Paper presented at the John Deutsch Institute Conference, "Health Services Restructuring: New Evidence and New Directions", November 17-18, 2005 at Queen's University. The final version is published in a book entitled *Health Services Restructuring: New Evidence and New Directions*, edited by C.M. Beach, R.C. Chaykowski, S. Shortt, F. St-Hilaire, and A. Sweetman, 2006 (Kingston: John Deutsch Institute, Queen's University). Published in cooperation with McGill-Queen's University Press and available at: http://mqup.mcgill.ca .

Deliberative processes and evidence-informed decision-making in health care

Tony Culyer

Institute for Work & Health

It has often been remarked that the evidential base in support of the proposition that evidence-informed decision making is, in broadest terms, a 'good thing' is itself distressingly thin. If true, that would, of course, be a quite an irony, even an oxymoron – an un-evidenced case for the use of evidence? Often coupled with the advocacy of evidence-informed decision-making goes an advocacy of openness and transparency in decision-making, the design of decision-making processes, and the use of deliberative processes with lots of consultation of so-called 'stakeholders'. Again a question cries out for an answer – regardless of any intrinsic attractiveness it may have, what's the evidence that this way of doing things 'works'?

In this lecture I want to turn the Evidence-Informed Decision-making approach in on itself in order to address just one part of the agenda that this irony suggests. I shall ask what there is of an evidential kind to support the propositions that *deliberative processes*, *transparency of process* and the *wide extent of consultation* that are often urged ought to accompany

them are ‘effective’. Where evidence fails, I shall ask what prima facie reasons we might reasonably entertain for supposing them to be ‘effective’. At some stage I shall have, of course, to tell you what I mean when I describe something as ‘effective’ and how we might find out whether it actually. But that is for later.

I am much indebted for what I have to say to my colleagues Jonathan Lomas, Chris McCutcheon, Laura McAuley and Susan Law, with whom I have been having discussions on this subject for the past year or so in conjunction with reviewing systematically the literature claiming to assess the effectiveness of these methods of decision-making (see Lomas *et al*. 2005)[1]. I am also grateful to the members of a workshop held under the auspices of the Canadian Health Services Research Foundation at which many of the ideas here were discussed. I must of course absolve all of these people from any responsibility for the imperfect parts of what I shall say. The best bits are undoubtedly stolen from them.

In my mind’s eye a deliberative process is characterized by – of course – deliberation! That is, the careful, deliberate, consideration and discussion of the advantages and disadvantages of various options. In many contexts, and certainly in the context I envisage here and now, the deliberation is going to be about health care and access to it, its finance, provision, distribution and management. Moreover, one of the principal matters about which any deliberation will take place will concern evidence. So I shall go on to consider what it

[1] Two questions informed the literature search : (1) In addition ton research on health outcomes, what other forms of information count as evidence for clinical, management, or policy decision-making in the health sector? (2) How can various forms of evidence and stakeholder perspectives be combined through a deliberative process to yield evidence-informed guidance for health systems? Initially 2,243 items were found in relation to question (1) and 855 to question (2). 188 articles passed the inclusion tests in the case of (1) but only 17 in the case of (2) (later supplemented by six papers from bibliographies of included papers).

might be sensible to regard as evidence when decisions are being made about health care and then I shall move on to consider some arguments for deliberative processes, whether some types of decision lend themselves more appropriately to deliberative processes than others, what kind of processes are most likely to deliver satisfactory outcomes, and what the empirical evidence is that bears on these matters. I shall conclude with a suggested framework for future appraisals of deliberative processes.

I am going to assume throughout this discussion that the purpose of deliberative processes is to help people to make decisions. A deliberative process, like evidence itself, or argument, ethics, algorithms, concepts of social welfare and its maximization, cost-benefit analysis, and like the concept of a fair or just distribution of health or of health services, does not have its justification in order to dictate or prescribe or even constrain the outcome of the process. It is there to facilitate the consideration of matters deemed to be relevant. Some of these are evidential and factual in nature. Others are conceptual, theoretical and – like deliberative processes themselves – procedural. None of these is there as a *substitute for thought*. Rather, they are all there as *aids to thought*, *aids to judgment*, *aids to better thinking*: thinking that is more comprehensive in the issues embraced, more consistent in the way they are embraced and more engaging of the people affected by the outcome. It is not even necessarily entirely consequentialist, in the sense that the goodness or badness of deliberative processes is to be judged only in terms of the goodness or badness of their outcomes. Though the case for it is certainly to some extent consequentialist, it can also be claimed that another purpose is for it to be socially integrative, or that the process is itself inherently good, regardless of any improvement that may result in terms of the quality of

decisions reached. For current purposes, however, I shall take it that it is the consequences that matter and that the consequences with which we are especially concerned are the decisions that the process enables.

So, having laid at least some of my cards on the table, let us turn to the question of evidence.

What is evidence?

One's first recourse in considering the nature of 'evidence' might be to a quasi-legal notion. In a court of law, evidence is what someone says is or was the case under oath. The statement made under oath may about what events (what actually happened), about descriptive characteristics of aspects of events (like a person's apparent identity or state of mind), or about relationships (like the scientific grounds for believing that x will cause y). So evidence is empirical, to do with the 'facts of the case' and the kind of evidence that is relevant in any situation consists in material facts that help establish the truth of someone's assertions ('are they lying?') or the cause of a consequence ('did he do it?'). While this is, indeed, a useful concept of evidence - and is one lying at the heart of the jurisprudential approach to evidence - it will not quite do for our purposes. One reason for this, which may strike you as contentious, is that the legal notion of admissible evidence *excludes* hearsay, which I do not wish to exclude axiomatically from that which may be counted as 'evidence', given its long traditions in medicine and management as 'the case-study'. A more powerful reason is that the legal idea of evidence is always locked into the particular circumstances of the case: 'did *he* do *it*?'. While the law allows judgment to be exercised

and does not normally require certainty or complete conclusiveness, there should be no reasonable doubt about any conclusion of consequence informed by evidence. It has a retrospective stance – '*did* he do it?' – requiring the establishment of particular historical events (for example, whether they actually took place) and the frame of mind of the accused at the time.

In health care decisions, however, one is nearly always concerned with forecasts, predicting and prognosticating. One is a good deal more uncertain about whether a particular dosage of, say, and ace inhibitor will be right for post heart attack patient Joe Bloggs, or whether the proposed referral and admissions criteria for arthroplasty really will enable the waiting times to be brought down, or whether combining fee-per-service with salaries for clinicians will induce them to better practice, or whether – to remind ourselves yet again of the irony with which I began – an evidence-informed approach to policy decision-making is better than something else. '*Did* he do it?' is in the past tense. '*Will it be* cost-effective?' is in the future tense. The late lamented John Eisenberg observed that "law relies on evidence of the instance; health care relies on evidence of the generalizable" (Eisenberg 2001 p.375) and "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" (*loc. cit.* p. 372). So the legal concept of evidence will not serve here. I'm tempted to go further. The legal concept of evidence, at least as it seems to have been applied in the Chaoulli case, seems perversely to downgrade the authority of scientific evidence, a point which Charles Wright has recently argued cogently (Wright 2005)..

I am going to assume that, at root, evidence is anything that claims to be an empirical fact which gives a reason for believing that thing, or something to which it relates, like a consequence that might be reasonably expected to flow from it. It may be claimed falsely, in which case it is false evidence. It may be asserted as fact without any empirics, in which case it merely asserted and not evidence at all. Statements describing evidence are not the same as purely logical statements, which do not need to contain any empirics at all and are to be judged by the analytical canons of logical rather than empirical truthfulness. Nor is evidence normative, even though it is often important to have evidence about normative things like the values people actually hold and even though political views may motivate the uncovering of the evidence in question and even though normative desiderata may provide a motive for collecting such evidence.

When people in the clinical, management or policy world are asked what they consider as 'evidence', they typically come up with a complex mixture of both scientifically general and locally idiosyncratic types of information — what my colleagues and I have come to call a 'colloquial' evidence (Lomas *et al.* 2005, p. 7) They "draw on multiple sources and define evidence broadly" (Davidoff *et al*. 1995). Clinical or program effectiveness data compete with assertion (sometimes claimed to be expert assertion), cost-utility algorithms sit alongside political acceptability, and public or patient attitude data are combined with vivid recollections of personal encounters. "What ministers call 'evidence' is what they get from their constituents at their Saturday surgery" (WHO Health Evidence Network 2004, Saarni and Gylling 2004).

What the law calls ‘hearsay’ is met in medicine and elsewhere as ‘case studies’ and might be unkindly called, more generally, ‘gossip’. As my old mentor, the Canadian economist Harry Johnson once said “Even a compendium of gossip is till gossip” (in conversation). Despite the pejorative overtones, I do not think that we ought to be instantly dismissive of ‘gossip’ offered in evidence. For one thing, when there is an absence of scientific evidence – and such absences are endemic - ‘gossip’ (or case-studies, or personal experience accumulated over a lifetime of professional practice) may be all we have. Another ground for caution against premature dismissal is that bodies which we have to take seriously themselves take evidence of the colloquial kind seriously. NICE (the National Institute for Health and Clinical Excellence in England & Wales) groups evidence into four classes: research, clinical experience, patient experience and evidence on what the public thinks (current NICE website slides). ‘Experience’ looms large here and there must be a reason for its presence. Most of the evidence that is not research-based is available only through colloquial channels.

So let us consider for a moment what merit there may be in admitting all these kinds of evidence into our decision-making machinery. The examples contain various combinations of what might be called scientific evidence and what I have called ‘colloquial’. The things that are ‘scientific’ about scientific evidence seem to be twofold: on the one hand the type of evidence deemed relevant is defined by the scientific hypothesis or theory being tested and, on the other, the way in which the evidence is created (as in controlled experiments) or gathered (as in censuses, social surveys or the use of secondary sources) and the way in which it is interpreted (as in multivariate regressions). It is emphatically *not* the ‘things’

about which evidence is sought that give scientific evidence its distinctive character. This mistaken thought arises because we use ‘science’ in two rather distinct senses. One is topical, descriptive and conventional in terms of the objects and phenomena typically studied by scientists, whether super novae or congenital diseases. The other is analytical and characterizes the way in which the phenomena – any phenomena – are studied. It is perfectly possible for non-scientists to talk non-scientifically but sensibly about scientific phenomena and it is equally possible for scientists to talk scientifically and sensibly about non-scientific topics. What makes evidence scientific is the manner of study, not the objects studied. In particular, you can have – and you often need – scientific evidence on people’s ‘health’, their ‘welfare’, their ‘values’, their social commitments, and scientific evidence about such *non-scientific* entities (in the conventional sense I have just described) can be gathered and analyzed scientifically. This is a point worth emphasizing, on account of its being often overlooked by enthusiasts (amongst whom I count myself) for evidence-informed practice.

The enormous power of scientific evidence derives in part from our knowledge of what happens in its absence, and especially unscientific assertions are articulated by people imbued with the aura of science. Recall Dr Spock’s advice to countless thousands of mothers: “There are two disadvantages to a baby’s sleeping on his back. If he vomits, he’s more likely to choke on the vomitus. Also he tends to keep his head turned towards the same side, this may flatten the side of his head…. I think it is preferable to accustom a baby to sleeping on his stomach from the start” (cited in Chalmers 2003, p. 23). As Iain Chalmers has recently commented, reflecting on his early days as a medical practitioner,

"No doubt like millions of Spock's other readers, I passed on this apparently rational, theory-based and authoritative advice. We now know from the dramatic effects of the 'Back to Sleep' campaigns in several countries that the practice promulgated by well-intentioned experts like Spock led to tens of thousands of avoidable sudden infant deaths" (Chalmers, 2005, p. 229). Absence of science is no less dangerous when the unevidenced assertions are made about things economic, or social, psychological or ethical as well as things medical. I include ethics with purpose. Although ethics is not science it is often possible to use scientific method to gather qualitative and quantitative information about the values that people have, or that they think ought to imbue a policy. This is self-evidently so *a fortiori* with the economic, managerial and other contextual dimensions of decisions.

To colloquial evidence, then, let us add scientific evidence of two kinds: one relating mostly to medicine and the biological sciences and the other relating to the context for decisions and their consequences. In some cases, the context may be social or political; in others it may be clinical ('does this clinical guideline apply to my patient here and now?') (Woolf *et al.* 1999). My colleagues and I have termed the former 'context-free' evidence and the latter 'context-sensitive'. A classic example of context-free evidence is in the results of a randomized clinical trial or of a meta-analysis of trials. A classic example of context-sensitive evidence is the famous Rand experiment on people's behaviour under different health insurance packages (Manning *et al*. 1987). Good examples of thorough integrations of the two, in which the sciences of epidemiology and economics are applied to generate not only the evidence but also the analytical frameworks used to analyze it, can be

found amongst the dozens of competently performed cost-effectiveness and cost-utility analyses increasingly being done for publicly-funded health care and managed health care systems.

Ought we to rank these three types of evidence: context-free scientific evidence, context-sensitive scientific evidence and colloquial evidence? At one level, the answer has to be yes. The scientific evidence must be ranked above the colloquial as far as its dependability is concerned. But that does not mean that it must invariably trump the colloquial. It is perfectly possible – and perfectly justifiable - that colloquial evidence on a major aspect of a major issue should (if there is a conflict) trump scientific evidence on a trivial aspect of the same issue. Moreover, scientific evidence is often incomplete (there may be lots of information about the performance of drug A but little on its principal competitors B and C) and only partially relevant (keyhole surgery seemed to have efficacy in centres of excellence but nothing was known about the training needs of surgeons unfamiliar with the technique in ordinary hospitals – efficacy versus effectiveness). So, although it is a popular pastime of systematic reviewers, I think one ought to be cautious in the use of words like 'trumping', or 'hierarchy of evidence' in assuming that one kind of evidence is at all times and for all purposes superior to another. The central issue concerning scientific and colloquial evidence is not one of dependability. The scientific plainly dominates the colloquial in that regard – when it is available. The colloquial comes into its own when the scientific evidence is not available or is incomplete in particular and relevant respects. These will most characteristically relate to context-sensitive matters, on which there is typically much less scientific research than the context-free. So colloquial evidence comes

into play in a significant fashion when the issue in whether, say, a medical procedure works in general (as demonstrated in US trials), but whether it is likely to work in Canada, or in community hospitals; or if it is believed to work in such places, does it work well enough to warrant public funding; or if it were introduced this year, could the local services cope with the expected demand; or what will be the consequences for hospital waits and what hospital admissions criteria will be needed to filter demand according to agreed protocols; or, since this procedure happens to work much better for some patients than for others, which sub-groups, should gain access to it? All of these are things about which scientific research evidence could be collected. It is not irrational that much of it is not collected since, after all, research resources need to be conserved like any other. The point I want to emphasize is that it is context-sensitivity that will normally warrant the use of colloquial evidence. If the guidance that is the essential objective of a deliberative process is to be offered as helpfully and comprehensively as possible, thus making that guidance plausible and implementable, then the colloquial evidence has the two essential functions of providing relevant context for the context-free science and of filling in gaps in the knowledge base – gaps that could be filled by scientific evidence but which have not been.

So the issue confronting any decision maker within a deliberative process is not so much how to balance the three types of evidence or to assess the weight to place on each as to allow each to perform its appropriate task:

- Scientific context-free evidence to constitute evidence about general potential
- Scientific context-sensitive evidence to constitute evidence about the particular in realistic scenarios

- Colloquial evidence to fit the first into the second and supply the best evidence short of scientific evidence to fill in any relevant gaps.

What is a deliberative process?

A deliberative process both elicits and combines evidence. It elicits it by virtue of the embodiment of the participants in a deliberative process and it combines it in a process of 'weighing it up' and considering the contexts in which it is to be used. Deliberative processes also increase, it is conjectured, the likelihood of achieving what the philosopher of health and health care Norm Daniels calls "sound and acceptable decisions". Daniels argues for processes that "account for reasonableness" so that they have a moral authority over and above that which customarily attaches to market or bureaucratic processes (Daniels 2000a, 200b). Some advocates see the rationale for such processes 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" as Rychetnik put it (Rychetnik et al. 2004). Others have emphasized the open participatory nature of a deliberative process: the environmentalist Judith Petts talks of "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" (Petts 2004).

Deliberation is commonly seen as desirable whenever the issues at stake are debatable. Deliberative processes stress the integration of technical analyses of clinical issues with

scientific analyses of social contexts, within an explicit decision making model with clear criteria, and involving stakeholder and lay public consultation and even participation, in contrast to the more traditional top-down approach. 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 This entailed 47 community meetings, 12 public hearings and 54 panel meetings for health care 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*. So deliberative processes and consultative processes are not synonymous but deliberative processes characteristically *include* consultation.

However, if one is to appraise the effectiveness of deliberative processes, one needs tighter definitions than these somewhat vague ones, whose fulfilment might be assessed in part in descriptive procedural terms but whose ultimate impact on the quality of decisions (if, indeed, that *is* the appropriate ultimate test) necessitates the defining of 'quality' of decisions. Moreover, to be of much use one would also want to be able to identify which components of a variable mix of characteristics are those generating the effective impact on quality – if there is any such impact. We have here a classic example of a challenge that might never succumb to the disciplines of evidence-informed practice, for the combination of complex processes with complex and qualitative outcomes is a formidable challenge indeed.

Objective of a deliberative process

The challenge as described is, however, an unreasonable one. It is as well to rule out the most ultimate outcomes altogether – outcomes like impact on the health of populations or population groups, or the distributional equity of the system. The main reason is less the difficulties inherent in measurement as the issue of attribution. A deliberative process might be as excellent as it could possibly be but the ultimate outcome fail to emerge because of failures elsewhere in the system – unless the deliberative process in question is close to such an endpoint and is part of a process whose participants are accountable for the delivery of a product at the endpoint. Appropriate outcomes against which the achievement of processes like deliberative ones ought to be assessed are those for which the process is primarily responsible.

What *would* be reasonable is to set more manageable and realistic objectives for deliberative processes and, given these, to be much more selective and specific about the criteria by which any process is to be judged. I shall discuss the criteria in a moment. First, let us consider a reasonable objective for deliberative processes.

It has already been postulated that the purpose of a deliberative process is to help people to make decisions. We have also postulated that evidence-informed decisions are in general better than un-evidenced ones. We have categorized evidence into three types: context-free scientific evidence, context-sensitive scientific evidence and colloquial evidence. We have outlined the role of each and, in particular, singled out colloquial evidence as having a special role to play to fit context-free evidence into the relevant context, wherever possible

using context-free research evidence, and to supply the best evidence short of scientific evidence to fill in any relevant gaps. These ingredients take us to the heart of what a deliberative process might be reasonably regarded as being there to achieve: to provide guidance to health care decision-makers that is informed by relevant scientific evidence, interpreted in a relevant context wherever possible with context-sensitive scientific evidence and, where not, by the best available colloquial evidence, and that is, as a result, feasible and implementable. If there is an ultimate test of the effectiveness of a deliberative process it is thus *not* whether the guidance emerging from it is implemented but whether it is *implementable*; it is not whether the guidance is perfect but whether it is *as good as it needs to be* for the purposes in hand and, in particular, at least as good as any alternative process of producing guidance (or doing without it).

When should we use deliberative processes?

One may conjecture the circumstances under which the use of deliberative processes is most warranted. The following are all conjectured to be candidates:

Participation

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

Involvement of wider social and cultural issues

- Issues of outcome, benefits and costs that go beyond conventional boundaries
- Substantial uncertainty about key values and risks that needs to be assessed and weighed
- There are other social and personal values not taken into scientific account
- There are issues of equity and fairness
- There are issues of implementability and operational feasibility
- Wide public and professional 'ownership' is desired
- There is doubt about feasibility and implementability

Scientific evidence on effectiveness is often summarized in the form of a narrative review, a systematic review, or a meta-analysis; scientific evidence on context, including values, might be gathered by controlled experiments, which in turn may also be summarized and synthesized; and colloquial evidence is often gathered through consultative processes including social surveys, public meetings and the hearing of witnesses, as well as directly from those participating in the deliberative process. Within this cascade of evidence are frequently embodied further deliberative processes (for example, during the process of systematic reviewing, when reviewers need to reach agreement of such matters as search terms, search engines, inclusion/'exclusion criteria, quality criteria and the like). Plainly, this is a more restricted form of deliberation, for the most part restricted to a select coterie of experts.

A good example of a deliberative process in which all the features mentioned so far are present is the method used by the National Institute of Health and Clinical Excellence (NICE) for evaluating health care technologies in England and Wales. There are formal submissions from interested parties, consultations and invited commentaries from consultees and commentators (between which NICE distinguishes carefully), systematic reviews, technical modelling exercises to cover some lacunae in the scientific evidence, multi-party representation in the (large) deliberative committee which hears witnesses, appeal possibilities, and various support groups using consensus methods on controversial issues of value and advising on whom to consult (Culyer 2005) .

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 (NICE 2004) embodied within it which was itself reaffirmed through a further specific deliberative process.

As an example of the algorithmic approach, NICE recommends the use of Quality-Adjusted Life-Years (QALYs) as the main outcome measure in the economic appraisals 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 five health state characteristics each measured on a three point scale and added together. The guidance explicitly states that an additional QALY should receive the same weight regardless of the other characteristics of the individuals receiving the health benefit. The selection of the five characteristics is value-judgmental, the relative weights assigned to them in calculating a

QALY are value-judgmental and the interpersonal comparison is value-judgmental. Incidentally, the algorithm is built upon prior scientific experiments and social surveys of the UK population.

As an example of a deliberative process, NICE referred one aspect of the QALY algorithm to its Citizens' Council, a form of consensus group to engage in a deliberative approach. This was a question regarding the weighting (if any) to be given to older people: should it be higher or lower than for other groups, contrary to the rule just cited as part of the algorithm? Their recommendations to the NICE Board included this: "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" (NICE 2004b)

Criteria for judging the effectiveness of deliberative processes

The ultimate product of a deliberative process is guidance shaped by judgment – judgment about an effect of doing something, its size, the ways in which it is likely to be achieved, for whom, for how long, its cost in terms of the resources used that would otherwise have been employed in other ways to achieve other good things, and how worthwhile it is. The test, therefore of a deliberative process is whether the resultant judgment is (or will be) more comprehensively 'evidence-informed', better matched to the context of application, more efficiently implementable and more widely acceptable to those affected by it.

Some evidence on deliberative processes

It all sounds so plausible! But, then, so did Spock. Is there any evidence that deliberative processes actually *work*? The short answer is 'not much'. A lot of the literature on deliberative processes has been and continues to be essentially advocacy rather than reports of the effectiveness of well-defined processes (e.g. Gibson *et al*. 2005). The evidence is also understandably less rigorous than that for more narrowly defined and experimentally controllable scientific matters of cause-and-effect. It is mostly qualitative and judgmental. There is virtually no replication of studies. Some evidence exists about designs of processes that are more or less likely to generate consensus, which have identified important roles for what we have called colloquial evidence and which have identified some reasons underlying failures to agree within groups and failures to heed context-free evidence – which include doubts about its applicability in particular context (an outstanding example showing all these characteristics is Raine *et al*. 2004). There is evidence that the provision of context-free scientific evidence tends to promote consensus, especially within professional panels (Lomas *et al*. 1988).

In comparing a deliberative process with other methods one author has written that: "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." (Kim 2002). Unfortunately, while all these outcomes 'can' or 'may' be true, they are not themselves evidenced 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’.

Although the evidence is not strong the following generalizations are probably not far from the mark: with a deliberative process

- Decision makers acquire a better grasp of the strengths and weaknesses of the underlying cases and can better defend their decisions
- Consensus building is enhanced
- The revelation of evidence gaps helps to inform downstream research programs
- Stakeholders and their peers are more likely to accept and implement decisions that they have had a hand in shaping
- Possible selection bias through the membership of decision-making panels becomes relatively obvious
- Context-free evidence can be re-interpreted in relevant contexts.

That is about as far as I would dare to go in claiming that there is evidence in the literature for the effectiveness of deliberative processes. However, never mistake absence of evidence for evidence of absence! My theoretical conjecture is still that deliberative processes, used in suitable situations, are preferable to other methods, and that the reasons for this conjecture are that the objectives of deliberative processes are better objectives than those of others and. That well-designed empirical research will bear this out.

I commented earlier on the formidable challenge posed by the evaluation of a process that is itself a combination of complex processes further complicated by complex and mostly qualitative outcomes. It is not, however, a challenge before which we can only lie supine and helpless. How, then might one best rise to it?. The key, I suggest, is by being suitably modest both in what we seek to establish empirically and what we claim for the process. 'Being modest' in this context means asking researchable questions and not applying criteria of completeness and rigour that we do not apply consistently everywhere else. To paraphrase Voltaire, let's not allow the perfect to become the enemy of the merely good[2]. In particular, let us not even begin to seek to evaluate the effectiveness of deliberative processes in the same way as we would assess the cost-effectiveness of a pharmaceutical.

I propose that we tackle the problem in a reductionist fashion by applying three principles:

- Select key outcomes of a deliberative process
- Select key characteristics of a deliberative process
- Have an explicit alternative process as comparator.

The first of these simplifies the complexity at the endpoint, the second identifies the variables to be investigated and the third ensures that we are not implicitly comparing a deliberative process with some unstated but ideal process which either does not exist or which is not a realistic alternative. Let us take each in turn.

---

[2] Le mieux est l'ennemi du bien. *Contes* (1772). This seems to be derived from an Italian proverb

Outcome selection

Based on our foregoing discussion, I suggest five principal outcomes which, taken together might be deemed suitable measures of success or failure of a deliberative process:

- A deliberative process is more likely to generate guidance that is consistent with the context-free scientific evidence set in a relevant context – relative to a comparator process
- A deliberative process is more likely to identify relevant clinical, social and political contexts for interpreting context-free scientific evidence – relative to a comparator
- Conversely, the quality and power of residual opposition to any guidance will be low
- A deliberative process is more likely to command a wide credibility in professional circles and beyond – relative to a comparator
- A deliberative process is more likely to identify impediments to the implementation of guidance and to propose solutions.

Process characteristics

Since deliberative processes, like other processes, have many components and are not uniquely defined by any particular combination and degree of each, it makes sense to select those aspects of the process that are conjectured to have an impact on the outcome and its size and its assessed value. Here there is likely to some overlapping with the outcome categories. Factors that may be contenders might be categorized as research quality, reasonableness of procedures and membership. They might include:

Research quality – relative to a comparator

- Quality of the scientific research (both context-free and context-sensitive) available at start of process and subsequently
- Quality of the colloquial evidence
- Availability of meta-analyses and systematic reviews of scientific evidence
- Clarity of the questions to be answered by the process
- Adequacy of the scoping of the questions to be answered
- Availability of research into the public's views on contextual and other 'non-scientific' matters of relevance
- Quality/quantity of support staffing

Reasonableness of procedures – relative to a comparator

- Quality of chairperson
- Clarity and openness of process
- Reasonableness of time-lines for submission and consideration of evidence
- Use of colloquial evidence to challenge context-free evidence, set contexts and plug gaps in science (but not to supplant scientific evidence of either kind)
- Availability of time for study, discussion and reflection before during and after 'meetings'
- Scope of opportunities for all interested parties to comment during the process
- Scope for members to request further information and take face-to-face oral evidence

- *In camera* discussions when desirable to encourage free expression of opinion; otherwise as open and transparent as possible (e.g. public sharing of agendas, data, publication of minutes)
- Opportunity for and use of appeal if process is flawed or decision appears unreasonable.

Membership – relative to a comparator

- Representativeness of expertise in the relevant scientific evidence among panellists
- Representativeness of breadth of colloquial sources of evidence
- Participation of respected people from the major communities of interest
- Willingness of members to share values openly
- Inclusivity of stakeholder consultation - opportunities for all affected parties to be represented.

Comparators

The principles applying to the selections of an appropriate comparator or comparators are similar to those used to select comparators in health technology assessment. They should be relevant in context and assessable in the same way as proposed for deliberative processes. They should be informed by policy considerations. Examples include comparator processes that:

- are the most likely alternatives to the deliberative process in question
- are the *status quo* (where this is not the deliberative process in question)
- are much less costly alternatives
- other deliberative processes having a different mix or balance of characteristics.

Neither relevance nor practicality require more than relative assessment – comparing a deliberative process with a non-deliberative one or one kind of deliberative process with another. It is much less taxing task than the comprehensive assessment of the merits of deliberative processes in general or of any one in particular. So far as I am aware, no one has ever done this in the way proposed here, with these criteria for success and in comparative terms, so the field lies open for a reputation to be made. It is also more informative to the designers of processes to know the impact of different features of the design on the character of the deliberations and their eventual outcomes. Studies of this sort could also be prospective and, on occasion side-by-side, analogous to head-to-head clinical trials.

Envoie

The application of the basic methods of health technology assessment to non-clinical technologies, of which decision-making processes are one important class, has barely begun. I hope to have shown that it is easy to do studies of this sort poorly but that it is also easy to overstate the difficulties of doing them well. You may not discover the general theory of everything in this way, and you may have to put up with merely learning about what is good compared with what is perfect but, if you can achieve thereby better decisions, and better decisions that are generally recognized as such, and in addition find processes that do not cost the earth by eliminating costly features that add little to outcomes, then as a health service researcher you will have probably contributed more to human welfare than is

customarily given to such researchers to contribute. What more could anyone ask of any researcher?

References

Chalmers, Iain (2003) Trying to do more good in policy and practice: the role of rigorous, transparent, up-to-date evaluations, *Annals of the American Academy of Political and Social Science*, 589, 22-40.

Chalmers, Iain (2005) If evidence-informed policy works in practice, does it matter if it doesn't work in theory? *Evidence & Policy*, 1, 227-242.

Culyer Anthony (2005) Involving stakeholders in health care decisions – the experience of the National Institute for Health and Clinical Excellence (NICE) in England and Wales. *Healthcare Quarterly*, 8, 56-60.

Daniels, Norman (2000a) Accountability for reasonableness in private and public health insurance, in Angela Coulter and Chris Ham (eds.) *The Global Challenge of Health Care Rationing*, Buckingham/Philadelphia: Open University Press, 89-106.

Daniels, Norman (2000b) Accountability for resonableness, *British Medical Journal*, 321, 1300-1301.

Davidoff F., Haynes Brian, Sackett David, Smith, R. (2005) Evidence based medicine, *British Medical Journal*, 310, 1085-1086.

Eisenberg, John (2001) What does the evidence mean? Can the law and medicine be reconciled? *Journal of Health Politics, Policy and Law*, 26, 369-381.

Gibson, J. L., D. K. Martin, P.A. Singer (2005*)* Evidence, economics and ethics: resource allocation in health services organizations. *Healthcare Quarterly*, 8, 50-9

Kim, M-S. Cloning and deliberation: Korean consensus conference. *Developing World Bioethics*, 2, 159-72.

Lomas, Jonathan, Geoffrey Anderson, Murray Enkin, Eugene Vayda, Robin Roberts and Betsy MacKinnon (1988) The role of evidence in the consensus process: results from a Canadian consensus exercise, *Journal of the American Medical Association*, 259, 3001-3005.

Lomas, Jonathan, Tony Culyer, Chris McCutcheon, Laura McAuley and Susan Law (2005) *Conceptualizing and Combining Evidence for Health System Guidance*, Ottawa: Canadian Health Services Research Foundation.

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, Arleen Leibowitz and M. Susan Marquis, 'Health insurance and the demand for medical care:

evidence from a randomized experiment', *The American Economic Review*, 77, 3, 1987, pp.251-277.

NICE (2004a) *Guide to the Methods of Technology Appraisal*, London: National Institute for Health and Clinical Excellence.

NICE (2004b) *NICE Citizens Council Report on Age*, London: National Institute for Health and Clinical Excellence.

Petts, J. (2004) Barriers to participation and deliberation in risk decisions: evidence from waste management, *Journal of Risk Research*, 7 115-33.

Raine, Rosalind, Colin Sanderson, Andrew Hutchings, Simon Carter, Kirtsen Larkin and Nick Black (2004) An experimental study of determinants of group judgments in clinical guideline development, *The Lancet*, 364, 429-437.

Rychetnik, l., P. Hawe, E. Waters, A. Barratt, and M Frommer (2004) A glossary for evidence based public health, *Journal of Epidemiology and Community Health*, 58, 538-45.

Saani S. I., Gylling H. A. (2004) Evidence based medicine guidelines: a solution to rationing or politics disguised as science? *Journal of Medical Ethics*, 30, 171-175.

WHO Health Evidence Network (2004) *Evidence of Health Needs and Intervention*, Copenhagen: WHO Regional Office for Europe.

Woolf, H., R. Grol, A. Hutchinson, Martin Eccles and Jeremy Grimshaw (1999) Potential benefits, limitations and harms of clinical guidelines, *British Medical Journal*, 318, 527-530.

Wright, Charles (2005) Different interpretations of "evidence" and implications for the Canadian healthcare system, in Colleen Flood, Lorne Sossin and Kent Roach (eds.) *Access to Care, Access to ustice: the Legal Debate over Private Health Insurance in Canada*, Toronto: University of Toronto Press.