January 1990, 16 p.	Jonathan Lomas	5.00
No. 90-4	Financial Incentives and Medical Practice: Evidence From Ontario on Medical Care Utilization, February 1990, 153 p.	Roberta Labelle Jeremiah E. Hurley Thomas Rice	20.00
No. 90-5	Use of Help by the Elderly: A Gender Issue? April 1990, 20 p.	Francois Beland	5.00
No. 90-6	Producing Health, Consuming Health Care May 1990, 69 p.	Robert G. Evans Greg L. Stoddart	7.50
No. 90-7	Assessing Concepts of Quality of Care: Results of a National Survey of Five Self-Regulating Health Professions in Canada May 1990, 23 p.	Catherine Fooks Michael Rachlis Carol Kushner	5.00
No. 90-8	Ionic vs Non-Ionic Contrast Media: A Burden or a Bargain, May 1990, 15 p.	Amiram Gafni Carl J. Zylak	5.00
No. 90-9	Guide to Design and Development of Health-State Utility Instrumentation June 1990, 141 p.	William Furlong David Feeny George W. Torrance Ronald Barr John Horsman	20.00
No. 90-10	Finding Audiences, Changing Beliefs: The Structure of Research Use in Canadian Health Policy, September 1990, 29 p.	Jonathan Lomas	5.00
No. 90-11	Equity Considerations in Utility-Based Measures of Health Outcomes in Economic Appraisals: An Adjustment Algorithm, September 1990, 26 p.	Amiram Gafni Stephen Birch	5.00
No. 90-12	Words Without Action? The Production, Dissemination and Impact of Consensus Recommendations, September 1990, 26 p.	Jonathan Lomas	5.00
No. 90-13	Development and Application of a Needs-Based Methodology for calculating a Capitation Rate for a Comprehensive Health Organization November 1990, 119 p.	Stephen Birch Shelley Chambers John Eyles Jeremiah E. Hurley Brian Hutchison	20.00
No. 90-14	Episodes of Care and Long Term Trends in Individuals' Patterns of Utilization of Medical Care, November 1990, 22 p.	Francois Beland Greg L. Stoddart	5.00
No. 90-15	Using Willingness-to-Pay as a Measure of Benefits: What is the Relevant Question to Ask in the Context of Public Decision Making About Health Care Programs? December 1990, 19 p.	Amiram Gafni	5.00
No. 90-16	Turning Practice Guidelines Into Actual Practice: A Randomized Controlled Trial of Opinion Leaders and Audit/Feedback in Community Hospitals, December 1990, 19 p.	Jonathan Lomas Murray Enkin Geoffrey M. Anderson Walter J. Hannah Eugene Vayda Joel Singer	5.00

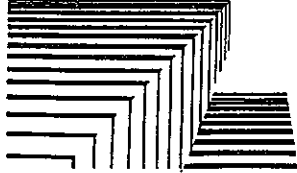
WORKING PAPER RELEASES - 1991

No. 91-1	The Effects of Co-payments in the Prescription Drug Market, January 1991, 29 p.	Jeremiah E. Hurley Nancy Arbuthnot Johnson	5.00
No. 91-2	The Search for Information - A Patient Perspective on Second Opinion, January 1991, 39 p.	Abraham Mehrez Yufei Yuan Amiram Gafni	5.00
No. 91-3	Implications of Basing Health Care Resource Allocations on Cost-Utility Analysis in the Presence of Externalities, April 1991, 25 p.	Roberta J. Labelle Jeremiah E. Hurley	5.00

CENTRE FOR HEALTH ECONOMICS AND POLICY ANALYSIS

HEALTH POLICY COMMENTARY RELEASES

<u>Number</u>	<u>Title</u>	<u>Author(s)</u>	<u>Price</u>
No. C89-1	Is the Canadian Health Care System Suffering from Abuse? August 1989, 25 p. (also listed as Working Paper # 19)	Christel A. Woodward Greg L. Stoddart	5.00
No. C90-1	Quality Assurance and Effectiveness in Health Care: An Overview, January 1990, 16 p. (also listed as Working Paper #90-3)	Jonathan Lomas	5.00
No. C91-1	Ethics and Efficiency in Health Care: Some Plain Economic Truths, January 1991, 27 p.	Anthony J. Culyer	5.00



WORKING PAPER REQUEST FORM

Please list below the working papers you wish to receive:

PAPER #	TITLE OR FIRST AUTHOR	# OF COPIES	COST

TOTAL COST

GST EXEMPT - # R119 035 988

NAME _____

ORGANIZATION _____

ADDRESS _____

POSTAL CODE _____ TELEPHONE NUMBER _____

PLEASE MAKE CHEQUE PAYABLE TO McMASTER UNIVERSITY,

AND SEND TO: Lynda Marsh
Centre for Health Economics & Policy Analysis
McMaster University
1200 Main Street West, Rm 3H26
Hamilton, Ontario
L8N 3Z5

PHONE: (416) 525-9140, Ext. 2135
FAX: (416) 546-5211
TELEX: 0618347
E-MAIL: CHEPA@SSCvax.McMaster.CA

