www.chsrf.ca / www.fcrss.ca

FINAL REPORT

CONCEPTUALIZING AND COMBINING EVIDENCE FOR HEALTH SYSTEM GUIDANCE

MAY 2005

Canadian Health Services Research **Foundation**
Fondation canadienne de la recherche sur les services de santé

...making research work
...pour que la recherche porte ses fruits

OUR PURPOSE

Vision
Our vision is a strong Canadian healthcare system that is guided by solid, research-based management and policy decisions.

Mission
To support evidence-based decision-making in the organization, management, and delivery of health services through funding research, building capacity, and transferring knowledge.

Strategy
To establish and foster linkages between decision makers (managers and policy makers) and researchers in the governance of the foundation and in the design and implementation of programs to support research, develop researchers, and transfer knowledge.

For more information on the Canadian Health Services Research Foundation, contact the foundation at:
1565 Carling Avenue, Suite 700
Ottawa, Ontario
K1Z 8R1
E-mail: communications@chsrf.ca
Telephone: (613) 728-2238
Fax: (613) 728-3527

Pour obtenir de plus amples renseignements sur la Fondation canadienne de la recherche sur les services de santé, communiquez avec la Fondation :
1565, avenue Carling, bureau 700
Ottawa (Ontario)
K1Z 8R1
Courriel : communications@fcrss.ca
Téléphone : (613) 728-2238
Télécopieur : (613) 728-3527

CHSRF FCRSS

- Final Report -

Conceptualizing and Combining Evidence for Health System Guidance

Jonathan Lomas, Tony Culyer, Chris McCutcheon, Laura McAuley, and Susan Law
for the
Canadian Health Services Research Foundation

May 2005

Additional assistance in literature compilation and review provided by Michèle Ménard, Becky Skidmore, Bonney Elliott, Robin Grant, Nicole Robinson, Joseph Rosen, and Jacqueline Tetroe

This report was prepared for the Task Force on Wait Times, Provincial/Territorial Conference of Deputy Ministers of Health with funds provided through the Canadian Institutes of Health Research. This work represents the conclusions of the authors and not necessarily those of the sponsoring or funding organizations.

Table of Contents

Main Messages

- At a basic level, the notion of evidence concerns facts (actual or asserted) intended for use in support of a conclusion.

- Most decision makers view evidence colloquially — "anything that establishes a fact or gives reason for believing something" — and define it by its relevance. Most researchers view evidence scientifically — the use of systematic, replicable methods for production — and define it by its methodology.

- Scientists' views on the role of evidence divide into those who emphasize more context-free universal truths (identified closely with "evidence-based medicine") and those who emphasize a context-sensitive role for evidence in a particular decision (identified closely with the applied social sciences).

- The appropriate methods for obtaining scientific evidence on context factors are not the same as those for obtaining scientific evidence on program or intervention effectiveness — but this makes such evidence no less "scientific."

- Scientific evidence on context can usefully be divided into evidence about attitudes, implementation, organizational capacity, forecasting, economics/finance, and ethics.

- Colloquial evidence can usefully be divided into evidence about resources, expert and professional opinion, political judgment, values, habits and traditions, lobbyists and pressure groups, and the particular pragmatics and contingencies of the situation.

- These three different forms of evidence — colloquial evidence, scientific evidence on effectiveness, and scientific evidence on context — will not combine of themselves to produce health system guidance; combining and interpreting them requires a deliberative process.

- A deliberative process is participative and often follows a period of consultation with relevant stakeholders; it entails both the eliciting and the combining of various types of evidence in order to reach an evidence-based judgment.

- There is little evidence on the effectiveness of deliberative processes, though there is much to be said in favour of them on grounds of principle.

- The design of a deliberative process is not neutral and may well influence the relative weights assigned to each of the three forms of evidence, thus influencing the extent to which guidance is "evidence-based."

- Characteristics of a deliberative process likely to ensure evidence-based guidance include consultation with all parties affected by the outcome, fair representation of scientists and stakeholders, high-quality syntheses of the scientific evidence, and skillful chairing.

Executive Summary

The worthy objective of introducing more science into decision-making started with evidence-based medicine and has spread to management of the health system and policy-making by government. As more activity in the health system is linked to the evidence-based imperative, health system stakeholders are being challenged by ambiguity surrounding the term "evidence." It is wise, then, to ask the question "what counts as evidence for health system guidance, and how can different types of evidence be combined to produce that guidance?"

The Canadian Health Services Research Foundation undertook a systematic review to examine how the concept of evidence is treated by those who produce scientific evidence, those who formulate guidance — guidelines, standards, benchmarks, targets, advisory reports, and so on — and those who make decisions. An additional review was conducted that examined deliberative processes for combining different forms of evidence to produce health system guidance.

Views of evidence

Evidence can be considered either colloquial or scientific. Outside the research community the colloquial definition of evidence dominates; that is, evidence is "anything that establishes a fact or gives reason for believing in something." Researchers tend to be more restrictive, confining the term evidence to information generated through a prescribed set of processes and procedures recognized as scientific. In this case, evidence is knowledge that is explicit (codified and propositional), systemic (uses transparent and explicit methods for codifying), and replicable (using the same methods with the same samples will lead to the same results). Decision makers are more likely to use the broadly inclusive, colloquial definition of evidence, though the evidence-based decision-making movement has engendered a greater regard for scientific forms of evidence.

When evidence is defined as science, its inclusion as part of guidance is determined through methodological tests. When it is defined colloquially, its inclusion is determined through relevance. Despite these differences, most authors covered in the review agreed that there is a need for evidence to be interpreted; the interpretation of evidence depends on who does the interpreting; and the legal definition of evidence is not very helpful for evidence-based health system guidance.

There are two distinct views on the role of science in health system guidance. One view is that science reveals universal truths (identified closely with evidence-based medicine). This view provides a glimpse of what might be achieved under ideal circumstances and creates context-*free* guidance.

Context-*sensitive* guidance, on the other hand, is embedded more strongly in the social sciences. This view of science is that evidence has little meaning or importance for

decision-making unless it is adapted to the circumstances of its application. In this view, scientific evidence on what works should be combined with scientific evidence on context.

In other words, context-free guidance indicates what we know works in general. Context-sensitive guidance shows both what works and how (or whether) it might be implemented in the specific circumstances under consideration. Thus the answer to "what is evidence?" depends on whether the objective of guidance is to create a context-free aspirational standard or context-sensitive actionable steps.

The methods for obtaining scientific evidence on context factors are just as challenging as those for obtaining scientific evidence on program effectiveness. Scientific evidence on context can be divided into evidence about implementation, organizational capacity, attitudes, forecasting, economics, and ethics. So given the complex elements of context, a multiplicity of methods might be needed to create scientific evidence.

If the goal of a given guidance-producing exercise is not the creation of a "pure" aspirational standard but the development of context-sensitive guidance, a significant challenge remains — how to combine colloquial evidence with the scientific evidence to enable a final conclusion to be reached in a way that gives due weight to each of the different forms of evidence. Technical approaches do exist, but they are unlikely to fairly balance different forms of evidence because they build in biases regarding which forms deserve more or less weight. What is needed rather than technical weighting is some form of deliberative process with appropriate representation of interests made explicit for the categories of evidence. The relative weighting of forms of evidence is left to these participants, within whatever structures or constraints are provided by the process.

Deliberative processes for combining the different forms of evidence

Our review of the literature on deliberative processes used for combining different forms of evidence led to the conclusion that little research has been done in this area beyond descriptions and assertions of "best practices." However, the descriptions do suggest that a deliberative process would be an effective tool for generating evidence-based, context-sensitive guidance, and they point to design features that are likely to be successful.

Deliberation is commonly used when there is uncertainty and the issues at stake are seen as debatable. Participative and consultative, a deliberative process "has clear objectives; is inclusive and transparent; challenges science; promotes dialogue, and directly impacts [sic] on the decision itself." Using such a mechanism elicits and combines the various types of evidence to reach an evidence-based judgment to increase the likelihood of making solid decisions.

How a deliberative process is designed undoubtedly will affect the outcome of how the evidence is considered. Important design features include consideration of topic selection, size of the group, participants, chair, types of meetings, scientific evidence inputs, framing effects, and "publicness" of the process. To get the most meaningful results, the

deliberative process includes consultation with relevant parties, fair representation of scientists and other stakeholders, high-quality syntheses of the scientific evidence, and skillful chairing.

By design, a deliberative process is not neutral and may influence the relative weights assigned to the three types of evidence — context-free, context-sensitive, and colloquial — thereby influencing the extent to which guidance is seen to be evidence-based. Thus, this process is likely to yield a judgment that is evidence-informed, better matched to the context of application, more efficiently implemented, and more widely acceptable.

Conclusion

There are differing views on what the "evidence" in evidence-based healthcare should be. This systematic review uncovered three categories of evidence: medical effectiveness research (context-free scientific evidence); social science-oriented research (context-sensitive scientific evidence); or the expertise, views, and realities of stakeholders (colloquial evidence). These views of evidence are not incompatible and each has a role to play in producing evidence-based guidance for the health system.

Under usual circumstances, there are no magic technical processes available to combine these different forms of evidence to create health system guidance. Thus each form of evidence must be entered into a deliberative process, with representation from both the scientific and stakeholder communities, if they are to be converted into a final consensus around appropriate, feasible, and realistic guidance for the health system. To date, there is little research to prove the promise held by the deliberative process. Nevertheless, there are enough investigations to suggest some design parameters that are likely to create a balanced consensus — that is, guidance that respects both scientific integrity on the one hand and its implementability in a specific health system context on the other.

Conceptualizing and Combining Evidence for Health System Guidance

The Issue

Governments have become interested in "evidence-based decision-making," particularly in healthcare, primarily due to concerns about cost-containment, quality improvement, and accountability. The worthy objective of introducing more science into decision-making started with evidence-based medicine and has spread to management of the health system and to policy-making by government. Indeed, the catalyst for this review was the commitment made in the September 2004 Canadian First Ministers' 10-year plan for healthcare to establish "evidence-based benchmarks for medically acceptable wait times." Senior civil servants recognized the challenge presented by this and other instances where there is ambiguity around the term "evidence." This concept can be used in excessively narrow or broad terms, leading to an incomplete or compromised form of evidence-based decision-making. As more and more activity in the health system is linked to the evidence-based imperative, it is wise to ask the question "what counts as evidence for health system guidance and how can it be combined to produce that guidance?"

The Approach

We undertook a systematic review (see Appendices 1-3 for details). We reviewed how the concept of "evidence" is treated by those who produce the scientific evidence, those who formulate guidance — guidelines, standards, benchmarks, targets, advisory reports, and so on — and those who make decisions (Section 1). We also reviewed articles that were about health sector deliberative processes for combining different forms of evidence to produce guidance (Section 2).

SECTION 1: What Counts as Evidence?

Two Views of Evidence — Colloquial versus Scientific

Not surprisingly, outside the research community the colloquial definition of evidence dominates. The Oxford American Dictionary captures the colloquial view as simply: "anything that establishes a fact or gives reason for believing something."[1]

When those in the clinical, management, or policy world are asked what they consider to be "evidence," they respond with a complex mélange of both scientifically verifiable and locally idiosyncratic types of information — a "colloquial" interpretation.[2-8] They "draw on multiple sources and define evidence broadly."[4] Clinical or program effectiveness data competes with expert assertion, cost-benefit calculation sits alongside political acceptability, and public or patient attitude data are combined with the vivid traces of personal encounter. "What ministers call 'evidence' is what they get from their constituents at their Saturday surgery."[6,9] In bringing evidence to bear on guidance, these decision makers are "sensitive to both scientific rationality and the local rationality of the workplace."[8]

Nevertheless, there is clearly some degree of defensiveness among decision makers about overly frequent recourse to non-scientific forms of evidence; the evidence-based decision-making movement has engendered a greater regard for the more scientific forms of evidence. For instance, after clearly documenting through direct observation the extensive use of both colloquial and scientific types of evidence by a hospital drugs and therapeutics committee, investigators noted that the guidance from the committee was "written so as to account for the decision in terms of scientific rationality … rather than the local rationality that was actually employed." They go on to comment that "this is not a duplicitous activity, but reflects how members of the medico-scientific community have been taught to account for their activities."[8]

The research community's view of evidence, although not uniform, is narrower and generally restricted to information generated through a prescribed set of processes and procedures recognized as scientific.[10-14] In this case the various tenets from philosophy of science determine what is evidence and can be summarized as knowledge that is:

- explicit (that is, codified and propositional);
- systematic (that is, uses transparent and explicit methods for codifying); and
- replicable (that is, following the same methods with the same samples will lead to the same results).

One author summarizes it this way: "Evidence … can be defined as information or facts that are systematically obtained, i.e. obtained in a manner that is replicable, observable, credible, verifiable, or basically supportable."[15]

"When evidence is defined as science its inclusion as part of guidance is determined through methodological tests. When it is defined colloquially its inclusion is determined through tests of local relevance."

Hence, when evidence is defined as science its inclusion as part of guidance is determined through methodological tests. When it is defined colloquially its inclusion is determined through tests of local relevance. "Evidence could therefore be seen to possess two facets: the scientific, factual facet and the more personal, contextual facet."[16]

Common Themes from either Colloquial or Scientific Views of Evidence

Despite the differences we found there are at least three areas of commonality in these two views of evidence. These are a) there is a need for evidence of any kind to be interpreted; b) the interpretation of evidence depends upon who is doing the interpreting; and c) the notion of evidence used in the law is not very helpful for producing health system guidance.

Most but not all authors agree that "evidence itself does not recommend its own interpretation."[17] A single piece of evidence, whether of the scientific or colloquial type, is rarely complete enough to create the guidance by itself. This is for a number of reasons:

- the inherent uncertainty that nearly always accompanies evidence:

"the occurrence of definitive studies is comparatively rare"[17]

"evidence rarely attains absolute certainty … there is no such thing as *'the'* evidence"[10]

- the complexity of the decisions for which guidance is being offered and, therefore, the likelihood that evidence will never be comprehensive:

  "evidence of the optimal combination of agents to treat Alzheimer's disease would require 127 randomized controlled trials, 63,500 patients and 286 years"[18]

- the need for actors to create the meaning and interpret the evidence before placing it as *knowledge* in guidance:

  "knowledge is a product of social, historical, cultural and political processes"[19]

  "like all other forms of human practice, research itself necessarily relies on judgment and interpretation: it can never be *governed*, but only *guided*, by methodological rules"[20]

"Evidence is inherently uncertain, dynamic, complex, contestable, and rarely complete."

This implies that evidence is inherently uncertain, dynamic, complex, contestable, and rarely complete. Therefore, for it to become part of guidance, some form of deliberative process is likely required to assess the relative merits and limitations of the evidence in light of the issue at hand. This deliberative process must be able to combine and interpret the population of evidence (however defined) for the purpose intended. We return to this important point in the sections below.

Next, maintaining a common understanding of what evidence is becomes increasingly difficult as further interest groups or stakeholders are added to any guidance-producing process. Conversely, the more homogeneous the group — in professional background and level of responsibility — the less tension and disagreement exists on what constitutes permissible evidence. Walshe and Rundall point out that:

> "The clinical culture is highly professionalised, with a formal body of knowledge which is shared by all members of the profession and which acts as frame of reference for intraprofessional dialogue and debate ... In contrast, health care managers are a highly diverse group drawn from different professional and disciplinary backgrounds, and they often lack even a shared language or terminology with which to describe and discuss what they do."[21]

"This implies that it will be increasingly difficult to negotiate and adopt a common definition of 'evidence' for guidance at ever higher levels of health system governance."

This implies that it will be increasingly difficult to negotiate and adopt a common definition of "evidence" for guidance at ever higher levels of health system governance, as increasingly heterogeneous groups become concerned about and/or affected by the outcome. Rosen, for instance, questions the assumption of many of

evidence-based medicine's supporters that "research can be applied equally to all forms of healthcare decisions, despite differences between the individual patient focus of evidence-based medicine, the population perspective taken by most health planners and the organizational viewpoint of most hospital managers."[3]

Finally, legal definitions have only a limited (if any) role in determining what is permissible as evidence within guidance for the health system.[22] For the more generous colloquial view, the law's exclusion of "hearsay" hurts; for the more definitive scientific view, the law's reliance on individual testimony misleads — "law relies on evidence of the instance; health care relies on evidence of the generalisable."[22] In addition, the law usually relies on scrutiny of what happened in a past event (benefiting from 20/20 hindsight); healthcare guidance is about shaping future events. Thus legal definitions of evidence seem disconnected from the reality of constructing evidence-based guidance for the health system: "the difference in the way in which evidence is approached creates a cultural divide between medicine and the law, a conflict with its roots in different epistemologies of evidence."[22]

"Legal definitions of evidence seem disconnected from the reality of constructing evidence-based guidance for the health system."

The individual case focus of legal evidence — "the instance" — has little appeal for those constructing generalizable guidance for the health system. At the other end of the spectrum, however, does scientific evidence offer the absolute or universal truths that would readily facilitate the construction of useful, widely applicable, evidence-based guidance for the health system? To explore this question we set aside the colloquial view of evidence and review those perspectives we found in the literature on the role of scientific evidence in guidance.

Two Views on the Role of Scientific Evidence — Context-free versus Context-sensitive

Two distinctly different views emerged of the role of science in health system guidance.[23,24] One view — emanating from the roots of "science for guidance" in evidence-based medicine — is that science does indeed reveal universal truths, that is, its role is independent of context. Aspirin works for any individual, sutures hold together anyone's cut, MRI scanners reveal any suspicious mass.

In this view science answers the question "can it work?" without paying attention necessarily to "will it work?" or "is it worth it?"[12] The role of science is somewhat detached from, and unconcerned with, its application to specific circumstances — that is the role of various deliberative processes in the clinical, managerial, or political realms. The science provides a glimpse of what might be achieved under ideal circumstances; it creates ***context-free guidance*** to which the "real world" can aspire.

A second view is embedded more strongly in the social sciences and is offered very much in counterpoint to the views of evidence-based medicine proponents. This view of science is that evidence has little meaning or importance for decision-making unless it is adapted to the circumstances of its application: "evidence itself does not recommend its own interpretation."[17] Confirmation of effectiveness for an expensive new drug for HIV/AIDS means something quite different to New Brunswick parliamentarians than it does to healthcare workers in Botswana.

Science cannot inform potential action without helping to also ascertain what is feasible action. Science creates evidence on both the "what" of effective therapeutic or program possibilities for a particular patient or group and the "how" of achieving that possibility for all or parts of a system when faced with specific professional cultures, financing regimes, public attitudes, patient responses, and so on. This view is well captured by Greenhalgh:

> "The scientific tools for dissecting the decision-making process are not those of the clinical epidemiologist but those of the sociologist, the psychologist, the qualitative researcher … It is time that extremists on both sides recognized that there is a science — albeit a *social* science — to the art of medicine, and that through science (and specifically through decision science) we can systematically explore, validate and refine the art of medicine and integrate it with our diagnostic, therapeutic and epidemiological expertise."[24]

Indeed, social science can be put to use to answer not only questions like "will it work in this context?" but also questions about the processes that might yield answers to such questions. Thus it can be used to inform the design of decision-making procedures that are suitable in defined contexts and for defined purposes and to answer questions like "is the use of an algorithm for assigning patients a priority rating in a waiting list likely to be more acceptable to the community than some alternative?" In this wider view of what scientific questions might embrace, evidence from the social sciences becomes integral to, not separate from, deliberative processes for creating ***context-sensitive guidance*** on feasible actions.

For the provision of guidance to the health system these two views imply quite different answers to the question "what counts as evidence?" Science for context-free guidance is largely restricted to the evidence on clinical or program outcomes and is exemplified by the growing availability of meta-analyses and other forms of systematic review in researcher-driven databases such as the Cochrane (www.cochrane.org) or Campbell (www.campbellcollaboration.org) collaborations. To create context-sensitive guidance, context-free science needs to be integrated into the science on "local" variables such as public attitudes, patient preferences, professional proclivity, managerial capacity, economic feasibility, geographic location, and so on in the search for consensus around what might be achievable rather than what might be a universal clinical truth.

"The role of science in the case of context-free guidance is to indicate what we know works in general; in the case of context-sensitive guidance it is to illuminate both what works and how (or whether) it might be implemented in the specific circumstances under consideration."

The role of science in the case of context-free guidance is to indicate what we know works *in general*; in the case of context-sensitive guidance it is to illuminate both what works and how (or whether) it might be implemented in the specific circumstances under consideration. Thus the answer to the question "what is evidence?" depends on whether the objective of guidance is to create a context-free aspirational standard or context-sensitive actionable steps.

In the case of context-free guidance, driven very much by the view of clinical research methods, evidence tends to be narrowly defined, as in this definition of evidence-based medicine: "the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients."[25] It is also generally defined with a clear hierarchy of greater and lesser forms of science: "identifying the best evidence means using epidemiological and biostatistical ways of thinking."[26]

Creating context-sensitive guidance is more likely to generate encompassing definitions of evidence such as the two outlined in Box 1.

Box 1
Example Definitions of Evidence from Guidance-producing Organizations

World Health Organization — Europe
"Findings from research and other knowledge that may serve as a useful basis for decision-making in public health and health care."[27]

U.K. Government Policy Hub
"The raw ingredient of evidence is information. Good quality policy making depends on high quality information, derived from a variety of sources — expert knowledge; existing domestic and international research; existing statistics; stakeholder consultation; evaluation of previous policies; new research, if appropriate; or secondary sources, including the internet. Evidence can also include analysis of the outcome of consultation, costings of policy options and the results of economic or statistical modeling."[28]

Some bemoan the epistemological warfare that has broken out between these two views of science, with some disciplines seeking preferred status for evidence produced by their own methods: "the rules of evidence appear to be tied to disciplines not projects"[19] or "the literature around EBM [evidence-based medicine] has an abundance of clear hierarchies and methods for grading the quality of evidence, but it deals less with the question of evidence versus non-evidence."[29] The weight of writing in this area, however, argues a "horses for courses" approach with "methodological pluralism rather than continuing paradigmatic antagonisms, seeking complementary contributions from different research designs rather than epistemological competition"[30] or "different stages of the policy process [possibly calling] for different types of evidence."[31]

Clearly "context," and its relevance and importance for the applicability of a particular piece of guidance, is the fulcrum around which swing these two views of the role of science. On the one hand contextual factors are seen as the domain of political and other forms of judgment; science is restricted to defining the universal truth of "what works:"

> "the *philosophical-normative* orientation towards what constitutes evidence is unconstrained by context."[23]

On the other hand, context is seen as both generating scientific questions in its own right (for example, "is this outcome measure a valid measure of what we are seeking?"), as well as itself being as open to scientific investigation as are therapeutic or program effectiveness:

> "the *practical-operational* orientation to what constitutes evidence is context-based, with evidence defined with respect to a specific decision."[23]

Context, however, is an elusive concept and begs a question of the existence of some boundaries within which science investigation can occur. Without such boundaries scientific evidence on context may quickly drift out of the research domain and into the colloquial view of evidence. For those wishing to adopt a "context-sensitive" role for science in guidance, the dimensions of context evidence, and the scientific methods used to elucidate them, are needed. This is the topic of the next section.

Dimensions of Context and Scientific Evidence

For some, trying to capture the concept of context is like catching clouds: "the context in which health care practice occurs can be seen on one level as infinite as it takes place in a variety of settings, communities and cultures that are all influenced by (for example) economic, social, political, fiscal, historical and psychosocial factors."[34] Indeed, some authors appear to implicitly assume that experiential and context evidence are too elusive to be captured by means of research: "this paper outlines … four types of evidence … research, clinical experience, patient experience and information from the local context."[10]

Many others, however, see a broad range of factors that can be illuminated with research methods, beyond the health effects of clinical programs or interventions: "For … improvements in public health decision-making to occur, the systematic evaluation of research on potential interventions ***and*** the contextual factors, such as acceptability to stakeholders and implementation constraints, need to be considered together."[7]

Box 2
Different Categories or Dimensions of Evidence

Defined by method of collection:
- observational, experimental, extrapolated, experiential[7]
- experimental, quasi-experimental, survey, administrative, qualitative, economic, ethical/philosophical, systematic review[32]
- legal, epidemiologic[22]
- quantitative, qualitative[33]
- clinical epidemiology, decision science[24]

Defined by general purpose:
- problem identification/description, effectiveness, implementation[5]
- culture, leadership, measurement[34]
- philosophic-normative, practical-operational[23]

Defined by source:
- research, organizational capacity, political acceptability[31]
- research, clinical experience, patient experience, local context[10]
- scientific, theoretic, practical, expert, judicial, ethical[35]

The variety of evidence dimensions or categories proposed in the articles of this review are summarized in Box 2. Three general approaches emerged. One group focused on categories according to the *method of collection* for the evidence — experimental or survey, for

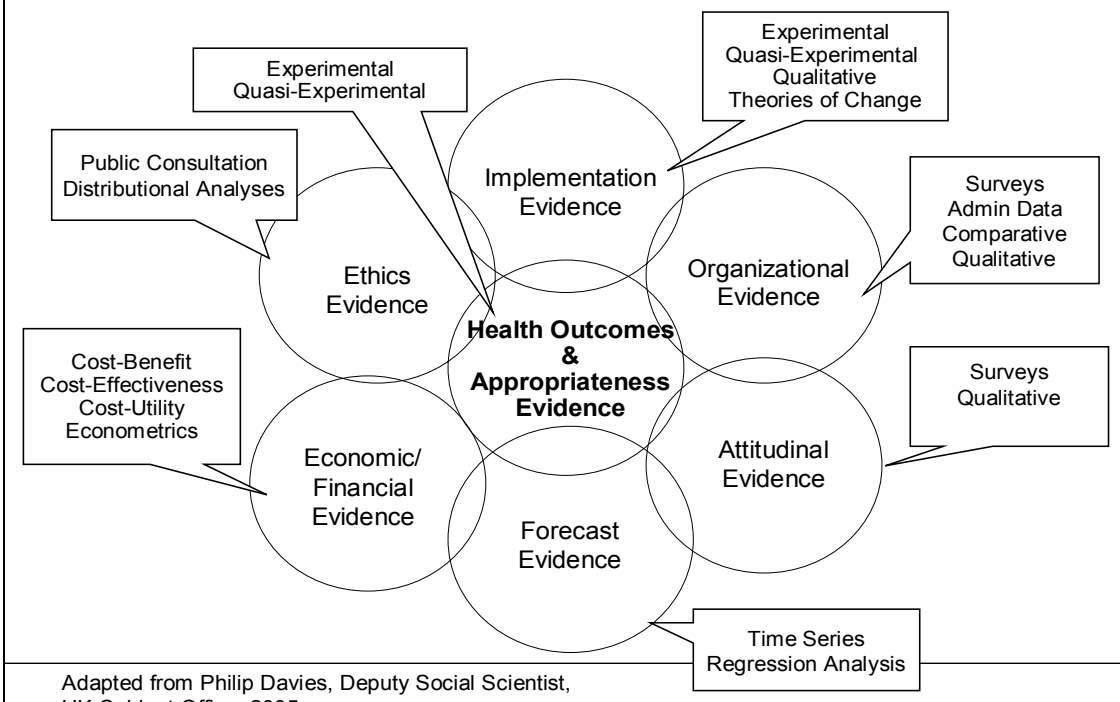
instance. Another focused on the *general purpose* to which the evidence would contribute — such as identifying a problem versus measuring the effectiveness of an intervention. The third emphasized *source*, usually distinguishing research from colloquial forms of evidence such as clinical experience.

"Given the potentially diverse elements of context, it is not surprising that a multiplicity of methods might be needed to create scientific evidence across even a few of the dimensions."

Given the potentially diverse elements of context, it is not surprising that a multiplicity of methods might be needed to create scientific evidence across even a few of these dimensions. As one author noted after describing ethical, judicial, expert, practical, and theoretic dimensions to providing guidance, "these dimensions tend to operate within different frameworks that seek to answer different questions in different ways, based on different evaluative criteria"[35] — the "horses for courses" view of methodology outlined earlier.

Figure 1 summarizes and makes explicit the various dimensions of context that emerged from the literature. It keeps separate the dimensions of context and the methods used most often to generate scientific evidence on each one.[36]

Figure 1: Types of scientific evidence (and their methods) for context-sensitive health system guidance



Adapted from Philip Davies, Deputy Social Scientist, UK Cabinet Office, 2005

At the core is health outcomes and appropriateness evidence for a particular program or intervention, obtained through experimental and quasi-experimental methods. The surrounding context evidence is placed into six categories: implementation, organizational, attitudinal (which may be attitudes of patients, the public, professionals, or other stakeholders), forecasting, economic, and ethics, each identified with its most common scientific methods.

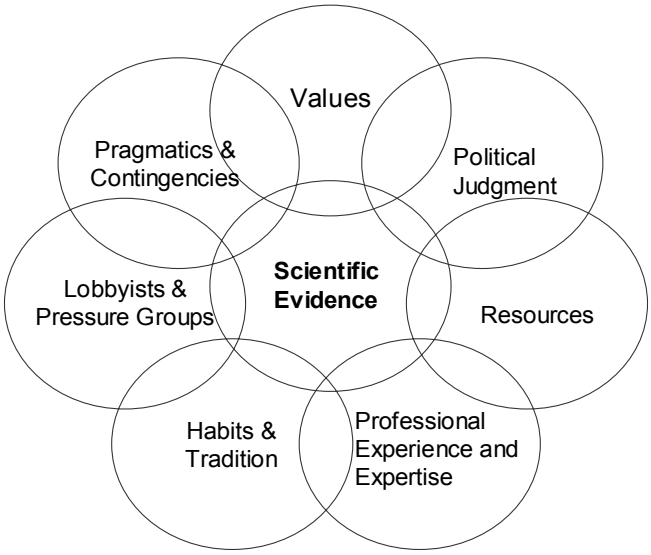
While there are obviously an infinite number of ways to characterize context dimensions relevant to health system guidance, the approach in Figure 1 offers a parsimonious summary of those encountered in the literature under review. It also underlines the fact that research methods are available to ensure that at least some scientific evidence is possible to inform each of the dimensions.

What to do about Colloquial Evidence?

"Colloquial evidence … can complement, or substitute for missing scientific evidence on context."

Thus far we have outlined three types of evidence: scientific evidence on effectiveness ("what works"), scientific evidence on context ("how or whether it works" in a particular circumstance), and colloquial evidence. For scientific evidence, Figure 1 provides a schematic of what might be included and how it might be obtained. For colloquial evidence, however, there is a tendency to lump it all together as a single entity. Figure 2 provides a schematic for disaggregating categories of colloquial evidence that can complement, or substitute for, missing scientific evidence on context.[2]

Figure 2: Types of colloquial evidence available for context-sensitive health system guidance



Adapted from Davies, 2005

In Figure 2 all the forms of scientific evidence presented in Figure 1 are collapsed into a single central circle, now surrounded by various forms of colloquial evidence. Although a parsimonious categorization could restrict these forms to *values* (including both values and political judgment), *practical operational considerations* (including resources, professional experience/expertise, and habits/traditions), and *interests* (including lobbyists/pressure groups and pragmatics/contingencies), they are divided into their sub-components for greater clarity in Figure 2. Qualification for inclusion of these forms of evidence is, by definition, not methodologically determined but driven by relevance and the sources available to the guidance producers.

If the goal of a particular exercise is not the creation of a "pure" aspirational standard but the development of context-sensitive guidance, a significant challenge remains — how to combine this colloquial evidence with the scientific evidence to enable a final conclusion to be reached in a way that gives due weight to each of the different forms of evidence.

There are those that have created complex technical approaches to combining these different types of evidence, using assigned relative weights.[37] This technical, algorithmic approach "tends to bury under a series of assumptions many value judgments that may or may not reflect those of the broader population of users and payers."[38] It also does not recognize that "knowledge from personal experience and from new research evidence must each be evaluated in its own terms, and then combined in some way that takes account of their distinctive characteristics as sources of knowledge."[20]

Thus what is more commonly adopted is some form of deliberative process rather than technical weighting, with appropriate representation of interests made explicit for the categories of evidence. The relative weighting of forms of evidence is left to these participants, within whatever structures or constraints provided by the deliberative process. It is this kind of approach which is proposed by Klein:

> "Given conflicting values, the process of setting priorities for health care must inevitably be a process of debate. It is a debate, moreover, which cannot be resolved by an appeal to science and where the search for some formula or set of principles designed to provide decision-making rules will always prove elusive. Hence the crucial importance of getting the *institutional* setting of the debate right ... the right process will produce socially acceptable answers — and this is the best we can hope for."[39]

"Deliberative processes … are not neutral in their design."

In other words, deliberative processes — or "the institutional setting of the debate" — are not neutral in their design. Some will favour one form of scientific evidence over another, others will favour colloquial evidence over scientific evidence or vice-versa. Evaluating the literature on different deliberative processes, and the forms of evidence they appear to favour or impede, is therefore the subject of the concluding part of this paper.

SECTION 2: Deliberative Processes for Combining Forms of Evidence

Why Use Deliberative Processes?

Deliberative processes are mechanisms for *eliciting and combining evidence* — "a more fundamental means by which the public can influence the generation of data and the derivation of the policy options as well as discussing acceptable decisions, thus, taking account of public as well as expert knowledge."[44] In addition, deliberative processes have been urged on other grounds (see Box 3). In the literature reviewed here some saw their rationale in terms of *democratic governance*: "a move away from the unilateral, technocratic, regulatory model of risk management and decision-making toward more inclusive, democratic, non-regulatory processes, reflecting the democratic ideal that people should be involved in their own governance."[41] Most pertinent to the provision of health system guidance, however, is the claim that deliberative processes increase the likelihood of achieving "sound and acceptable decisions."[49] Daniels, a philosopher of health and healthcare, argues for processes that "account for reasonableness" so they have a moral authority greater than that customarily attached to market or bureaucratic processes.[50]

"Deliberative processes increase the likelihood of achieving sound and acceptable decisions."

What is a Deliberative Process?

Deliberation is commonly seen as desirable whenever the issues at stake are debatable. Deliberative processes stress "an integration of technical analysis and stakeholder and lay public deliberation, contrasting with the traditional 'top-down' or 'bureaucratic-rationalistic' policy orientation."[44] It is similar to the "co-operative discourse" model of decisions used in Germany and Switzerland (for example, Schneider *et al*.,[51] Webler[52]). It is a "participatory process that has clear objectives; is inclusive and transparent; challenges science; promotes dialogue between all parties; promotes a consensus about the potential decision, and directly impacts [sic] on the decision itself."[44]

Box 3
Arguments for Deliberative Processes

Eliciting and Combining Evidence
- To bring evidence together and weigh it all up[40]
- To determine what risks are acceptable[41]
- Exposing and/or resolving conflict over evidence (and/or over values)[42]
- To facilitate interdisciplinary dialogue between "experts"[41]
- To reveal "evidence" not otherwise available[43]
- To be seen to be taking care over evidence[44]
- To enable quality to be addressed[44]

Democratic Governance
- Democracy[42]
- Involvement of people in their own governance[42]
- Accountability[45]
- Transparency[40]
- A check on the partiality of "experts"[46]
- To create a learning public[42]
- To embody the public's values[47]

Creating Acceptable Guidance
- To get potential opposition inside the tent[43]
- To let all stakeholders have their say[42]
- To be plausible to the public and professionals[42]
- To maintain public and professional commitment and confidence[9,47]
- To embody implementation issues of specific contexts[48]

Box 4
Main Conditions for a Deliberative Process

Participation

- Evidence from more than one expert discipline is involved
- Evidence from more than one profession is involved
- Stakeholders have conflicting interests
- Resolving technical disputes
- Evidence may be scientifically controversial
- Evidence gathered in one context is to be applied in another

Wider social and cultural issues are involved

- Issues of outcome and costs that go beyond the narrowly scientific
- Uncertainty exists and risks that need to be assessed and valued
- Other social and personal values
- Issues of equity and fairness
- Issues of implementability and operational feasibility
- Wide public and professional "ownership" is desired

Box 4 lists the principal features of a decision context that is likely to warrant using a deliberative process.

The factors embraced for deliberation include the three forms of evidence we outlined in Section 1: scientific evidence on effectiveness; scientific evidence on context features (such as a survey to generate "a profile of perceived health care needs and problems facing the community and the stakeholders who respond to the survey"[53]); and colloquial evidence. Others have referred to the colloquial evidence brought to the deliberative table by the participants as interpretive data: "data reflecting the perceptions of local health care leaders…"[53]

Scientific evidence on effectiveness is often summarized in the form of a narrative review, systematic review, or meta-analysis; scientific evidence on context, including values, might be gathered by controlled experiments, which in turn may be summarized; and colloquial evidence is often gathered through consultative processes including social surveys, public meetings, and the hearing of witnesses, as well as from participating deliberative panellists.

Deliberative processes are not the same as consultative processes. A prominent example of a consultative process is the Oregon priority-setting exercise initiated in 1989.[45] This entailed 47 community meetings, 12 public hearings, and 54 panel meetings for healthcare providers, the information from which was then delivered to a committee (the Oregon Health Services Commission) for prioritization of procedures. Thus many were *consulted* but relatively few *participated*.

Some of the questions that need to be asked prior to the creation of a deliberative process are set out in Box 5.

Box 5
Checklist Questions for Deliberative Process

- Is the topic or issue best dealt with through a deliberative process?
- What groups or people are best involved in the process?
- How is the process of gathering and disseminating evidence best conducted?
- How open ought the processes and deliberations to be?
- How much control over the types of evidence sought and presented ought the members of the deliberative process to have?
- To whom ought the parties in the deliberation be accountable and how can they be held accountable?

Two Examples of Deliberative Processes

A good example of a deliberative process in which all these features are present is the methods used by the National Institute for Health and Clinical Excellence (NICE) when evaluating healthcare technologies for England and Wales. These methods include formal consultations and invited commentaries, reviews, multi-party representation in the (large) deliberative committee which hears witnesses, appeal possibilities, and support groups using consensus methods on controversial issues of value.[54]

Although matters of value and fairness are often suitable candidates for resolution by a deliberative process, it does not follow that they invariably are. NICE offers examples of both a deliberative process and an algorithmic approach embodied within it, which was itself reaffirmed through a further specific deliberative process, illustrated in Box 6.

Box 6
Algorithms versus Deliberative Processes in NICE

Algorithm
NICE recommends the use of Quality-Adjusted Life-Years (QALYs) as the main outcome measure in the economic appraisals that are presented to its multi-disciplinary and multi-professional appraisals committee. The particular form of QALY recommended is the EQ-5D, which is an algorithm embodying various health state characteristics (5) measured on a three-point scale and added together. The guidance explicitly states *"an additional QALY should receive the same weight regardless of the other characteristics of the individuals receiving the health benefit."*[55]

Deliberation
NICE has also referred an aspect of this algorithm to its citizens council, a form of consensus group to engage in a deliberative approach regarding the weighting (if any) to be given to older people. Their recommendations to the NICE board included: *Overall, the majority of us on the Citizens Council [22 out of 29] felt very strongly that no judgement should be made about being more generous to certain age groups because of the social roles those age groups tend to fulfil.*[56]

The Texas Department of Health and Mental Retardation has utilized a deliberative process to generate clinical guidance.[47] It also created a subtle and more complex combination of deliberative procedures to establish the "bundle of benefits" for publicly funded patients. This included a senior staff meeting to determine broad categories of benefit (such as employment-related, housing-related, substance abuse-related); a series of multi-stakeholder meetings and teleconferences to identify, with the help of national experts, the scientific evidence on practices for these categories; and a state-wide consensus conference of 200 stakeholders meeting in plenary and subsequently in a smaller select but representative group of 40 to unearth the colloquial evidence, followed by an implementation and testing stage at selected sites reviewed by relevant stakeholder groups.

"The deliberative process yields a judgment – about what is likely to be achieved, in what ways, for whom, how worthwhile it is, for how long, and at what cost."

In the final analysis, the deliberative process yields a judgment —— about what is likely to be achieved, in what ways, for whom, how worthwhile it is, for how long, and at what cost (in terms of the resources used that would otherwise have been employed in other ways to achieve other good things). Done according to the principles of its advocates, the

resultant judgment should be more comprehensively "evidence-informed," better matched to the context of application, more efficiently implemented, and more widely acceptable.

Design Issues

There has been little evaluation of the best way to design deliberative processes. Therefore, in what follows we outline some of the opinions to be found under key issues. In particular, we emphasize design features that address how to appropriately combine scientific and colloquial forms of evidence in order to achieve consensus around evidence-based health system guidance.

Topic selection

A wide variety of methods is used to select the topics for deliberative processes: government departments, with or without "horizon scanning" or other sub-group input; professional groups; principal investigators; and civil servants. When on familiar territory expert groups are tempted to be unnecessarily exhaustive. For example, Black *et al*.[57] note that just 2.7 percent of the three hundred theoretical indications for coronary angiography account for more than half the cases that actually occur in clinical practice. In such a case, adding further "topics" — and deliberation time — is subject to severely diminishing returns.

Size of group

There is no consensus on the matter of ideal size for a guidance-producing group. McGlynn *et al.*[58] report sizes of consensus panels of between nine and 18 in an international review. Black *et al.*[57] suggest a norm of 10. NICE's citizens council and its appraisals committee can have as many as 30 participants, sometimes including non-members giving evidence. Large size is believed to reduce active participation by all, although it does potentially enlarge the range of viewpoints when dealing with complex matters. As mentioned earlier, a large heterogeneous group can give rise to conflict and difficulty in achieving consensus on guidance.[57]

Participants

Distinctions can be made between the members of the deliberative panel itself — those with a right to be consulted or to make comments, those who will bring colloquial evidence to bear — and those who might represent the scientific evidence. In all the cases reviewed by McGlynn *et al.*[58] both "scientists" and "lay people" were represented. Participants with high social status tend to dominate others.[57] Clinical specialists tend to favour their own specialty.[57] Particpants may be invited to represent an interest group officially or they may sit as individuals who, although they come with a background from a particular interest group, represent only themselves. A failing noted for some consensus conferences is the arbitrariness in the selection of panel members[59] leading to selection bias.

"…the crucial selection factor is the inclusion of participants skilled in, and able to advocate for consideration of, scientific evidence alongside 'lay' participants."

For the purposes of evidence-based guidance panels, the crucial selection factor is the inclusion of participants skilled in, and able to advocate for consideration of, scientific evidence alongside "lay" participants. Under the assumption that stakeholders will bring colloquial evidence to bear on the topic, the role of scientists is to ensure that this form of evidence does not trump scientific evidence unless there is extensive uncertainty or dispute about the validity of the science.

Chair

The chair is generally selected/appointed by a due process external to the deliberative panel itself, though not invariably.[58] It seems clear that large groups and complex processes will demand a high order of chairing skill. Black *et al.*[57] report that there is little work on the characteristics of a good facilitator or its impact on group decision-making.

Types of meetings

While some meetings are face-to-face (as implied above) others can be virtual. For example, some panels use the Delphi method in which participants never meet directly but respond to questionnaires, usually on a Likert scale, and approach consensus through iteration.[57] In face-to-face meetings various degrees of structuring are possible. For example, the Nominal Group Technique involves the collection of ideas from each participant and then their systematic playing back to the group by a facilitator in such a way as to ensure that all ideas are openly addressed. Structures that explicitly separate consideration of different types of evidence increase the degree to which final guidance can be traced back to the scientific evidence.[60]

Scientific evidence inputs

This can be presented in a variety of ways as background material and/or as the focal point of attention. Writing in 1990, McGlynn *et al.*[58] comment that systematic attempts to summarize or synthesize scientific evidence are rare, but we conjecture that this observation would no longer be made today — not at any rate for clinical topics. Indeed, many deliberative processes now take care to separate the scientific from the colloquial forms of evidence input. Presentation formats include narrative and systematic reviews, meta-analyses, informal reviews, textbooks, collections of selected papers, abstracts, expert and other oral presentations, specially commissioned primary research, and modelling exercises that extend the primary research results.

Framing effects

Although it is well-known from experimental psychology and experimental economics that the way in which questions are put can radically affect the response, there is little evidence that this matter has been explored in the context of deliberative processes. Black *et al.,*[57] however, demonstrated the need to attend to framing effects by showing that a

group's interpretation of the same data presented in four different ways varied quite substantially.

Publicness of the process

Most processes seek a balance between transparency/openness and the privacy usually felt necessary for discussion to flow freely and for the participants to have the confidence to admit ignorance, enquire openly, try out wild ideas, and so on. Temporary dissent is a common part of a process whose ultimate outcome is consensus; not all parties (and this includes scientists as well as "lay participants") are equally able to engage in principled discussion in public; some "play to the audience;" and team-building is, anyway, not an activity most helpfully done in the public gaze. For example, the Health Council of the Netherlands is transparent on the procedural side, while the actual deliberations are closed.[40]

Do Deliberative Processes Work?

"Much of the literature on deliberative processes has been and continues to be essentially advocacy rather than reports of the effectiveness of well-defined processes."

What makes for an effective deliberative process? Much of the literature on deliberative processes has been and continues to be essentially advocacy rather than reports of the effectiveness of well-defined processes (for example, Gibson *et al*.[61]). The evidence is understandably less rigorous than that for more narrowly defined "scientific" matters of cause-and-effect. It is mostly judgmental.

In comparing a deliberative process with other methods Kim[42] writes:

> "It clearly has some advantages. 1. It can solve the problem of political legitimacy… 2. It can act as a check to the partiality which expert groups in biotechnology and ethics may have. It promotes the dialogue between experts … and between experts and ordinary citizens. 3. It enables us to make informed and responsible decisions. 4. It results in education of citizens' preferences."

Unfortunately, while all these outcomes "can" or "may" be true, they are not themselves evidence and have to be suspect coming as they do from a manifest enthusiast for the method. Other authors are more sceptical, arguing consensus processes place too much emphasis on group process, public relations, and "back room politics."[60]

Some conclusions that seem a little more factual are based on authors' experiences with deliberative processes in a variety of topic areas:

- "Having participated in the research process, panel members knew well the purpose, strength and limitations of the new classification tool. … 'bathing' was eliminated from the classification tool … [and] … members, credible to their peers, explained to colleagues the rationale for deletion."[43]
- "Stakeholders are more likely to accept and implement a decision they have participated in shaping."[41]

- "…selection bias, particularly with respect to the choice of questions and panelists — remains a significant threat to the credibility of the consensus process."[59]
- "By the time the Alberta research was completed, the two key stakeholder associations were ready to endorse the classification tool." [43]

McGlynn *et al*.[58] conducted an examination of the processes of consensus conferencing in the U.S., Canada, Denmark, Finland, the Netherlands, Norway, Sweden, Switzerland, and the U.K. To maximize the use of the evidence they recommended the more frequent use of systematic literature synthesis, meta-analysis, more formal voting (polling) systems to identify disagreements and help their resolution, and more time for deliberation and reflection.

One study specifically investigated the role of evidence in a deliberative consensus process. It showed that if attainment of consensus during the process was defined as the objective, then the deliberative process worked best in those areas where scientific evidence was available and presented.[60]

NICE's deliberative processes have been subject to a World Health Organization investigation,[62,63] which "largely affirms NICE as a leading organisation internationally in the use of evidence about clinical and cost-effectiveness to inform decisions in the health sector."[62] NICE's procedures represent an institutionalized set of deliberative arrangements at the level of a country. They seem to work — in the sense that both the procedures and conclusions reached by using them have commanded respect and broad assent, but that is a weak test. It underlines the difficulty, however, of answering questions like "does it work?" in the context of inherently complex and controversial decision situations.

We conclude this section with a summary in Box 7 of some features of deliberative processes that we conjecture to be

Box 7
Conjectured Success Factors for Deliberative Processes

Background Resources
- Well-conducted scientific research
- Well-defined questions to be answered
- Well-resourced support staffing
- Meta-analyses and systematic reviews of scientific evidence
- Availability of research into the public's views on ethical and other value issues

Rules and Expectations
- Clear timelines for submission and consideration of evidence
- Separated consideration of scientific and colloquial evidence
- Clear deadlines for decisions
- Time for study, discussion, and reflection
- Opportunities for all interested parties to comment during the process
- Ability for members to request further information and take oral evidence
- Opportunity for appeal if process has been flawed or decision appears unreasonable

Participant Selection
- Participant selection adequately represents expertise in the relevant scientific evidence
- Participant selection adequately represents breadth of colloquial sources of evidence
- Participation of recognized and respected people from the major communities of interest
- Opportunities for all affected parties to be represented

conducive to success. These are not, however, well-tested evidence-based conclusions but are drawn from the limited literature that has assessed such processes and descriptions (such as Bal *et al*.,[40] Culyer,[54] Lomas *et al*.[60]). Perhaps the best overall summary comment on the promise of deliberative processes for health system guidance comes from an author who engineered a deliberative process with the specific intent of combining the scientific and colloquial forms of evidence in the area of air pollution: "… reviews of oxygenates in gasoline demonstrated the effectiveness of combining scientists and stakeholders in a manner that was able to maintain the integrity of the science while addressing stakeholder concerns and assuring stakeholder 'buy-in.'"[41]

Conclusion

This paper has addressed two highly related topics: "what counts as evidence in the provision of health system guidance?" and "how is evidence (however defined) best combined to create health system guidance?"

In an attempt to defuse some of the polemical and rhetorical debate around concepts of evidence the paper provides a framework, based on a systematic review of the literature, that gives credence and value to three forms of evidence:

- medically oriented effectiveness research (context-free scientific evidence);
- social science-oriented research (context-sensitive scientific evidence); and
- the expertise, views, and realities of stakeholders (colloquial evidence).

We argue that each form of evidence has a role to play in producing context-sensitive, evidence-based guidance for the health system.

However, we also note that under usual circumstances there are no magic technical processes available to combine these different forms of evidence to create health system guidance. Thus each form of evidence must be entered into a deliberative process, with representation from both the scientific and stakeholder communities, if they are to be converted into a final consensus around appropriate, feasible, and realistic guidance for the health system.

Our review of the literature on deliberative processes, with special attention to their capacity for appropriately combining forms of evidence, led to the conclusion that little research has been done in this area beyond descriptions and assertions of "best practice." Nevertheless, there are enough investigations to suggest some design parameters for deliberative processes that are likely, although not guaranteed, to create a balanced consensus. By balanced consensus we mean guidance that respects both scientific integrity on the one hand and its implementability in a specific health system context on the other. A balanced consensus, obtained by careful consideration of all relevant evidence, and involving a good range of those best qualified to assess it and those most likely to be affected by the outcome, is a worthy objective — and one that the literature, taken as a whole, seems to indicate can be achieved.

References

1. *Oxford American dictionary*. 1st ed. New York: Avon Books; 1980.

2. *What counts? Interpreting evidence-based decision-making for management and policy: report of the 6th CHSRF Annual Invitational Workshop, Vancouver, British Columbia, March 11, 2004*. Ottawa: Canadian Health Services Research Foundation; 2005. Available: http://www.chsrf.ca/knowledge_transfer/pdf/2004_workshop_report_e.pdf.

3. Rosen R. Applying research to health care policy and practice: medical and managerial views on effectiveness and the role of research. *J Health Serv Res Policy* 2000;5(2):103-8.

4. Mitton C, Patten S. Evidence-based priority-setting: what do the decision-makers think? *J Health Serv Res Policy* 2004;9(3):146-52.

5. Rychetnik L, Hawe P, Waters E, Barratt A, Frommer M. A glossary for evidence based public health. *J Epidemiol Community Health* 2004;58(7):538-45. Available: http://jech.bmjjournals.com/cgi/content/full/58/7/538.

6. Petticrew M, Whitehead M, Macintyre SJ, Graham H, Egan M. Evidence for public health policy on inequalities: 1: the reality according to policymakers. *J Epidemiol Community Health* 2004;58(10):811-6.

7. Swinburn B, Gill T, Kumanyika S. Obesity prevention: a proposed framework for translating evidence into action. *Obes Rev* 2005;6(1):23-33.

8. Jenkings KN, Barber N. What constitutes evidence in hospital new drug decision making? *Soc Sci Med* 2004;58(9):1757-66.

9. Lomas J. Making clinical policy explicit. Legislative policy making and lessons for developing practice guidelines. *Int J Technol Assess Health Care* 1993;9(1):11-25.

10. Rycroft-Malone J, Seers K, Titchen A, Harvey G, Kitson A, McCormack B. What counts as evidence in evidence-based practice? *J Adv Nurs* 2004;47(1):81-90.

11. Scott-Findlay S, Pollock C. Evidence, research, knowledge: a call for conceptual clarity. *Worldviews Evid Based Nurs* 2004;1(2).

12. Atkins D, Siegel J, Slutsky J. Making policy when the evidence is in dispute. *Health Aff* 2005;24(1):102-13.

13. Whitehead M, Petticrew M, Graham H, Macintyre SJ, Bambra C, Egan M. Evidence for public health policy on inequalities: 2: assembling the evidence jigsaw. *J Epidemiol Community Health* 2004;58(10):817-21.

14. Norheim OF. The role of evidence in health policy making: a normative perspective. *Health Care Anal* 2002;10(3):309-17.

15. Rycroft-Malone J, Stetler CB. Commentary on evidence, research, knowledge: a call for conceptual clarity. *Worldviews Evid Based Nurs* 2004;1(2):98-101.

16. Zarkovich E, Upshur RE. The virtues of evidence. *Theor Med Bioeth* 2002;23(4-5):403-12.

17. Upshur RE. Seven characteristics of medical evidence. *J Eval Clin Pract* 2000;6(2):93-7.

18. Upshur RE. If not evidence, then what? Or does medicine really need a base? *J Eval Clin Pract* 2002;8(2):113-9.

19. McQueen DV, Anderson LM. What counts as evidence: issues and debates. *WHO Reg Publ Eur Ser* 2001;(92):63-81.

20. Hammersley M. Is the evidence-based practice movement doing more good than harm? Reflections on Iain Chalmers' case for research-based policy making and practice. *Evid Policy* 2005;1(1):85-100. Available: http://docserver.ingentaconnect.com/deliver/cw/tpp/17442648/v1n1/s5/p85.pdf?fmt=dirpdf&tt=10934&cl=33&ini=connect&bini=&wis=connect&ac=0&acs=&expires=1115782089&checksum=3A6D5FB82A4C0846A98E5F5CEA5B3929&cookie=1898466304.

21. Walshe K, Rundall TG. Evidence-based management: from theory to practice in health care. *Milbank Q* 2001;79(3):429-57, IV-V.

22. Eisenberg JM. What does evidence mean? Can the law and medicine be reconciled? *J Health Polit Policy Law* 2001;26(2):369-81.

23. Dobrow MJ, Goel V, Upshur RE. Evidence-based health policy: context and utilisation. *Soc Sci Med* 2004;58(1):207-17.

24. Greenhalgh T, Worrall JG. From EBM to CSM: the evolution of context-sensitive medicine. *J Eval Clin Pract* 1997;3(2):105-8.

25. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996;312(7023):71-2.

26. Davidoff F, Haynes B, Sackett D, Smith R. Evidence based medicine. *BMJ* 1995;310(6987):1085-6.

27. Health Evidence Network. *Evidence on health needs and intervention*. Copenhagen: Regional Office for Europe, World Health Organization; 2004. Available: http://www.euro.who.int/evidence/policy/20040623_1.

28. Strategic Policy Making Team, Cabinet Office. *Professional policy making for the twenty-first century*. London: The Office; 1999. Available: http://www.policyhub.gov.uk/docs/profpolicymaking.pdf (accessed 2005 Mar).

29. Saarni SI, Gylling HA. Evidence based medicine guidelines: a solution to rationing or politics disguised as science? *J Med Ethics* 2004;30(2):171-5.

30. Nutley S, Davies H, Walter I. *Evidence based policy and practice: cross sector lessons from the UK* [Working paper 9]. London: ESRC UK Centre for Evidence Based Policy and Practice; 2002. Available: http://www.evidencenetwork.org/Documents/wp9b.pdf.

31. Klein R. Evidence and policy: interpreting the Delphic oracle. *J R Soc Med* 2003;96(9):429-31.

32. *How research and evaluation evidence contributes to policy making*. London: Policy Hub, Cabinet Office, Government Social Research Unit; 2004. Available: http://www.policyhub.gov.uk/evalpolicy/how_res_eval_evid.asp.

33. Upshur RE, VanDenKerkhof EG, Goel V. Meaning and measurement: an inclusive model of evidence in health care. *J Eval Clin Pract* 2001;7(2):91-6.

34. McCormack B, Kitson A, Harvey G, Rycroft-Malone J, Titchen A, Seers K. Getting evidence into practice: the meaning of 'context'. *J Adv Nurs* 2002;38(1):94-104.

35. Buetow S, Kenealy T. Evidence-based medicine: the need for a new definition. *J Eval Clin Pract* 2000;6(2):85-92.

36. Davies P. Evidence-based government: how can we make it happen? [oral presentation]. CHSRF 7th annual invitational workshop--leveraging knowledge: tools and strategies for action; 2005 Mar 3; Montreal.

37. Eddy DM. Selecting technologies for assessment. *Int J Technol Assess Health Care* 1989;5(4):485-501.

38. Lomas J, Fulop N, Gagnon D, Allen P. On being a good listener: setting priorities for applied health services research. *Milbank Q* 2003;81(3):363-88.

39. Klein R, Williams A. Setting priorities: what is holding us back--inadequate information or inadequate institutions? In: Coulter A, Ham C, editors. *The global challenge of health care rationing*. Buckingham (UK): Open University Press; 2000. p.15-26.

40. Bal R, Bijker WE, Hendriks R. Democratisation of scientific advice. *BMJ* 2004;329(7478):1339-41.

41. Charnley G. *Enhancing the role of science in stakeholder-based risk management decision-making*. Washington: HealthRisk Strategies; 2000. Available: http://www.riskworld.com/Nreports/2000/Charnley/NR00GC00.htm.

42. Kim MS. Cloning and deliberation: Korean consensus conference. *Developing World Bioeth* 2002;2(2):159-72.

43. Charles C, Schalm C, Semradek J. Involving stakeholders in health services research: developing Alberta's resident classification system for long-term care facilities. *Int J Health Serv* 1994;24(4):749-61.

44. Petts J. Barriers to participation and deliberation in risk decisions: evidence from waste management. *J Risk Res* 2004;7(2):115-33.

45. Garland MJ. Rationing in public: Oregon's priority-setting methodology. In: Strosberg MA, Wiener JM, Baker R, Fein IA, editors. *Rationing America's medical care: the Oregon plan and beyond*. Washington: Brookings Institution; 1992. p.37-59.

46. Veatch RM. Consensus of expertise: the role of consensus of experts in formulating public policy and estimating facts. *J Med Philos* 1991;16(4):427-45.

47. Cook M. Evidence-based medicine and experience-based practice--clash or consensus? *Med Law* 2004;23(4):735-43.

48. Browman GP, Snider A, Ellis P. Negotiating for change. The healthcare manager as catalyst for evidence-based practice: changing the healthcare environment and sharing experience. *Healthc Pap* 2003;3(3):10-22.

49. Stern PC, Fineberg HV, editors, Committee on Risk Characterization, National Research Council. *Understanding risk: information decisions in a democratic society*. Washington: National Academies Press; 1996.

50. Daniels N. Accountability for reasonableness. *BMJ* 2000;321(7272):1300-1. Available: http://bmj.bmjjournals.com/cgi/content/full/321/7272/1300.

51. Schneider E, Oppermann B, Renn O. Experiences from Germany: application of a structured model of public participation in waste management planning. *Interact J Public Participation* 1998;4(1):63-72.

52. Webler T. 'Right' discourse in citizen participation: an evaluative yardstick. In: Renn O, Webler T, Wiedemann P, editors. *Fairness and competence in citizen participation: evaluating models for environmental discourse* [Technology, risk and society v 10]. Dordrecht (The Netherlands): Kluwer Academic; 1995. p.35-86.

53. Palmer JS. Prioritization in community health planning: combining methods to achieve implementable priorities. *J Health Hum Serv Adm* 1998;21(1):109-34.

54. Culyer AJ. Involving stakeholders in health care decisions – the experience of the National Institute for Clinical Excellence (NICE) in England and Wales. *Healthc Q* 2005;8(3):54-8.

55. National Institute for Clinical Excellence. *Guide to the methods of technology appraisal*. London: The Institute; 2004. Available: http://www.nice.org.uk/pdf/TAP_Methods.pdf.

56. *NICE Citizens Council report on age*. London: National Institute for Clinical Excellence; 2004. Available: http://www.nice.org.uk/pdf/Citizenscouncil_report_age.pdf.

57. Black N, Murphy M, Lamping D, McKee M, Sanderson C, Askham J, et al. Consensus development methods: a review of best practice in creating clinical guidelines. *J Health Serv Res Policy* 1999;4(4):236-48.

58. McGlynn EA, Kosecoff J, Brook RH. Format and conduct of consensus development conferences. Multi-nation comparison. *Int J Technol Assess Health Care* 1990;6(3):450-69.

59. Wortman PM, Vinokur A, Sechrest L. Do consensus conferences work? A process evaluation of the NIH consensus development program. *J Health Polit Policy Law* 1988;13(3):469-98.

60. Lomas J, Anderson G, Enkin M, Vayda E, Roberts R, MacKinnon B. The role of evidence in the consensus process. Results from a Canadian consensus exercise. *JAMA* 1988;259(20):3001-5.

61. Gibson JL, Martin DK, Singer PA. Evidence, economics and ethics: resource allocation in health services organizations. *Healthc Q* 2005;8(2):50-9.

62. Devlin N, Parkin D, Gold M. WHO evaluates NICE. *BMJ* 2003;327(7423):1061-2. Available: http://bmj.bmjjournals.com/cgi/content/full/327/7423/1061.

63. Hill S, Garattini S, van Loenhout J, O'Brien BJ, de Joncheere K. *Technology appraisal programme of the National Institute for Clinical Excellence: a review by WHO*. Geneva: World Health Organization; 2003. Available: http://www.nice.org.uk/pdf/boardmeeting/brdsep03itemtabled.pdf.

Appendix 1

Methods and Results

Search Strategy

The search was done between January and April 2005. First, the project team assembled a core list of relevant papers and experts. Experts were then contacted using a standardized but personalized form letter. The letter included a list of selected, relevant citations and requested that they suggest key literature as well as identify other experts. Input from these new experts was solicited using the same form letter.

The resulting list of citations was analysed to identify appropriate controlled vocabulary terms (e.g. MeSH) and the more frequently occurring key words. Two main electronic search strategies were devised. Due to the difficulty of searching efficiently for the very general question of "what counts as evidence?" relevant articles were identified through three sources: 1) the literature identified by the project team and experts; 2) web sites of appropriate agencies; and 3) the results found using the two electronic strategies.

The first strategy addressed perspectives on evidence, that is, what are health system managers' and decision makers' perspectives on what constitutes evidence. The second strategy explored deliberative processes that combine stakeholder perspectives and different forms of evidence. Due to the tight timeframe involved and the large volume of literature available, all controlled vocabulary was limited to major descriptors (the descriptor is a major focus of the article). When appropriate, frequency operators were also used for abstract keywords to increase relevancy and precision in search results.

Several electronic databases were searched: BIOSIS, Cumulative Index to Nursing & Allied Health Literature (CINAHL), HealthSTAR/Ovid Healthstar, MEDLINE® (including In-Process & Other Non-Indexed Citations), and PsycINFO were searched on Ovid. The Psychology and Behavioral Sciences Collection was searched on EBSCOhost, and System for Information on Grey Literature in Europe (SIGLE) was searched on Silverplatter. OVID EMBASE was searched for the deliberative process strategy. Results were limited to the publication dates 1990-2005 in the perspectives on evidence strategy (1980-2005 for the deliberative process strategy) and, where possible, the human population. All references were downloaded into Reference Manager® bibliographic software, coded as to which question they addressed, and duplicates removed. The first strategy retrieved 1,694 unique hits following de-duplication in Reference Manager; the second strategy retrieved 834.

Additional grey literature was sought through the web sites of relevant health services research organizations (such as AHRQ, Evidence for Policy and Practice Information, WHO Health Evidence Network) and health technology assessment agencies (such as CCOHTA, Centre for Reviews and Dissemination, NICE) and through searching the

bibliographies of relevant articles, reports, and monographs. The detailed strategy is illustrated at the end of this appendix, using MEDLINE as an example.

Including the core articles selected by the research team, expert-recommended articles, and the results of the electronic search, a total of 2,243 articles were considered to answer question 1. It was necessary to go beyond our initial search strategy on deliberative processes; the details of this additional search are explained in the section below. In total, 855 articles were considered to answer question 2.

Screening of Search Results

Included core and expert-recommended articles

The initial set of papers was based on the research team's suggestions. For question 1 a core set of six papers was identified; for question 2 a core set of three papers was identified. At the beginning of the project the research team felt that expert-recommended articles could automatically be selected for inclusion. However, the quality and relevance of the recommended articles varied substantially. Most were pooled together with the results of the electronic search for the screening of titles and abstracts. Some of the citations seemed highly relevant and were immediately retrieved for full review: for the first question, three articles were added to the set of six; for the second, one article was added to the set of three. The full research team met to discuss these articles and agreed that all were of high enough quality and relevance to be used in the final paper.

Screening of titles and abstracts

For both research questions, two authors independently reviewed title and abstract of early scoping searches. Papers were identified as meeting, possibly meeting, or not meeting question-specific inclusion/exclusion criteria. A third author independently assessed a sub-set of the scoping sample for verification. Those meeting or potentially meeting the criteria were used to refine the search and were retrieved to inform the project.

Two authors applied the screening criteria to all the expert-recommended citations and the results of the final electronic search. Consensus was sought and reached. All papers meeting the inclusion/exclusion criteria were retrieved. Papers potentially meeting the criteria were retrieved when the likelihood of relevance, as determined by the reviewers, was high.

In the case of question 1, 188 articles were selected for inclusion. Only 17 articles passed the inclusion/exclusion criteria for question 2. Due to the low yield, two authors re-screened the articles rated as possibly meeting the criteria. Then the bibliographies of all the included papers were searched for relevant citations. The authors subsequently retrieved these papers and applied the criteria to them. Six additional papers were identified this way.

There was some concern that no articles on the Oregon Health Plan process passed the screening step. None of the identified or retrieved literature concerning the Oregon Health Plan detailed the deliberative process used. Two authors re-screened all identified articles and books related to Oregon and reviewed their reference lists. When this step proved unsatisfactory, a specific but informal web search was conducted to find literature summarizing Oregon's deliberative process. One book and four additional papers were identified this way.

Inclusion/Exclusion Criteria

Question 1
In addition to research on health outcomes, what other forms of information count as evidence for clinical, management, or policy decision-making in the health sector?

Literature to be included must focus on:
1. Definitional aspects of evidence for health decision-making

OR

2. Relevance of sources/forms of evidence for health decision-making

Excluded:
1. Articles focused only on hierarchies or levels of evidence related to study design
2. Articles on the quality of individual studies or types (designs) of studies
3. Articles focused on the transfer or use of research (or other) evidence
4. Articles focused on a particular health practice or condition that is not generalizable

Question 2
How can various forms of evidence and stakeholder perspectives be combined through a deliberative process to yield evidence-informed guidance for health systems?

Included literature must:
1. a) Define what a deliberative process is

   OR

   b) Describe a tried deliberative process for using research evidence to guide the health system

   AND

2. Be evaluative (that is, it must in some way judge the merit of the process)

Excluded:
1. Theoretical articles that are not definitional
2. Processes that only include public or patient input
3. Processes that do not include research evidence
4. Articles that are only descriptive (that is, not evaluative)

Final Review of Screened Articles

Given the difference in the number of papers meeting the inclusion/exclusion criteria, different approaches were used for the two research questions. Some of the papers were more difficult to retrieve and could not be secured in time to meet the project timelines (12 for question 1; one for question 2).

Question 1

A standardized data collection form was created. The form included a subjective assessment of relevance (high, medium, or low). The authors completed the form for one of the core papers selected at the beginning of the project. The project team then recruited five individuals to read the papers and complete data collection forms. The sample data collection form and associated paper were provided for reference. Readers were asked to complete five papers at the beginning; two authors assessed their work and provided guidance where necessary. Each reader collected data from a minimum of 30 papers.

One author read all papers ranked high by the readers (N=53). Sixteen of the 53 papers were forwarded to the writing team. This author then scanned the data collection forms from papers ranked medium (N=67) to identify papers appearing sufficiently relevant for further review. Thirty medium-ranked papers were read in full and four were forwarded to the writing team. Fifty-six papers were ranked low; a sample (N=10) of these papers was also considered to ensure consistent and accurate assessment. No papers ranked low were forwarded to the writing team. Papers forwarded to the writing team for inclusion were those felt to be of highest relevance, meaning they were directly relevant to the topic and presented novel ideas (only representative samples of papers repeating the same themes were chosen). The writing team then received all data abstraction forms for review and consideration. The writing team requested no further papers.

The 20 papers passing in-depth review were added to the nine core and expert articles selected at the beginning of the project (see Appendix 2 for the full list of 29 articles).

Question 2

Due to the low number of included papers, a data abstraction form was not needed. All retrieved papers were read by one of two authors. The two authors then met to discuss which papers were most relevant and would be sent to the writing team. Relevance was again determined by how well the paper met the inclusion criteria and how novel the ideas were. Eleven papers passed in-depth review. These were added to the four core and expert articles selected at the beginning of the project (see Appendix 3 for the full list of 15 articles).

Search Strategy Illustration

This illustration uses MEDLINE as an example. Vocabulary and syntax were adapted according to the database and platform used. Further information is available by contacting the project team.

Exp	Explode (i.e., subject heading)
Adj	Adjacent, any word order
$	Truncation symbol
*	Focus (i.e., major descriptor)
/	link to subheading
freq	Frequency
ti	Title
ab	Abstract
yr	Publication year

Strategy #1: What Counts as Evidence? (performed 29 Jan 2005):

Ovid MEDLINE(R) <1966 to January Week 3 2005>

1 *Evidence-Based Medicine/
2 (((Evidence-base$ or evidence) adj base$) or EBM).ti..
3 (((Evidence-base$ or evidence) adj base$) or EBM).ab. /freq=3
4 *Health Services Research/
5 1 or 2 or 3 or 4
6 exp *Decision Making/
7 (((((decision-maker$ or Decision$) adj1 maker$) or decision-making$ or decision$) adj1 making) or decisionmaker$ or decisionmaking).ti.
8 (((((decision-maker$ or Decision$) adj1 maker$) or decision-making$ or decision$) adj1 making) or decisionmaker$ or decisionmaking).ab. /freq=3
9 exp *Health Policy/
10 (((health adj policy) or health) adj policies).ti.
11 (((health adj policy) or health) adj policies).ab. /freq=2
12 exp *Policy Making/
13 exp *POLITICS/
14 ((((((((((policy adj1 maker$) or policy) adj1 making) or policies) adj1 maker$) or policies) adj1 making) or policymaker$ or policymaking$ or policy) adj1 develop$) or policies) adj1 develop$).ti.
15 (((((((policy adj1 maker$) or policy) adj1 making) or policymaker$ or policymaking$ or policy) adj1 develop$) or policies) adj1 develop$).ab. /freq=2
16 ((((healthcare or health) adj care) or health-care) adj manager$1).ti.
17 (managerial adj practice$).ti,ab.
18 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17
19 5 and 18

20 ((Evidence or (research adj finding$)) adj2 (policymaking or decisionmaking or policy-making or decision-making or ((policy adj1 making) or (decision adj1 making) or policymaker$ or (policy adj1 maker$) or policy-maker$ or decisionmaker$ or (decision adj1 maker$)) or (decision-maker$ or (manage$ adj decision$) or (political adj decision$) or (policy adj1 decision$) or (public adj polic$))))).ti.
21 ((Evidence or (research adj finding$)) adj2 (policymaking or decisionmaking or policy-making or decision-making or ((policy adj1 making) or (decision adj1 making) or policymaker$ or (policy adj1 maker$) or policy-maker$ or decisionmaker$ or (decision adj1 maker$)) or (decision-maker$ or (manage$ adj decision$) or (political adj decision$) or (policy adj1 decision$) or (public adj polic$))))).ab. /freq=2
22 (EBDM or EBPM).ab. /freq=2
23 20 or 21 or 22
24 19 or 23
25 limit 24 to yr=1990 - 2005
26 limit 25 to humans

Strategy #2: Deliberative Processes (performed 26 Mar 2005):

Ovid MEDLINE(R) <1966 to March Week 3 2005>

1 exp *Evidence-Based Medicine/
2 exp *Practice Guidelines/
3 exp *Public Health Administration/
4 exp *Consensus Development Conferences/
5 exp *Health Planning Guidelines/
6 *Health Planning/og [Organization & Administration]
7 exp *Health Planning Organizations/
8 exp *State Medicine/og [Organization & Administration]
9 ((healthcare or (health adj care) or health) adj service$ adj3 deliver$).ti.
10 ((healthcare or (health adj care) or health) adj service$ adj3 deliver$).ab.
11 (evidence or evidence-based or (evidence adj based) or EBM or EBDM).ti,ab.
12 ((practice adj guideline$) or cpg or cpgs).ti,ab.
13 (Scientific adj (knowledge or advice or fact$ or proof$)).ti,ab.
14 (Health adj2 administration).ti,ab.
15 ((state or social$) adj1 medicine).ti,ab.
16 (NHS or (national adj health adj service$)).ti,ab.
17 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16
18 exp *Organizational Policy/
19 exp *Policy Making/
20 exp *Public Policy/
21 exp *Decision Making/
22 exp *Decision Support Techniques/
23 exp *Decision Making, Organizational/

24 ((policy adj1 making$) or (policies adj1 making$) or policymaking$ or (policy adj1 develop$) or (policies adj1 develop$)).ti.
25 ((policy adj1 making$) or (policies adj1 making$) or policymaking$ or (policy adj1 develop$) or (policies adj1 develop$)).ab.
26 ((public or health or healthcare or (health adj care)) adj1 (planning or policy or policies)).ti,ab.
27 ((decisionmaking$ or decision-making$ or decision$) adj1 making$).ti,ab.
28 ((decision-support$ or decision$) adj1 support$).ti,ab.
29 exp *Health Care Rationing/
30 exp *Health Priorities/
31 ((healthcare or (health adj care)) adj1 (rationing or priority or priorities)).ti,ab.
32 (Priority adj1 setting).ti,ab.
33 (setting adj1 priorities).ti,ab.
34 (prioritiz$ or prioritis$).ti,ab.
35 (Accountability adj1 reasonableness).ti,ab.
36 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35
37 *Group Processes/
38 exp *"Dissent and Disputes"/
39 exp *CONSENSUS/
40 exp *NEGOTIATING/
41 exp *DEMOCRACY/
42 exp *Community-Institutional Relations/
43 exp *Interinstitutional Relations/
44 ((stakeholder$ or (stake adj holder$) or participant$ or public or citizen$ or multilateral or multi-lateral or multiple) adj1 (collaborat$ or contribut$ or consensus$ or cooperat$ or deliberat$ or engage$ or input or inputs or involve$ or participat$ or partner$)).ti.
45 ((stakeholder$ or (stake adj holder$) or participant$ or public or citizen$ or multilateral or multi-lateral or multiple) adj1 (collaborat$ or contribut$ or consensus$ or cooperat$ or deliberat$ or engage$ or input or inputs or involve$ or participat$ or partner$)).ab.
46 (deliberative or deliberation$ or iterative or iteration$ or dissent$ or dispute$ or dissension$).ti,ab.
47 (democracy or democrati$).ti,ab.
48 (Citizen$ adj (jury or juries)).ti,ab.
49 ((Group adj process$) or (consensus adj process$)).ti,ab.
50 (Nominal adj group adj technique$).ti,ab.
51 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50
52 36 and 51
53 17 and 52
54 limit 53 to yr=1980 - 2005

Appendix 2

Articles Passing Screen and Final Review for Question 1 (What Counts as Evidence)

Atkins D, Siegel J, Slutsky J. Making policy when the evidence is in dispute. *Health Aff* (Millwood) 2005;24(1):102-13.

Buetow S, Kenealy T. Evidence-based medicine: the need for a new definition. *J Eval Clin Pract* 2000;6(2):85-92.

What counts? Interpreting evidence-based decision-making for management and policy: report of the 6th CHSRF Annual Invitational Workshop, Vancouver, British Columbia, March 11, 2004. Ottawa: Canadian Health Services Research Foundation; 2005 Available: http://www.chsrf.ca/knowledge_transfer/pdf/2004_workshop_report_e.pdf.

Dobrow MJ, Goel V, Upshur RE. Evidence-based health policy: context and utilisation. *Soc Sci Med* 2004;58(1):207-17.

Eisenberg JM. What does evidence mean? Can the law and medicine be reconciled? *J Health Polit Policy Law* 2001;26(2):369-81.

Forbes A, Griffiths P. Methodological strategies for the identification and synthesis of 'evidence' to support decision-making in relation to complex healthcare systems and practices. *Nurs Inq* 2002;9(3):141-55.

Greenhalgh T, Worrall JG. From EBM to CSM: the evolution of context-sensitive medicine. *J Eval Clin Pract* 1997;3(2):105-8.

Jenkings KN, Barber N. What constitutes evidence in hospital new drug decision making? *Soc Sci Med* 2004;58(9):1757-66.

Klein R. Evidence and policy: interpreting the Delphic oracle. *J R Soc Med* 2003;96(9):429-31.

McCormack B, Kitson A, Harvey G, Rycroft-Malone J, Titchen A, Seers K. Getting evidence into practice: the meaning of 'context'. *J Adv Nurs* 2002;38(1):94-104.

McQueen DV, Anderson LM. What counts as evidence: issues and debates. *WHO Reg Publ Eur Ser* 2001;(92):63-81.

Mitton C, Patten S. Evidence-based priority-setting: what do the decision-makers think? *J Health Serv Res Policy* 2004;9(3):146-52.

Norheim OF. The role of evidence in health policy making: a normative perspective. *Health Care Anal* 2002;10(3):309-17.

Nutley S, Davies H, Walter I. *Evidence based policy and practice: cross sector lessons from the UK* [Working paper 9]. London: ESRC UK Centre for Evidence Based Policy and Practice; 2002. Available: http://www.evidencenetwork.org/Documents/wp9b.pdf.

Petticrew M, Whitehead M, Macintyre SJ, Graham H, Egan M. Evidence for public health policy on inequalities: 1: the reality according to policymakers. *J Epidemiol Community Health* 2004;58(10):811-6.

Rosen R. Applying research to health care policy and practice: medical and managerial views on effectiveness and the role of research. *J Health Serv Res Policy* 2000;5(2):103-8.

Rychetnik L, Hawe P, Waters E, Barratt A, Frommer M. A glossary for evidence based public health. *J Epidemiol Community Health* 2004;58(7):538-45. Available: http://jech.bmjjournals.com/cgi/content/full/58/7/538.

Rycroft-Malone J, Seers K, Titchen A, Harvey G, Kitson A, McCormack B. What counts as evidence in evidence-based practice? *J Adv Nurs* 2004;47(1):81-90.

Rycroft-Malone J, Stetler CB. Commentary on evidence, research, knowledge: a call for conceptual clarity. *Worldviews Evid Based Nurs* 2004;1(2):98-101.

Saarni SI, Gylling HA. Evidence based medicine guidelines: a solution to rationing or politics disguised as science? *J Med Ethics* 2004;30(2):171-5.

Scott-Findlay S, Pollock C. Evidence, research, knowledge: a call for conceptual clarity. *Worldviews Evid Based Nurs* 2004;1(2).

Small N. Viewpoint. Knowledge, not evidence, should determine primary care practice. *Clin Governance Int J* 2003;8(3):191-9.

Sturm R. Evidence-based health policy versus evidence-based medicine. *Psychiatr Serv* 2002;53(12):1499.

Swinburn B, Gill T, Kumanyika S. Obesity prevention: a proposed framework for translating evidence into action. *Obes Rev* 2005;6(1):23-33.

Upshur RE, VanDenKerkhof EG, Goel V. Meaning and measurement: an inclusive model of evidence in health care. *J Eval Clin Pract* 2001;7(2):91-6.

Upshur RE. If not evidence, then what? Or does medicine really need a base? *J Eval Clin Pract* 2002;8(2):113-9.

Upshur RE. Seven characteristics of medical evidence. *J Eval Clin Pract* 2000;6(2):93-7.

Whitehead M, Petticrew M, Graham H, Macintyre SJ, Bambra C, Egan M. Evidence for public health policy on inequalities: 2: assembling the evidence jigsaw. *J Epidemiol Community Health* 2004;58(10):817-21.

Zarkovich E, Upshur RE. The virtues of evidence. *Theor Med Bioeth* 2002;23(4-5):403-12.

Appendix 3

Articles Passing Screen and Final Review for Question 2 (Deliberative Processes for Combining Forms of Evidence)

Bal R, Bijker WE, Hendriks R. Democratisation of scientific advice. *BMJ* 2004;329(7478):1339-41.

Black N, Murphy M, Lamping D, McKee M, Sanderson C, Askham J, et al. Consensus development methods: a review of best practice in creating clinical guidelines. *J Health Serv Res Policy* 1999;4(4):236-48.

Browman GP, Snider A, Ellis P. Negotiating for change. The healthcare manager as catalyst for evidence-based practice: changing the healthcare environment and sharing experience. *Healthc Pap* 2003;3(3):10-22.

Charles C, Schalm C, Semradek J. Involving stakeholders in health services research: developing Alberta's resident classification system for long-term care facilities. *Int J Health Serv* 1994;24(4):749-61.

Charnley G. *Enhancing the role of science in stakeholder-based risk management decision-making*. Washington: HealthRisk Strategies; 2000. Available: http://www.riskworld.com/Nreports/2000/Charnley/NR00GC00.htm.

Cook JA, Toprac M, Shore SE. Combining evidence-based practice with stakeholder consensus to enhance psychosocial rehabilitation services in the Texas benefit design initiative. *Psychiatr Rehabil J* 2004;27(4):307-18.

Garland MJ. Rationing in public: Oregon's priority-setting methodology. In: Strosberg MA, Wiener JM, Baker R, Fein IA, editors. *Rationing America's medical care: the Oregon plan and beyond*. Washington: Brookings Institution; 1992. p.37-59.

Gibson JL, Martin DK, Singer PA. Evidence, economics and ethics: resource allocation in health services organizations. *Healthc Q* 2005;8(2):50-9.

Kim MS. Cloning and deliberation: Korean consensus conference. *Developing World Bioeth* 2002;2(2):159-72.

Lomas J, Anderson G, Enkin M, Vayda E, Roberts R, MacKinnon B. The role of evidence in the consensus process. Results from a Canadian consensus exercise. *JAMA* 1988;259(20):3001-5.

Lomas J. Making clinical policy explicit. Legislative policy making and lessons for developing practice guidelines. *Int J Technol Assess Health Care* 1993;9(1):11-25.

McGlynn EA, Kosecoff J, Brook RH. Format and conduct of consensus development conferences. Multi-nation comparison. *Int J Technol Assess Health Care* 1990;6(3):450-69.

Palmer JS. Prioritization in community health planning: combining methods to achieve implementable priorities. *J Health Hum Serv Adm* 1998;21(1):109-34.

Petts J. Barriers to participation and deliberation in risk decisions: evidence from waste management. *J Risk Res* 2004;7(2):115-33.

Wortman PM, Vinokur A, Sechrest L. Do consensus conferences work? A process evaluation of the NIH consensus development program. *J Health Polit Policy Law* 1988;13(3):469-98.