CHE
Centre For Health Economics

Part of
UNIVERSITY of York

Deliberation for Better Decisions in Health Technology Assessment: Conjectures and Tests

Ahmet Küçükuncular and Anthony J Culyer

CHE Research Paper 202

Deliberation for Better Decisions in Health Technology Assessment: Conjectures and Tests

[a] **Ahmet Küçükuncular**
[b] **Anthony J Culyer**

[a] Faculty of Economics and Administrative Sciences, Near East University, Nicosia; Department of Economics and Related Studies
[b] Centre for Health Economics, University of York

May 2026

Background to series
CHE Discussion Papers (DPs) began publication in 1983 as a means of making current research material more widely available to health economists and other potential users.

The CHE Research Paper series takes over that function and provides access to current research output via web-based publication. The RPs can take various forms and results and ideas do not necessarily represent the final position and may include work in progress not subject to peer review at the time of publication.

CHE RP 202 is an early version of a paper intended for later submission to an academic journal.

Further copies
Only the latest electronic copy of our reports should be cited. Copies of this paper are freely available to download from the CHE website www.york.ac.uk/che/outputs/. CHE Research Papers can be reused and distributed for non-commercial purposes provided the source is properly cited.

Please use https://doi.org/10.15124/yao-xt97-3x42 if you reference this paper.

Centre for Health Economics
Alcuin College
University of York
York,
YO10 5DD, UK

www.york.ac.uk/che

Abstract

Health technology assessment increasingly relies on formal tools such as cost-effectiveness analysis, Quality-Adjusted Life-Years (QALYs), thresholds and structured evidence reviews. These tools are indispensable, but they cannot by themselves settle questions of relevance, interpretation, fairness, implementation or public justification. We argue that deliberative processes in health technology assessment matter not because deliberation is inherently virtuous, nor because algorithms are unhelpful, but because decisions about health technologies involve heterogeneous evidence, contested values, uncertainty and context-sensitive trade-offs that cannot responsibly be resolved by algorithm alone. The paper is organised around two linked tasks. The first is conjectural: to identify when deliberation is likely to be useful by distinguishing it from algorithmic reasoning, consultation and commenting, and by explaining why evidence does not combine itself into guidance. The second is evaluative: to propose tests for judging whether deliberation improves, or is likely to improve, decision quality. On our account, a better decision is one that is more comprehensively evidence-informed, better matched to the context of application, more efficiently implementable and more widely acceptable to those affected by it, whether positively or negatively. We further argue that deliberation may be ethically constitutive of legitimacy in public healthcare decision making, since allocation decisions require reasons that are public, contestable and defensible. We conclude by setting out a future agenda concerned with comparing deliberative and alternative processes, assessing their cost-effectiveness, building case-based institutional memory, improving stakeholder involvement and making anonymous opportunity cost victims more visible.

Keywords: Health technology assessment; Deliberative processes; Evidence-informed decision making; Health care priority setting; Cost-effectiveness analysis; Public reason; Legitimacy; Resource allocation.

I CONJECTURES

1 INTRODUCTION: FOR WHAT MAY DELIBERATIVE PROCESSES IN HTA BE GOOD?

Health technology assessment (HTA) increasingly depends on formal methods for making difficult allocation decisions explicit. Cost-effectiveness analysis, QALYs, thresholds, structured evidence reviews and decision criteria are indispensable tools. They discipline judgement, expose trade-offs and help decision makers avoid arbitrariness. But they do not remove the need for judgement. Decisions about health technologies involve heterogeneous evidence, uncertainty, contested values, institutional constraints, professional responsibilities and consequences for patients, publics and anonymous others whose care may be displaced. The question, therefore, is not whether HTA should use formal tools or deliberative processes. It is whether, when and how deliberation can improve the quality of decisions that formal tools alone cannot settle.

This is not the usual political case for deliberation. Burke's account (1958) of a deliberative assembly, for example, emphasised the restraint of sectional interest in favour of judgement directed towards the whole. Habermas's theory of deliberative democracy, by contrast, links deliberation to communicative rationality, democratic legitimacy and the public justification of law and policy (Habermas, 1998). Both traditions matter because they show why deliberation has often been treated as politically and morally significant. Our concern, however, is narrower and more institutional. We ask whether deliberation can improve decision making in HTA by helping decision makers interpret evidence, expose assumptions, weigh values, assess context and justify recommendations. The claim is therefore not that deliberation is good because it is democratic, but that it may be good when, and because, it improves decisions.

For some, that might be enough. There are doubtless those for whom deliberation and all that it implies is inherently a good thing. And some would seem to think of it as a veritable cornucopia of political good things: more civic engagement, more tolerance, better reasons, enhanced legitimacy, greater understanding of opposing points of view, more public trust, and decisions that are more considered and more informed (cf. Carpini et al., 2004; Chambers, 2003; Fishkin, 2011; Gutmann & Thompson, 2004). We record no curmudgeonly dissent regarding the desirability of any of these things, nor from the hope that they may indeed sometimes be the fruits of deliberative processes. But they are not the essence of our present argument. Nor is our purpose to contribute directly to the extensive literature following Habermas's (1998) Theory of Deliberative Democracy, in which deliberation is closely tied to communicative rationality, democratic legitimacy and the public justification of law and policy.

Burke and Habermas both matter because they show why deliberation has often been treated as politically and morally significant. Our concern, however, is narrower and more institutional. We ask whether deliberation can improve decision making in HTA by helping decision makers interpret evidence, expose assumptions, weigh values, assess context and justify recommendations. The claim is therefore not that deliberation is good because it is democratic, polite, inclusive or morally attractive in itself, but that it may be good when, and because, it improves the quality of institutional decisions.

1.1 A more fundamental case for deliberation

The essence of our point is that a rather strong case can be made for deliberative processes on grounds that they lead to better decisions, regardless of whatever other desirable outcomes may be their consequences. That immediately raises two questions. What is a better decision? And how would one recognise one if one saw one? These are not idle questions. In HTA, decisions are commonly taken under conditions of uncertainty, conflicting evidence, institutional constraint, and value disagreement. They are decisions about technologies, but they are also decisions about patients, publics, opportunity costs, professional authority, and the uses to which evidence may properly be put.

HTA has increasingly become one of the principal institutional sites in which modern health systems attempt to make such decisions explicit. It is a place where cost- effectiveness, clinical effectiveness, budget impact, equity, feasibility, need, severity, public acceptability, and political accountability all press upon the decision maker at once. It is therefore unsurprising that recent work has emphasised that HTA is not merely a technical exercise but an explicitly value-based activity requiring normative reasoning as well as empirical evidence (Charlton et al., 2024). Nor is it surprising that contemporary guidance on deliberative processes in HTA defines them as procedures and activities that support the informed and critical examination of an issue and the weighing of arguments and evidence to guide decision making (Oortwijn, Husereau, et al., 2022). That formulation sits comfortably with the older claim that deliberative processes in health care are to be judged by whether they produce guidance informed by relevant scientific evidence, interpreted in context wherever possible, and supplemented where necessary by the best available forms of contextual and colloquial evidence (Culyer & Lomas, 2006).

In stark contrast to the deliberative process stands the algorithm. An algorithm is a systematic mathematical process sequentially linking various strands in a decision problem to an outcome. A good example for present purposes is the Quality-Adjusted Life-Year (QALY), which combines a set of predefined characteristics of good health, measurable at a variety of intensities, and weighted in a preset fashion in order to measure a health outcome. The algorithm can be made as complicated as is desired, at least in principle, by adding characteristics, refining intensities, changing weights, including probabilities and uncertainty, discounting future health changes, modifying the threshold against which results are judged and so on. Every element of the algorithm may also be moderated by consultation with patients, clinicians, economists, or members of the public. The process remains, however, mechanical and unidirectional. It is not deliberative. Consultation is not deliberation.

We are not against algorithms. Indeed, health economics would be a poorer discipline without its capacity to structure difficult choices and make trade-offs visible. The National Institute for Health & Care Excellence's (NICE's) use of cost-effectiveness has long been defended not as a mindless application of a formula, but as part of a broader process in which scientific and social value judgements are made explicit, contested, and justified (Culyer, 2006; Rawlins & Culyer, 2004). Nor is there any virtue in leaving hard choices to intuition, rhetoric, or the accidental distribution of influence. A good algorithm may be an excellent aid to thought. But an aid to thought is not a substitute for thought. Cost-effectiveness analysis, QALYs, severity modifiers, thresholds, and multi-criteria decision tools may all assist choice; they do not make choices, nor do they relieve decision makers of the responsibility of explaining why one claim was accepted and another rejected.

This distinction matters because the central difficulty in HTA is not merely that decisions are complex. The mere complexity of the elements of a decision is neither a necessary nor a sufficient ground for supposing that deliberation is called for. Some decisions may be technically complex yet morally routine, while others may be technically simple yet ethically troubling. A single threshold rule may seem simple, but beneath it may lie controversies about evidence quality, generalisability, opportunity cost, equity, clinical significance, unmet need, public confidence, and the scope of legitimate exception. Conversely, a formal multi-criteria model may appear more capacious than a single threshold, but it too may require deliberation if decision makers need to challenge its structure, its weights, its assumptions, its evidential base, or the values smuggled into its apparent neutrality.

We therefore begin from a deliberately modest claim. We are not advocating the indiscriminate use of deliberative processes. They are costly. They may not be appropriate. Even where appropriate, they may not be worth their cost. Indeed, we are not advocating anything, apart from good decision making. Our argument is that there is a reasonable expectation that deliberation can enhance the quality of decisions in HTA and that it is possible to make coherent conjectures about the circumstances in which this is likely to be so, without endowing the process with any further claims of virtue.

The need for such caution has only increased as HTA has travelled across institutional settings and jurisdictions. Global advocacy for HTA has rightly emphasised the value of systematic evidence, explicit priority setting, and transparent decision making, but it has also shown that the local realities of governance, legitimacy, capacity, and political judgement cannot be bypassed by more evidence alone (Chalkidou et al., 2016; Charlton et al., 2024). Recent studies of stakeholder engagement likewise suggest that involvement is often justified by reference to better evidence, improved interpretation, and more acceptable decisions, but that these aims are unevenly specified and unevenly realised in practice (Carter et al., 2025). The point is not that all stakeholder engagement is deliberation. Much of it is not. The point is that the same practical problem keeps recurring: evidence must be interpreted, values must be made explicit, and decisions must be justified in ways that can survive scrutiny by those who disagree with them.

This is also where the argument connects with extra-welfarism and with refined utilitarianism. Extra welfarism in health economics made it possible to ask not only how much welfare is produced in the usual preference satisfaction sense, but what forms of benefit, need, equity, and health-related capability ought to matter in public decision making (Brouwer et al., 2008; Culyer & Wagstaff, 1993). Refined utilitarianism, as recently formulated in relation to the NHS, similarly treats rules, public trust, legitimacy, and context as central to the welfare consequences of institutional practice rather than as external constraints imposed upon it (Küçükuncular, 2026). In this view, deliberation is not merely a procedural supplement to optimisation; it is one of the institutional settings in which evidence, values, and public reasons are brought into relation with one another. In contexts where difficult allocation decisions must be justified to those who will live with their consequences, deliberation may be part of what makes those decisions publicly defensible.

The argument of this paper is therefore as follows. Deliberative processes in HTA are needed not because algorithms are useless, nor because deliberation is self-evidently virtuous, but because decisions of this kind involve heterogeneous evidence, value conflict, uncertainty, and context sensitive trade-offs that cannot be responsibly resolved by algorithm alone. A deliberative process is valuable when it prevents three common errors: treating consultation as though it were itself decision making, treating evidence as though it mechanically yielded judgement, and treating technical appraisal as though it were sufficient for public justification. Its value should be judged not by the warmth of the process, nor by whether all participants emerge in agreement, but by whether the resulting decision is better. We suggest that a better decision is one that is more comprehensively evidence-informed, better matched to the context of application, more efficiently implementable, and more widely acceptable to those affected by it, whether positively or negatively. These are conjectures, not dogmas but they provide a starting point for asking whether deliberation in HTA is more than merely good manners, and whether it can be deemed a genuine contributor to good decision-making.

The remainder of the paper develops this case in stages. First, we distinguish deliberation from algorithmic reasoning, consultation and commenting. Second, we examine why HTA decisions require the interpretation and combination of several kinds of evidence rather than the mere aggregation of data. Third, we ask when deliberation is likely to be useful and when it is likely to be unnecessary or wasteful. Fourth, we develop criteria for judging whether deliberation has improved the quality of a decision. Finally, we consider the institutional characteristics that make deliberative processes more likely to succeed, including the quality of evidence, the design of the process, the composition and conduct of the membership, and the possibilities for accountability, appeal and learning by doing.

2. DELIBERATION, ALGORITHM, CONSULTATION AND COMMENTING

If deliberative processes in health technology assessment are to be defended, they must first be distinguished from other procedures with which they are often casually, and sometimes misleadingly, conflated. The word deliberation is one that acquires an aura before it acquires a definition. It may

be used approvingly to describe almost any process in which more than one person appears to have considered more than one consideration. That will not do. In HTA especially, where evidence is gathered, consulted upon, commented upon, modelled, weighed, filtered and finally translated into guidance, precision about what counts as deliberation is indispensable.

A deliberative process entails the careful, deliberate consideration and discussion of the advantages and disadvantages of various options. It is not inherent in the concept that deliberation must always be face to face, though the possibility of interaction remains central. What matters is not physical co-presence as such, but the opportunity for informed exchange in which claims, evidence, assumptions and values may be challenged, defended, revised and, if need be, abandoned. Recent guidance from the HTAi and ISPOR task force puts the matter well: deliberation in HTA consists in the informed and critical examination of an issue and the weighing of arguments and evidence to guide a subsequent decision (Oortwijn, Husereau, et al., 2022). That formulation is useful because it captures both the epistemic and the practical dimensions of the process. Deliberation is not mere discussion. It is discussion directed towards judgement.

In stark contrast to the deliberative process stands the algorithm. An algorithm is a systematic mathematical process sequentially linking various strands in a decision problem to an outcome. A familiar example for present purposes is the QALY, which combines a set of predefined characteristics of health, measurable at varying intensities and weighted in a preset fashion to generate a metric for use in decision making. No matter how complex the algorithm becomes, the process remains mechanical and unidirectional. It is not deliberative. Consultation is not deliberation. A cost-effectiveness model may be an input into deliberation, a subject of deliberation, or indeed an object of dispute within a deliberative process. But it is not itself deliberation. The role of an algorithm is to be an aid to thought, not a substitute for it. An account of NICE's appraisal procedures likewise makes clear that cost-effectiveness analysis, far from replacing judgement, functions within a broader institutional process in which scientific evidence and social value judgements must be interpreted and justified (Culyer, 2006).

This distinction is not merely rhetorical; it has practical consequences. An algorithm may rank options, summarise consequences, or expose trade-offs that might otherwise remain hidden. For these reasons it is often indispensable. But it cannot decide whether the model has used the right inputs, whether the chosen threshold is normatively acceptable, whether the relevant comparator has been correctly identified, whether uncertainty is tolerable, whether transferability across settings is plausible, whether severity or unmet need warrant exception, or whether the resulting guidance can be publicly defended. These are matters of judgement. An algorithm may form an element within a deliberation, or even be the subject of one, but it is not itself deliberative. Recent discussions of normative reasoning in HTA have therefore emphasised that formal methods do not displace evaluation at the level of reasons. On the contrary, they make that evaluation more pressing, because apparently technical results often conceal unresolved disputes about value, fairness and relevance (Charlton et al., 2024). A model may present itself as the end of debate when it is in truth only the beginning of one.

Nor should deliberation be confused with consultation. Consultation is often an important part of a good decision process. It may gather evidence, identify interests, reveal neglected considerations, expose likely objections, and widen the evidential base on which decision makers proceed. But consultation is not deliberation. In the early 1990s the US state of Oregon conducted a public consultation to help determine decisions about the benefits to be included in the state's Medicaid program. There were 47 public meetings, 12 hearings and 54 panel meetings which generated information, testimony and opinion. Any deliberation occurred elsewhere, namely in the body charged with making the decision (Culyer, 2009; Fox & Leichter, 1991; Oregon Health Services Commission, 1991, p. 21). Those consulted were not themselves participants in the actual decision-making but were important providers of material subsequently used within it. Recent reviews of stakeholder engagement in HTA show that consultation with stakeholders, including

patients, industry and clinicians, is widespread, but that such engagement often occurs through submissions, hearings or opportunities for comment rather than through participation in the actual weighing of reasons that leads to guidance (Carter et al., 2025). Consultation may enrich deliberation; it does not by itself constitute it.

There is also commenting. NICE provides opportunities for people to comment on draft scopes, draft recommendations and consultation documents. These comments may sometimes be very important. They may uncover errors, correct misunderstandings, introduce new evidence, or expose the practical consequences of a proposed recommendation. Yet again, the activity is not itself deliberative. Commenting, like consultation, tends to involve limited interchange, uneven participation, and little or no shared responsibility for the eventual judgement. It is an input into the process rather than the process itself. This is not to belittle it. On the contrary, a good deliberative process may depend heavily upon consultation and commenting if it is to avoid insularity or professional self-enclosure. But to say that such procedures are valuable is not to say that they are deliberative. A deliberative process can, unlike these two types of procedure, be intrinsic to the actual decision-making process. That is what makes deliberative processes critically different.

The point may be sharpened further. Deliberation is not merely a larger quantity of consultation, nor is it commenting with better manners. It is a distinct institutional practice with its own logic. It requires the possibility that participants may change their minds as evidence is adduced, assumptions exposed, and competing values articulated. It requires that what participants bring to the process, their expertise, experience, interests, doubts and values, may count in the weighing of the case. And it requires some degree of common orientation to judgement rather than simple advocacy. Oortwijn et al. (2022) are right to insist that deliberative processes in HTA involve more than a single meeting or event. They include the structures, procedures and activities by which deliberation is supported across the process. But that broader framing should not obscure the central distinction. A process is not deliberative merely because others are invited to speak. It is deliberative when reasons are exchanged under conditions that make mutual scrutiny possible and consequential.

It follows that the argument for deliberation in HTA is not an argument against quantification, evidence synthesis or public input. It is rather an argument against supposing that any of these, singly or together, can relieve decision makers of the need to judge. Cost-effectiveness analysis may discipline intuition. Consultation may widen the evidential field. Commenting may improve the quality of draft recommendations. All this is welcome. Yet none of it answers the question of what ought to count, how heterogeneous evidence is to be interpreted, how competing values are to be weighed, or why one recommendation rather than another should ultimately stand. Those are the questions to which deliberation is addressed. It is precisely because HTA routinely confronts such questions that consultation is not enough, an algorithm is not enough, and commenting is not enough. Deliberation is needed where judgement must be exercised and justified.

This is why the distinction matters for the structure of the paper that follows. If deliberation were simply another name for consultation or a more sociable wrapper around algorithmic appraisal, there would be little more to say. But if deliberation is the setting in which evidence, values and reasons are brought into relation under conditions oriented towards judgement, then it becomes possible to ask the more serious questions: what kinds of evidence require such a process, when is it likely to be useful, what makes it succeed or fail, and how might one tell whether it has led to a better decision. It is to those questions we now turn.

3. WHY HTA NEEDS DELIBERATION: EVIDENCE DOES NOT COMBINE ITSELF

A deliberative process can be a mechanism for eliciting, combining and interpreting evidence. That statement is deceptively simple, since almost everything turns on what is meant by “evidence”, whose evidence is to count, how its relevance is to be judged, and how evidence of different kinds is

to be brought to bear upon a decision that must finally be made. In HTA, these questions are not incidental; they are the ordinary material of decision making.

At a basic level, evidence may be understood as facts, actual or asserted, intended for use in support of a conclusion. But decision-makers do not all mean the same thing by this. Most decision-makers view evidence colloquially and eclectically, as anything that increases their degree of belief in a fact or proposition relevant to the decision before them. They define it by its resonance with experience and its relevance to the kinds of decisions they are required to make. Scientists, by contrast, characteristically view evidence as the product of systematic and replicable methods of research investigation, and they tend to classify it by method, design, discipline, and its contribution to scientific understanding. These are not merely differences of vocabulary. They are differences in the conventions by which claims are made credible.

It is helpful to think of three broad forms in which evidence may come: colloquial evidence, context-free scientific evidence, and context-sensitive scientific evidence (Culyer & Lomas, 2006). Context-free scientific evidence is readily associated with the traditions of evidence-based medicine: randomised trials, systematic reviews, meta-analyses, controlled comparisons and other designs intended to establish, as far as possible, whether an intervention works under specified conditions. Context-sensitive scientific evidence concerns the circumstances in which an intervention may be introduced, sustained, afforded, organised, accepted or resisted. It includes, for example, evidence about implementation, organisational capacity, local epidemiology, workforce availability, financial feasibility, health system structure, equity effects, public attitudes and ethical constraints. Colloquial evidence embraces professional experience, clinical judgement, patient and carer testimony, managerial knowledge, political judgement, social values, local habits and traditions, pressure group claims, and the particular pragmatics and contingencies of the situation. It is not always scientific evidence, but it is not for that reason always irrelevant. Sometimes it is precisely the sort of evidence without which a technically correct recommendation becomes practically useless.

These distinctions are not intended to create a hierarchy in which one kind of evidence permanently dominates the others. Rather, they clarify why HTA decisions cannot be made by evidence accumulation alone. The appropriate methods for obtaining scientific evidence about context factors are not the same as those for obtaining evidence on programme or intervention effectiveness. That does not make such evidence less important. A technology may be efficacious but not implementable. It may be cost-effective in one jurisdiction but not in another. It may look persuasive in a model while resting on assumptions about uptake, adherence, comparators or service capacity that are locally implausible. It may deliver aggregate health gain while raising serious concerns about fairness, severity, disability, unmet need or public trust. In all these cases, the question is not merely whether evidence exists, but what sort of evidence it is, what role it should play, and how far it travels.

This is one reason why evidence-informed deliberative processes have become so important in the recent HTA literature. Baltussen, Jansen and Oortwijn (2022) describe such processes as a practical approach for enhancing the legitimacy of benefit package design by bringing together stakeholder involvement, evidence-informed evaluation, transparency and appeal. They similarly emphasise that evidence-informed deliberative processes are not simply a way of adding more voices to an appraisal, but a way of identifying, reflecting on, and learning about the meaning and importance of values informed by evidence on those values (ibid.) Evidence needs a process because evidence does not arrange itself into guidance.

In practice, the range of ideas about what constitutes evidence is likely to be very wide. It may be defined by method of collection (observational, experimental, quasi experimental, extrapolated, survey, experiential or administrative). It may be defined by disciplinary or theoretical framework, such as quantitative, qualitative, economic, ethical, legal, epidemiological, clinical, decision analytic or philosophical evidence. It may be defined by purpose, such as problem identification, description,

scoping, cost containment, efficacy, effectiveness, cost-effectiveness, implementability, cultural fit, leadership, measurement, or practical operation. It may also be defined by source, as with primary or secondary data, administrative data, clinical experience, patient and carer experience, political necessity, local managerial experience, expert opinion, professional judgement, or public values.

HTA decisions are often made precisely at the point where different kinds of evidence meet. Evidence about relative treatment effect may need to be weighed alongside evidence about severity. Cost-effectiveness may need to be considered alongside financial risk, uncertainty, system capacity and health inequality. Patient experience may challenge the adequacy of an outcome measure. Clinical judgement may question the generalisability of trial data. Public values may bear upon whether a technically efficient rule is acceptable in a public health system. None of this is a departure from evidence-informed decision making. It is the very essence of evidence-informed decision-making when the decision is institutional, public and allocative.

Recent work on values in HTA reinforces this point. Charlton et al. (2024) argue that HTA is an inherently value-based activity that makes use of normative reasoning alongside empirical evidence. They also observe that the language used to articulate its normative aspects is often insufficiently precise. Values are too often treated as late arriving complications, as though they matter only after the evidence has done its work. But values are already present in the selection of outcomes, the choice of comparators, the handling of uncertainty, the definition of benefit, the treatment of opportunity cost, the status of severity, and the weight given to equity. To deliberate about evidence is therefore also, often, to deliberate about the values that determine what counts as relevant evidence in the first place.

Regardless of which of the three broad types of evidence one is considering, further complexities arise. All evidence needs to be interpreted. Its relevance needs to be assessed. Its quality needs to be assessed. Its applicability in the current context, as compared with that in which it was generated or collected, needs to be assessed. Its completeness needs to be assessed. Qualitative evidence needs to be weighed alongside quantitative evidence. Technical controversy over its standing may need to be settled, or at least exposed. The precision of estimates of effectiveness needs to be considered. The robustness of results needs to be tested by sensitivity analysis. Evidence, of whatever kind, needs to be assessed in relation to values for determining priorities, worthwhileness, what ought to be done and by whom.

Nor does any single piece of evidence, whether scientific or colloquial, usually contain enough to enable guidance to be created without further evidence and assessment. A trial may tell us something about efficacy. It may not tell us enough about service delivery. A systematic review may summarise treatment effect. It may not settle which comparator is morally or institutionally appropriate. A model may extrapolate benefits. It may not resolve whether the assumptions built into the extrapolation are acceptable. A patient submission may reveal burdens neglected by trial endpoints. It may not establish a generalisable effect. A clinician's judgement may illuminate practical feasibility. It may also carry professional assumptions that ought to be challenged. The point of deliberation is not to flatten these differences, but to make them visible and assessable.

To be useful, then, a deliberative process must be able to combine and interpret the population of evidence, however defined, for the purpose intended. It must enable those engaged in it to explain why they decided as they did. This requires more than adding up inputs. It requires a process through which relevant evidence can be identified, research commissioned where necessary, evidence brought together and assessed rather than merely aggregated. It requires a process that allows participants to change their minds as evidence is adduced. It requires a means through which colloquial evidence, context-free science and context-specific science can be assessed and combined. It also requires that what people bring to the deliberation may count: their values, experience, expertise, attitudes to risk, and attitudes to not knowing.

This is also where a link with refined utilitarianism becomes useful (Küçükuncular, 2026). If refined utilitarianism in healthcare treats rules, trust, public acceptability and institutional legitimacy as part of the long run welfare effects of decision making, then evidence cannot be confined to clinical effectiveness or cost per QALY alone. Evidence about social acceptance, trust, procedural fairness, implementation, public reasoning and institutional credibility becomes relevant not as decoration, but as part of what determines whether a rule or recommendation can do the work expected of it. A recommendation that is technically efficient but publicly unintelligible, practically unimplementable, or institutionally corrosive may fail even on broadly consequentialist grounds. Deliberation is one of the settings in which those wider consequences can be brought into view.

Maintaining a common understanding of evidence is likely to become increasingly difficult as further stakeholders are added to any guidance-producing process. Conversely, the more homogeneous the group in terms of professional background and level of responsibility, the less tension and disagreement there is likely to be over what constitutes permissible evidence. But homogeneity has costs of its own. It may make decision making smoother at the price of making it narrower. It may reduce conflict by excluding the very forms of knowledge that would have complicated, corrected or improved the judgement. The object is not to maximise disagreement, but neither is it to purchase consensus by leaving out awkward evidence.

This is why deliberation is epistemically valuable. It gives decision-makers an organised way of asking whether the evidence base is complete enough, whether the relevant forms of expertise have been heard, whether disagreement is empirical, methodological or normative, and whether the decision can be explained in terms that are faithful both to the evidence and to the values at stake. The real world presents decision-makers with a cascade of ambiguous evidence, and the cascade itself generates a need for organised management of the volume and kind of information. In health technology evaluation, one typically confronts summarised scientific evidence on efficacy, other scientific evidence on context, artificial evidence from models that extrapolate beyond experimental time periods, and colloquial evidence from public meetings, patient groups, expert witnesses, professional experience and local knowledge. None of this evidence is totally authoritative. Much of it already contains judgement. To interpret, use and compare these different kinds of evidence, one needs further judgement, and that judgement is likely to be better informed if it is disciplined by a deliberative process.

The conclusion of this section may be put simply. HTA needs deliberation because evidence does not combine itself. The three forms of evidence identified above (i.e. colloquial evidence, context-free scientific evidence and context-sensitive scientific evidence) will not combine by themselves to produce health system guidance. Combining and interpreting them requires a process, and the most suitable process may be deliberative. The next question, therefore, is not whether deliberation is always desirable. It is when such a process is likely to be useful, and what features of a decision problem make deliberation more than a costly adornment.

4 WHEN DELIBERATION IS LIKELY TO BE USEFUL

We have argued so far that deliberation is not algorithm, consultation or commenting, and that HTA needs deliberation because evidence does not combine itself. But it does not follow that deliberation is always needed, still less that it is always worth its cost. We therefore need to ask a more discriminating question: when is a formal process of deliberation likely to be useful?

The answer cannot be simply that it is useful when the issue is complex. As was observed above, the mere complexity of the elements of a decision is neither a necessary nor a sufficient ground for supposing that deliberation is called for. Some complex matters may be resolved through well-established rules, settled methods and uncontroversial evidence. Some apparently simple matters may conceal disagreements about value, relevance, fairness or institutional responsibility that require careful deliberation. The question is not whether a decision contains many moving parts.

It is whether those parts require interpretation, contestation and kinds of judgements that cannot be responsibly supplied by an algorithm, a consultation exercise or a written comment process alone.

Two broad classes of circumstance may be distinguished in which deliberation is likely to add value. The first is epistemic: deliberation is useful when the evidence is heterogeneous, incomplete, contested, uncertain or difficult to transfer from one setting to another. The second is institutional and normative: deliberation is useful when the decision involves contested values, contested measures of outcome, questions of equity and fairness, issues of implementability and operational feasibility, professional ownership, public credibility or political trust. The two classes are analytically distinct, but in practice they commonly overlap.

4.1 Epistemic triggers

One set of characteristics likely to make a formal process of deliberation useful arises from consideration of evidence. This is especially so when the evidence comes from more than one expert discipline, such as epidemiology and economics, medicine and surgery, ethics and statistics, or when it is in part quantitative and in part qualitative. It is also likely whenever the decision-making group may not share the same analytical or statistical framework, the same conventions about evidence, or even the same language for describing what is being assessed. These are all potential barriers to communication, barriers that are unlikely to be overcome by simply adding more evidence to the file.

Evidence from more than one profession can be seen as threatening, especially if it is presented in unfamiliar ways or has intimidating technical qualities. Professional competence may become questioned in the light of evidence that was not normally considered, or that appears to reveal professional inadequacy. Feeling threatened is dangerous, not only for the person threatened but for the whole system of making decisions. A well-conducted deliberative process, together with appropriate prior training, can be one in which people's interests and specific skills are identified, protected where justifiable, and drawn on through debate and discussion to the advantage of the whole.

Deliberation is therefore more likely to be useful when there are technical disputes to resolve, whether in the evidence itself, in the methods used to generate it, or in the methods used to summarise it. Disputes over trial quality, indirect comparison, external validity, uncertainty, model structure, discounting, outcome measurement, subgroup analysis, patient-reported outcomes, or systematic review methods may look technical, but they are often also matters of judgement. Rivalries and differences exist within as well as across professions and disciplines.

Incomplete evidence and absent evidence are also triggers for deliberation. Evidence in HTA is rarely complete enough to tell decision-makers what to do without further interpretation. A treatment may have strong evidence of efficacy but poor evidence of implementation. It may have adequate trial data but weak evidence on quality of life. It may have a plausible model but uncertain long-term extrapolation. It may have persuasive patient testimony but limited comparative evidence. It may have strong evidence in one jurisdiction but weak evidence for the setting in which guidance is being made. These problems do not disappear by being named. They require discussion and exploration.

A further epistemic trigger concerns context. Evidence gathered in one context may be adopted for application in another. The difficulty here is not merely that a number must be adjusted or a parameter varied. The difficulty is that the meaning of the evidence may depend upon clinical pathways, workforce capacity, institutional incentives, public attitudes, legal duties, affordability and delivery structures. A technology that is effective in a trial may not be achievable in a service. A model that is plausible for one health system may be misleading in another. A value judgement that is acceptable in one institutional culture may be unacceptable elsewhere. Deliberation is a way of asking how far evidence travels and what must be known before it is allowed to do so.

This is close to the logic of evidence-informed deliberative processes as they have developed in recent HTA literature. Oortwijn et al. (2022) describe deliberation in HTA as the informed and critical

examination of an issue and the weighing of arguments and evidence to guide a subsequent decision, while also stressing that good deliberation must be fit for local contexts. Simangolwa, Mbonigaba and Govender's (2024) scoping review of evidence informed deliberative processes in sexual and reproductive health and rights benefits package design in sub-Saharan Africa illustrates the problem sharply. They found unclear committee terms of reference, poorly defined participation and deliberation processes, exclusion of patients, the public and producers of health technology, weak use of decision criteria, and dissatisfaction with policy recommendations. The lesson is not merely that more evidence was needed. It is that evidence needed to be embedded in a clearer process of selection, scoping, appraisal, stakeholder involvement, deliberation, communication and appeal.

The real world presents decision makers with a cascade of ambiguous evidence, and that cascade itself generates a need for organised evidence management. In HTA one typically confronts summarised scientific evidence on efficacy, often in the form of narrative reviews, systematic reviews or meta-analyses, each of which will itself have involved judgement. There is other scientific evidence about context, artificial evidence from models that extrapolate beyond experimental time periods, and colloquial evidence that comes through public meetings, hearings of witnesses, patient organisations, professional experience and local knowledge. None of this evidence is totally authoritative. Much of it has already been shaped by judgement. In order to interpret, use and compare the different kinds of evidence, one needs further judgement, and that judgement is likely to be better informed if it is exercised in a deliberative process.

4.2 Institutional and normative triggers

A second general ground for expecting deliberation to be useful hangs on the institutional and normative complexity of many issues. These include questions about the measure of outcome, the balancing of benefits and costs, the definition of benefit itself, especially when it goes beyond conventional boundaries of medicine, uncertainty about key values, risks that need to be assessed and weighed, social and personal values not taken into scientific account, issues of equity and fairness, issues of implementability and operational feasibility, the desire for wide professional ownership, the need for public credibility, and the need for political trust.

This part of the argument is important because deliberation helps keep open management issues that an agency's management might prefer to avoid. HTA decisions are allocative public decisions. They do not simply describe what works. They determine what should be funded, for whom, under what conditions, and at whose opportunity cost. Such decisions nearly always offend or hurt some group. A recommendation to fund a technology may displace resources from anonymous others. A recommendation not to fund may disappoint identifiable patients and clinicians. A recommendation to restrict may generate claims of unfairness. A recommendation to wait for further evidence may be felt by some as abandonment. These are not accidents at the margins of HTA. They are part of what HTA is for.

The idea that the process has been one in which there has been consultation, intensive deliberation, and where important issues have been explicitly raised and weighed, helps to create political trust both above and below. It does so by exposing and countering arbitrariness and self-interest, whether personal, professional, institutional, commercial or political. It makes it easier for people to understand why a particular decision has been reached, even when they dislike the decision. It may also reduce the quality and power of residual opposition to clinical guidance, not by silencing it, but by making it harder to sustain the claim that relevant voices were never heard, relevant evidence never considered, or relevant values never weighed.

This is also where professional ownership matters. Guidance that is technically elegant but professionally alienating may fail in practice. If clinicians, managers, patients or commissioners experience a recommendation as something invented elsewhere and thrust upon them, implementation may be slow, resentful or partial. If, by contrast, the design of the deliberative

process itself has been shaped by a wider range of interested parties, they are more likely to be committed to it, to own it. One is less likely to be alienated by a process one helped to design than by an arbitrary process that somebody else invented. Deliberation can therefore generate guidance whose implementation will be speedier, more capable of identifying impediments to implementation, and more able to find solutions to those impediments and to identify knowledge gaps that might be resolved by further enquiry.

Recent work on stakeholder engagement and patient involvement supports this institutional claim. Carter et al. (2025) show that stakeholder engagement in HTA varies significantly across pathways and jurisdictions, and that its timing and form shape what it can contribute. This matters because engagement that occurs too late, or only as comment, may supply information without genuinely affecting judgement. Gousset et al. (2025) propose a three-domain framework for characterising the impact of patient involvement in HTA: impact on the basis of the HTA result or recommendation, impact on the HTA agency, and impact on patient participants (Gousset et al., 2024). This framework recognises that patient involvement may affect not only evidential interpretation but also institutional learning, transparency, expectation setting, participant confidence and process improvement. In other words, involvement matters not only because it adds data, but because it can alter the character and credibility of the deliberative process itself.

The institutional and normative triggers for deliberation are therefore not soft add-ons to the epistemic triggers; they are often inseparable from them. A dispute over an outcome measure may be simultaneously technical, clinical and ethical. A disagreement over cost-effectiveness may also be a disagreement over severity, disability, end of life value, opportunity cost and public acceptability. A question about implementation may involve evidence about workforce capacity, but also professional identity, organisational incentives and political feasibility. A question about equity may require data on distributional effects, but it also requires a judgement about what kinds of inequality matter, how they should be weighed, and whether departure from aggregate health maximisation is justified.

Deliberation need not deliver consensus to be useful. It may expose disagreement about evidence and enable reconciliation or inform research that may settle disputes. It may expose disagreement about values, for example what counts as equitable, what counts as a good outcome measure for the patient groups in question, or what weights should be applied to different groups, and enable reconciliation where possible. But deliberation is not required to deliver consensus. Identifying lasting disagreements may be better than leaving them implicit and unexplored. A disagreement made explicit may be openly managed, explained, appealed or revisited. A disagreement hidden beneath a technical formula or a procedural formality is more likely to return later as distrust.

This also means that deliberation can be useful for public credibility even when it does not satisfy everyone. HTA decisions are often made in morally awkward spaces, where every option has an opportunity cost and every appeal to need, benefit, equality or urgency has a possible counterclaim. The point of deliberation is not to make tragedy disappear. It is to make the reasons for acting in one way rather than another more complete, more explicit, more contestable and more accountable. In this sense deliberation is linked not only to better evidence but to better public justification.

However, deliberation is costly. It may consume time, staff, money and political attention. It may delay decisions. It may be captured by articulate interests, professional hierarchies, industry power, or the performative habits of people who play to the crowd. It may generate the appearance of inclusion without the reality of influence. It may overburden patients or citizens who are invited to contribute but not properly supported. For that reason, the case for deliberation must remain conditional. The question is not whether deliberation is good in itself. The question is whether, in the circumstances, it is likely to improve the decision enough to justify its cost.

We can now state the matter more compactly. Deliberation is likely to be useful when the decision problem is epistemically open, institutionally consequential and normatively contested. It is

especially useful when evidence comes from several disciplines, when methods are disputed, when evidence is incomplete or context-dependent, when values are unsettled, when equity and fairness are at stake, when implementation is uncertain, when professional ownership is needed, when public credibility matters, and when political trust is fragile. These are not rigid criteria. They are conjectures about the circumstances in which deliberative processes are most likely to enhance the quality of HTA decisions. The next task is therefore to say what such enhancement would amount to. If deliberation is said to produce better decisions, what are the tests of that claim? The first part of the paper has therefore offered conjectures about when deliberation is likely to matter. The second turns to the tests by which those conjectures might be examined.

II TESTS

5 BETTER DECISIONS AND HOW WE MIGHT RECOGNISE THEM

We have said that deliberative processes are likely to be useful when evidence is heterogeneous, methods contested, values unsettled, implementation uncertain, and public credibility or political trust at stake. But this remains incomplete unless we say what deliberation is supposed to improve. It is not enough to say that a deliberative process is more inclusive, more elaborate, more consultative, or more democratic in appearance. Those features may be welcome, but they are not the test. The claim being made here is sharper and more demanding. We conjecture that deliberative processes are good for making better decisions. The test of that claim is whether the resultant judgements are, 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 them, whether positively or negatively.

The tests may be used prospectively or retrospectively. Prospectively, they ask whether a deliberative process is likely to improve the decision before it is made. Retrospectively, they ask whether deliberation appears to have improved a decision after the event. Here, we give priority to the prospective use, because our main purpose is to clarify when deliberation is worth having and how it should be designed. Retrospective assessment remains essential, but it requires case evidence beyond the scope of this article.

These four tests are not offered as a closed theory of goodness. They are, rather, a way of making the claim for deliberation more accountable. If deliberation is to be defended in HTA, it must be possible to say not merely that the process was fuller, or that more people spoke, but that the decision was improved in recognisable ways. This is also the spirit of Culyer and Lomas's (2006) earlier question: deliberative processes and evidence informed decision making in health care may be attractive, but do they work, and how might we know? Their answer points toward an empirical and institutional test, namely whether a deliberative process produces guidance that is informed by relevant scientific evidence, interpreted in a relevant context wherever possible, supplemented where necessary by context-sensitive and colloquial evidence, and judged by outcomes that matter to those affected.

The four tests proposed herein ought therefore to be read as criteria of decision quality, not ceremonial virtues of deliberation. A deliberative process that fails to improve evidential comprehension, contextual fit, implementability or acceptability may yet be polite, inclusive and well-intentioned. It may even be politically useful. But it will not have succeeded in our sense. Conversely, a deliberative process need not be perfect to be valuable. Its value lies in whether it improves the decision compared with plausible alternatives, such as unstructured professional judgement, algorithmic thresholding alone, narrow expert appraisal, written consultation, or the status quo.

5.1 More comprehensively evidence-informed?

The first test is whether the decision is more comprehensively evidence-informed. This does not mean that the decision is based on the greatest possible quantity of data. Evidence accumulation is not the same as evidence-informed judgement. A decision may be surrounded by data and still fail to engage the evidence that matters. It may include extensive clinical evidence but neglect implementation. It may include economic modelling but fail to test the empirical assumptions on which the model rests. It may include patient submissions but fail to understand what they reveal about outcomes, burdens or acceptability. It may include contextual evidence but treat it as anecdote rather than as evidence bearing upon the feasibility and value of guidance.

To be more comprehensively evidence-informed, a decision must have identified, interpreted and integrated the forms of evidence bearing on the decision at hand. This includes context-free scientific evidence, such as trials and systematic reviews; context-sensitive scientific evidence, such as evidence about implementation, organisation, finance, system capacity and distributional effects; and colloquial evidence, such as professional experience, patient testimony, public values and local knowledge (Culyer and Lomas, 2006). The test is not whether each kind of evidence is given equal weight. It is whether each has been considered in a way appropriate to its relevance, quality and role in the decision.

Deliberation can improve comprehensiveness because it allows different evidential conventions to confront one another. Clinical evidence may be tested by economic modelling. Economic modelling may be challenged by patient experience. Patient experience may expose a weakness in an outcome measure. Contextual evidence may undermine a proposed route of implementation. Professional judgement may identify practical constraints that a model has not captured. Ethical reasoning may reveal that a technically persuasive option depends upon a contestable view of fairness. This is not an argument for indiscriminate pluralism. It is an argument for disciplined pluralism, in which the range of relevant evidence is made visible and assessed.

An important contributor to evidential comprehensiveness relates to values. HTA is an inherently value-based activity that uses normative reasoning alongside empirical evidence, and that its normative aspects, when not altogether ignored, are often described in ways that are unnuanced, imprecise or inconsistent (Charlton et al. 2024). A comprehensively evidence-informed decision cannot mean a decision in which values are absent. It means a decision in which the evidential and normative components of judgement have been made explicit enough to be examined. Where values determine what counts as a relevant outcome, comparator, uncertainty, severity claim or equity concern, they too must be brought within the field of reasoned assessment.

The purpose of deliberation is not to demote technical evidence by placing it alongside less disciplined forms of assertion. The purpose is to make technical evidence usable in public decision making. A randomised trial, a systematic review or an economic model may be the best available evidence for one part of the decision. But it is rarely the whole decision. To be comprehensively evidence informed is to know what each piece of evidence can bear, what it cannot bear, and how it relates to the other considerations that must enter guidance.

5.2 Better matched to the context of application?

The second test is whether the decision is better matched to the context of its application. Evidence is generated somewhere, under certain conditions, with particular comparators, populations, methods, time-horizons and assumptions. Guidance is applied elsewhere, in a health system with its own institutions, budgets, staff, legal obligations, public expectations, delivery capacities and opportunity costs. The movement from evidence to guidance is therefore not a simple transfer. It is a translation.

A decision is better matched to context when it attends to the conditions under which a recommendation must actually operate. It asks whether the intervention can be delivered, whether the relevant workforce exists, whether infrastructure is adequate, whether patients are likely to

accept the technology, whether clinicians are likely to use it as intended, whether the comparator is appropriate, whether the modelled pathway resembles the real pathway, and whether the distributional effects are consistent with the ethical and legal commitments of the system. Contextual fit is not a secondary matter to be considered after the real science has been completed. It is part of the science of making guidance accurate, achievable and acceptable.

This is one of the central insights of evidence-informed deliberative processes. The distinction between context-free and context-sensitive scientific evidence is made precisely because guidance is not produced by evidence about efficacy alone (Culyer & Lomas, 2006). A technology may work under trial conditions and yet fail under service conditions. It may be cost-effective in a model and yet infeasible in practice. It may be effective on average but unacceptable in a particular institutional or cultural setting. It may be clinically attractive but financially destabilising. It may be efficient yet publicly indefensible if it is perceived as unfair or discriminatory.

The importance of contextual fit is also visible in broader public health policy frameworks. Haby et al. (2025) propose an integrated framework intended to help policy makers use evidence in a structured and transparent way, while attending to effectiveness, fairness, feasibility, acceptability and accountability in public health decision-making. Although HTA is a more formalised process than many public health decisions, the same lesson applies. A recommendation is not good simply because it is supported by evidence somewhere. It is good only if that evidence has been interpreted in relation to the place, system and population in which the recommendation is to be applied.

Deliberation improves contextual fit by creating a setting in which those who know different parts of the context can test the reach of the evidence. Economists may see the opportunity cost. Clinicians may see the pathway. Patients may see burdens that outcome measures obscure. Managers may see bottlenecks. Ethicists may see normative assumptions. Commissioners may see budgetary consequences. Public representatives may see legitimacy problems. The point is not that each participant is always right about their domain. The point is that the deliberative process gives the decision a better chance of discovering what would otherwise be missed.

A decision poorly matched to context may fail even if its evidential base is impressive. It may be ignored, delayed, resisted, misapplied, appealed, litigated, or quietly undermined. By contrast, a decision better matched to context has a better prospect of being understood, implemented and sustained. Contextual fit is therefore not an ornament of deliberation. It is one of the ways in which deliberation can improve the decision itself.

5.3 More efficiently implementable?

The third test is whether the decision is more efficiently implementable. Implementation should not be treated as an administrative afterthought, to be handed over once the intellectually serious part of the process has ended. In HTA, a decision that cannot be implemented, or that can be implemented only slowly, wastefully, inconsistently or in ways that defeat its purpose, is a poorer decision than one whose implementation conditions have been considered from the beginning.

To say that a decision is more efficiently implementable is not merely to say that it is cheaper to carry out. It is to say that the decision has taken account of the practical arrangements through which guidance becomes action. These may include workforce capacity, training needs, referral pathways, procurement, patient identification, monitoring, data systems, administrative burden, financial incentives, regional variation, professional acceptance and the timing of adoption. A recommendation may look efficient at the point of appraisal and become inefficient at the point of implementation if these conditions have not been considered.

Deliberation is useful here because implementation problems are often known before they are formally measured. They may be known by clinicians, managers, patients, commissioners, pharmacists, carers or service users. They may appear as colloquial evidence, local experience or professional caution before they appear in formal studies. A narrow evidential process may dismiss

such knowledge as anecdotal. A deliberative process can ask whether it is a warning. The aim is not to give uncontrolled experience the status of decisive proof. It is to ensure that practical knowledge is examined before a recommendation is made rather than discovered after failure.

The three-domain impact framework for patient involvement in HTA (Gousset et al. 2024) widens the assessment of impact beyond the immediate recommendation to include impact on the HTA agency and on patient participants. That wider view matters for implementation. Patient involvement may affect the basis of a recommendation by improving relevance, clarifying outcomes or identifying burdens. It may affect the HTA agency by improving transparency, internal learning or the handling of future submissions. It may affect patient participants by improving understanding, confidence or trust. These are not all implementation effects in the narrow sense, but they shape the conditions under which recommendations are received and acted upon.

Efficient implementability also has a political and professional dimension. Guidance imposed upon professionals without explanation may be resisted even when technically sound. Guidance that appears to have ignored patients may lack public credibility. Guidance that creates administrative burdens disproportionate to expected benefit may be followed formally and defeated practically. Guidance that is ambiguous may produce variation and appeals. A deliberative process can reduce these risks by identifying impediments to implementation before they become failures of implementation.

The test is therefore this: will deliberation help produce guidance that can be acted upon more readily, consistently and intelligently than would otherwise have been the case? Will it identify obstacles? Will it adjust the recommendation to avoid avoidable failure? Will it clarify responsibility? Will it reveal where further research, monitoring or conditional implementation is needed? If it does, then deliberation improves the decision not only by making it more acceptable but by making it more practically capable of doing what the decision is meant to do.

5.4 More widely acceptable to those affected?

The fourth test is whether the decision is more widely acceptable to those affected by it, whether positively or negatively. This is the most easily misunderstood of the four tests. Acceptability is not popularity. Nor is it a requirement that all affected parties agree with the decision. HTA decisions often cannot satisfy everyone. Indeed, if a decision has no losers, no opportunity cost and no disappointed claimants, it may not have been a difficult HTA decision in the first place. The relevant question is whether the decision is acceptable in the sense of being publicly intelligible, procedurally fair, reasoned, contestable and legitimate.

A decision may be unpopular but acceptable. A decision not to fund a technology may disappoint patients, clinicians and manufacturers, yet still be acceptable if the reasons are clear, the evidence has been fairly considered, the opportunity costs have been acknowledged, the process has allowed relevant voices to be heard, and there is a route of appeal or revision. Conversely, a decision may be popular but unacceptable if it rests on prejudice, manipulated evidence, undisclosed influence or neglect of anonymous opportunity cost victims. Acceptability in the relevant sense is therefore a matter of public reason and institutional legitimacy, not the volume of applause.

Deliberation can improve acceptability because it can expose the reasons behind a decision. It can show that disagreement was not ignored, that contrary evidence was not suppressed, that patient experience was not merely collected and shelved, that professional judgement was challenged where necessary, and that values were not smuggled into the decision under technical language. If values are not named, they cannot be scrutinised. If they cannot be scrutinised, the decision may appear arbitrary even when it is not.

Acceptability also depends on the quality of participation. A process may consult widely and still fail if those consulted cannot see how their contributions mattered. It may invite patient involvement yet leave patients feeling used rather than heard. It may allow comments but give no meaningful

account of how comments affected the reasoning. Gousset et al.'s (2024) framework is again useful here, because it treats the impact of patient involvement as something that should be identified, evaluated and communicated, rather than merely assumed. If participation is to support legitimacy, institutions must be able to say what it changed, or why it did not change the recommendation.

Haby et al.'s (2025) emphasis on transparency and accountability in evidence-informed public health policymaking also supports this point. In public decisions, especially those involving scarcity, the legitimacy of the decision depends partly on whether those affected can see the relationship between evidence, criteria, reasons and outcome. The same applies in HTA. A recommendation that appears to arise from an opaque calculation may be distrusted even if technically defensible. A recommendation that emerges from a transparent deliberative process may be better able to withstand disagreement because those affected can see how the decision was reached.

The refined utilitarian perspective also has relevance. If rules, trust, public acceptability and legitimacy are part of the long run welfare effects of healthcare institutions, then a decision's acceptability is not external to its quality (Küçükuncular, 2026). On this refined utilitarian view, a technically efficient but publicly corrosive decision may still be a poor decision if it damages trust, cooperation, adherence or perceived fairness. Deliberation is one way of reducing that risk, not by pandering to every preference, but by making the decision more publicly defensible.

This fourth test is therefore not a soft concession to public opinion. It is a demanding institutional test. Will the deliberative process make the decision more intelligible to those affected? Will it expose and answer the most serious objections? Will it provide reasons that can be contested? Will it make clear what values are at stake? Will it reduce the suspicion of arbitrariness or self-interest? Will it leave space for appeal, revision or learning? If so, the decision may be more acceptable even to those who remain disappointed by it.

5.5 The four tests as an evaluative framework

Taken together, these four tests provide a practical way of asking whether deliberation has improved, or is likely to improve, a decision. A better HTA decision is one that is more comprehensively evidence-informed, better matched to the context of its application, more efficiently implementable, and more widely acceptable to those affected by it. These tests are mutually reinforcing, but they are not identical. A decision may be evidence rich but context poor. It may be context-sensitive but difficult to implement. It may be implementable but poorly justified. It may be acceptable in the short term but insufficiently evidence informed. The point of the framework is to prevent any one dimension from masquerading as the whole.

Nor should the tests be treated as a mechanical scorecard. That would defeat the argument of the paper. The tests are themselves matters for judgement. In some cases, evidence comprehensiveness may matter most. In others, contextual fit or acceptability may be decisive. In urgent circumstances, implementability may justifiably dominate. In morally contested cases, legitimacy may become central. The framework does not remove deliberation; it structures it.

The four tests also make possible a more serious comparison between deliberative and non-deliberative processes. One could ask whether a decision made by a deliberative committee was better evidence informed than one made by a threshold rule alone. One could ask whether patient involvement improved contextual fit. One could ask whether wider professional participation improved implementation. One could ask whether transparent reason give improved acceptability among disappointed parties. These are empirical questions, though difficult ones. They are also institutional questions, since the effects of deliberation may emerge not only in a single decision but across a body's habits of reasoning, record keeping, appeal, learning and future credibility.

A final caution is necessary. Deliberation may fail all four tests. It may produce decisions that are no better informed, no better matched to context, no more implementable and no more acceptable than the alternatives. It may even make decisions worse, by giving disproportionate influence to the

articulate, the powerful, the persistent or the strategically positioned. That is why deliberation should not be endowed with overblown claims of virtue. Its defence must remain conditional. But the same conditionality is also what gives the argument force. We do not claim that deliberation is always good. We claim that where deliberation improves evidence, context, implementation and acceptability, it improves decision quality in ways that matter for HTA.

The four tests therefore move the argument from advocacy to evaluation; from conjecture to empirical testing. They allow us to ask not only whether a process was deliberative, but whether it was worth having. They also prepare the ground for the next question. If these are the dimensions along which deliberation may improve decision making, what characteristics of a deliberative process are likely to make such improvement more probable?

6 THE NORMATIVE CASE: DELIBERATION, LEGITIMACY AND PUBLIC REASON

The argument so far has been mainly epistemic and institutional. Deliberation has been defended because it may make HTA decisions more comprehensively evidence-informed, better matched to context, more efficiently implementable, and more widely acceptable to those affected by them. But this does not quite exhaust the case. In public healthcare decision-making, deliberation matters not only because it may improve judgement. It may also be ethically constitutive of legitimacy in public healthcare decision-making.

This claim needs to be made carefully. It is not that every decision becomes legitimate merely because it has been preceded by a meeting, a consultation exercise, or an opportunity to comment. Nor is it that every person affected by a decision must agree with it. That would be impossible in HTA, where decisions about funding, restriction, non-adoption, severity, opportunity cost and exception almost always disappoint someone. The point is rather that, where public institutions make coercive or authoritative decisions about the allocation of finite healthcare resources, those decisions require more than technical accuracy. They require reasons that can be made public, contested, revised and defended. Deliberation is one of the means through which such reasons are formed.

This is where HTA differs from private optimisation. A private actor may rank options according to a preferred objective and act accordingly. A public healthcare institution must do something more demanding. It must explain why a particular claim to treatment, funding or priority is accepted while another is refused. It must explain why some forms of evidence counted, why others counted less, why some values were treated as decisive, and why a disappointed party is nevertheless not being treated arbitrarily. That explanatory burden cannot be discharged by a formula alone. Nor can it be discharged by saying that all interested parties were invited to submit comments. A decision becomes publicly defensible when the reasons for it have been exposed to scrutiny in a process capable of affecting the judgement reached.

A connection with the extra-welfarism occurs here. Extra-welfarism was never merely a technical adjustment to welfare economics. It opened a space in health economics for asking what sorts of benefits, capabilities, needs and equity considerations ought to matter in the evaluation of healthcare (Ako-Arrey et al., 2016; Brouwer et al., 2008; Culyer & Wagstaff, 1993; Rawlins & Culyer, 2004). That shift matters because it recognises that public healthcare decisions are not exhausted by preference satisfaction or by unqualified maximisation. They require an account of the evaluative space in which claims are compared. HTA therefore cannot avoid normative reasoning but it can make that necessary reasoning explicit.

NICE's use of cost-effectiveness is not best understood as the replacement of judgement by a threshold, but as the incorporation of cost-effectiveness within a process of appraisal, interpretation and value judgement (Culyer, 2006; Rawlins & Culyer, 2004). In later work on cost-effectiveness thresholds and expanded HTA, Culyer (2016a, 2016b) again pressed the point that HTA is not merely

an algorithmic device but an institutional process in which fairness and legitimacy have to be addressed. The present argument follows the same line. Algorithms may discipline judgement, but they do not legitimate it. Legitimacy requires that the relevant evidence and values be brought into a public order of reasons.

Charlton et al. (2024) argue that HTA is an inherently value-based activity that uses normative reasoning alongside empirical evidence, but that the language used to describe its normative dimensions is often imprecise and inconsistently employed. If that diagnosis is right, then one of the tasks of deliberation is not merely to collect values, but to articulate them. A committee does not simply ask what people want. It asks what reasons can be given for treating severity, unmet need, innovation, disability, age, rarity, end of life claims or equity as relevant, and how these reasons should be weighed against opportunity cost and overall health gain.

Deliberation also occurs at different levels within an agency or health system. At higher levels, which may include parliamentary, ministerial or agency-level deliberation, decisions may be made about the values, criteria and methods that govern future appraisals. At the appraisal level, deliberation concerns the particular technology, population, comparator, evidence base and recommendation. At the implementation level, deliberation concerns how guidance is translated into practice. These levels are connected: higher-level deliberations may bind lower-level decisions, while difficult cases at lower levels may expose the need to revise higher-level rules.

This point also resonates with accountability for reasonableness. Daniels and Sabin (1997) argued that, under conditions of reasonable disagreement about distributive principles, fair priority setting requires a process in which rationales are public, relevant to fair-minded people, revisable through appeal, and regulated to ensure these conditions are met. The point is not that process can substitute for justice. It is that, when substantive agreement is unavailable, legitimate public decisions require a fair process of reason giving in which disagreement is elicited, exposed and made answerable to reasons. HTA is one of the clearest contemporary settings in which this problem arises. People may disagree reasonably about whether the sickest should come first, whether more QALYs should dominate, whether rarity should matter, whether age may ever count, or whether an end-of-life premium is justified. Deliberation is not a magic solvent for these disagreements. It is a way of making them explicit.

The link with refined utilitarianism is also direct but should not be overstated. Refined utilitarianism, as developed in relation to the NHS, holds that rules, public trust, context and legitimacy are not external ornaments to healthcare allocation. They are part of the long run welfare effects of institutional decision making (Küçükuncular, 2026). A rule that maximises measurable health gain while undermining trust, intelligibility, cooperation or public justification may still be a poor rule if those effects are properly included in the evaluative frame. On this account, deliberation is not merely a procedural supplement to optimisation; it is one of the institutional practices through which a public code of allocation becomes sustainable, intelligible and acceptable under conditions of disagreement.

This is why deliberation matters for legitimacy rather than merely for acceptability. Acceptability, understood sociologically, may be too weak. It may depend on ignorance, resignation, deference, public relations or the absence of visible alternatives. Legitimacy is more demanding. It asks whether those subject to a decision have been offered reasons that are relevant, intelligible, contestable and institutionally serious. A decision may remain painful. It may still be opposed. But it is less likely to be experienced as arbitrary if the reasons for it have been made available, if objections have been heard, if the values at stake have been named, and if there is some possibility of appeal, revision or learning.

This also helps explain why deliberation need not end in consensus. Public reason does not require unanimity. It requires that coercive decisions be justified by reasons that can be offered to others as reasons, rather than by private preference, professional authority, institutional convenience or

political embarrassment alone. In HTA, this means that a recommendation should be defensible not only to economists, clinicians or manufacturers, but also to patients whose claims are refused, citizens whose resources are displaced, and professionals who are expected to implement the guidance. The relevant standard is not that everyone must be persuaded. It is that the decision should be capable of being defended in terms that show respect for those who disagree.

There is therefore a normative difference between a decision that happens to be right and a decision that is publicly defensible. A purely algorithmic decision may sometimes reach a sensible conclusion. A narrow expert process may sometimes judge well. But a public institution cannot rely on accidental correctness. It needs processes that expose the path from evidence to judgement. Deliberation contributes to legitimacy when it makes that path visible, criticisable and revisable.

This normative account also guards against an uncritical celebration of participation. Not every invitation to participate is legitimating. Participation that cannot affect the judgement may be decorative. Consultation that is ignored may deepen rather than reduce distrust. Patient involvement that is unsupported may transfer burdens to those least able to bear them. Public engagement that is engineered to secure a predetermined outcome is not deliberative acceptance. It is managed consent. For deliberation to contribute to legitimacy, it must have the possibility of changing reasons, modifying recommendations, revealing uncertainty, or identifying where disagreement remains unresolved.

This conclusion is modest but important. Deliberation matters in HTA because the decisions are not only technical decisions under uncertainty. They are public decisions about the distribution of benefits and burdens in circumstances where reasonable disagreement is unavoidable. Algorithms may help make trade-offs visible. Evidence may help discipline judgement. Consultation may broaden the field of information. But legitimacy requires that the movement from evidence to recommendation be publicly reasoned. Deliberation is valuable because it can make that movement more explicit, more accountable and more publicly defensible. It is to the institutional conditions of such a process that we now turn.

7 WHAT MAKES A DELIBERATIVE PROCESS GOOD?

If deliberative processes are to be defended on the ground that they may improve decisions, we must ask what characteristics of such processes are likely to make that improvement more probable. The question is not whether deliberation is desirable in the abstract. It is what sort of deliberation, under what conditions, with what evidence, participants, rules, support, transparency and opportunities for correction, is likely to produce guidance that is more comprehensively evidence-informed, better matched to context, more efficiently implementable, and more widely acceptable to those affected by it.

We therefore need a framework for assessing the quality of deliberative processes themselves. The following framework is not intended as a rigid checklist, still less as an algorithm for good deliberation. That would contradict much of the argument so far. It is instead an attempt to organise a set of conjectures about the characteristics likely to matter. Four dimensions seem particularly important: evidence quality, process quality, representational quality, and accountability and learning.

This way of putting matters is consistent with recent HTA guidance. The HTAi and ISPOR task force describes deliberative processes as procedures, activities and events supporting informed and critical examination of an issue and the weighing of arguments and evidence to guide decision-making (Oortwijn, Husereau, et al., 2022). Related practical guidance provides a checklist for governance, structure and the management of HTA stages, and groups the relevant questions by phases of a model deliberative process (Oortwijn, Jansen, et al., 2022). That checklist is not a scoring tool, and

nor should the framework we proposed be read as one. The point is to provide a disciplined way of asking whether a process has the features needed to make good judgement more likely.

7.1 Evidence quality

The first set of characteristics concerns research quality. The quality of both context-free and context-sensitive scientific research available at the start of the process matters, as does the quality of evidence that may come in at various stages of the process. High quality helps. This is true of trial evidence, observational evidence, systematic reviews, meta-analyses, economic models, budget impact analyses, implementation studies, qualitative work, patient evidence and evidence about social values. Poor evidence does not become better because it has been deliberated over. Deliberation may reveal its weakness, but it cannot abolish it.

The same applies to colloquial evidence. One of the difficulties in HTA is that much of the practically relevant material may not arrive in the form of a trial, a review or a model. It may come through clinical experience, patient testimony, local knowledge, managerial experience, professional judgement, or political understanding. Some of it may be partial, interested, fragmentary or anecdotal. But if the colloquial evidence all goes in the same direction, and is not obviously self-interested, there may be reason to attach greater weight to it than would be appropriate if it were isolated or plainly strategic. A good deliberative process does not treat colloquial evidence as decisive. It treats it as evidence to be appraised.

The availability and quality of systematic reviews and meta-analyses are also important. They are inevitably of variable quality, and there are reasonably standard ways of identifying higher and lower quality reviews. But even the best review is not free of judgement. Inclusion criteria, search strategies, comparator selection, outcome choice, interpretation of heterogeneity, and the handling of uncertainty all involve judgement. The better these things are done, the more useful the evidence will be. The worse they are done, the more deliberation will be required to identify the limits of what can properly be inferred.

Clear scoping is therefore crucial. The clarity of the questions to be answered by the process is not a trivial matter. It is about defining the technology, the patient group, the comparator, the relevant outcomes, the decision criteria and the institutional purpose of the appraisal. A badly scoped deliberative process may be energetic but directionless. It may collect much evidence while failing to answer the question that matters. Conversely, careful scoping can reduce unnecessary disagreement and allow remaining disagreement to be more precisely located.

The guidance of Oortwijn, Husereau, et al. (2022) is helpful because it treats the determination of need, scope, desired outcomes and context as part of the design of deliberation itself, not as preliminaries too obvious to require attention. Simangolwa, Mbonigaba and Govender's (2024) review illustrates the same point negatively: where evidence-informed deliberative processes for benefits package design in sub-Saharan Africa were weakly specified, they found, among other difficulties, unclear committee roles, poorly defined participation and deliberation processes, weak use of criteria, exclusion of key stakeholders and dissatisfaction with recommendations. The lesson is that evidence quality cannot be separated from the way a process identifies the question, defines the criteria and organises the evidence to be considered.

7.2 Process quality

The second set of characteristics concerns the quality of the process itself. A basic conjecture is deliberately simple but important: the quality of chairperson matters (Brouwer et al., 2008; 2016a). It may sound vague, but it is not empty. A chairperson can enable discussion or close it down. A chairperson can invite challenge or suppress it. A chairperson can manage dominant voices or allow them to occupy the room. A chairperson can distinguish useful disagreement from obstruction. A chairperson can keep the process directed towards judgement without forcing premature closure. These are not merely matters of personality. They are institutional functions.

A good chair or facilitator should help create the conditions under which participants can say what they know, disclose what they doubt, challenge what they find unconvincing, and revise their views without humiliation. This is especially important where participants differ in status, disciplinary confidence, technical fluency, institutional power or personal vulnerability. Without such management, deliberation may reproduce hierarchy rather than discipline it. It may become a forum in which the most confident voices defeat the more hesitant ones, irrespective of the quality of reasons.

Clarity and openness of process also matter. Participants need to know the purpose of the deliberation, the rules of engagement, the criteria to be used, the status of their contribution, the role of evidence, the route from discussion to recommendation, and what will be recorded or communicated. Openness need not mean that everything is public at every moment. There may be opportunities for in camera discussion to encourage free expression, especially where participants might otherwise be frightened to express a view, to look foolish, or to change their minds. But the use of private discussion must be justified by the needs of deliberation, not by institutional convenience.

There is an obvious tension here. The fewer private sessions there are, in one way, the better. Public reason giving matters. Yet occasionally private discussion may be necessary for free expression. Some committees include people who play to the crowd. Remove the crowd and they may behave like reasonable people. Or perhaps they will not. Some are impossible. The quality of the chair matters, but so too does the conduct of members. Members need not abandon their expertise, interests or institutional locations, but they must be willing to deliberate rather than merely advocate. That is, they must be able to treat the evidence and reasons offered by others as capable of changing the judgement.

Reasonable timelines are also central to process quality. Timelines cannot be infinitely long. There is pressure for decisions to be made. Delay may itself be harmful, especially where patients are waiting, budgets are uncertain or technologies are already in use. On the other hand, the shorter the timeline, the less opportunity there is for consultation, discussion, deliberation, reflection, further evidence gathering and changing one’s mind. Time is not a neutral administrative variable. It shapes the quality of judgement.

Agenda setting is also part of process quality. A deliberative body can be weakened by an agenda that asks too much of it. If too many technologies, issues or documents are placed before participants at once, deliberation may become hurried, selective or superficial. Overburdened agendas may encourage skimping, deference to prepared summaries or resignation by members whose contribution is most needed. Good deliberation therefore requires not only enough time, but a manageable sequence of questions.

A good deliberative process therefore needs opportunities for study, thought, discussion and reflection. Participants need enough time to understand the evidence, seek clarification, identify uncertainty and consider the implications of different options. They also need scope to request further information and, where appropriate, to take oral evidence when they have doubts about something. A process that does not allow participants to ask for more evidence may deliberate only on what happened to have been supplied. That is often not enough.

The HTAi and ISPOR task force is again relevant. Its guidance emphasises preparation, conduct, documentation, communication, monitoring and evaluation across the phases of deliberative process design (Oortwijn, Jansen, et al., 2022). It also notes that well-designed and well-executed deliberative processes may provide stakeholders with reasons to perceive an HTA process as fair based on procedural characteristics, even where they disagree with its outcome. This is the practical face of process quality. A process must be good enough not only to reach a judgement, but to make that judgement institutionally defensible.

7.3 Representational quality

The third set of characteristics concerns representation and membership. Deliberative bodies require representative expertise in the relevant scientific evidence, clinical and otherwise. They require people able to assess context-free and context-sensitive evidence. They need clinical expertise, economic expertise, methodological expertise, ethical expertise and, where appropriate, legal, managerial and implementation expertise. They also need representative breadth of colloquial sources of evidence, including practising clinicians, patients, carers, service managers and those with relevant topic specific experience.

The point is not to assemble a miniature parliament of every possible interest. Nor is it to suppose that every stakeholder must be present in every discussion. Membership must fit the question. But if the relevant forms of expertise and experience are absent, the deliberation will be narrower than the decision requires. If patients are absent, burdens and outcomes that matter to them may be misdescribed or ignored. If managers are absent, implementation barriers may be missed. If clinicians are absent, pathways may be misunderstood. If economists are absent, opportunity cost may be neglected. If ethicists or public representatives are absent, value judgements may remain implicit.

The importance of representation is not merely epistemic. It is also institutional. Without appropriate representation, the whole exercise may lack credibility and may undermine trust in experts, even where expert assessment is technically strong. Carter et al.'s (2025) scoping review of stakeholder engagement in HTA pathways is useful here because it shows that stakeholder involvement varies substantially in form, timing and purpose across HTA systems. The implication is not that every stakeholder must be involved in every deliberation, but that the process must be able to explain who was involved, at what stage, in what capacity and with what influence. A written submission is not the same as participation in deliberation. A hearing is not the same as shared responsibility for judgement. A patient story is not the same as structured patient evidence. Representational quality therefore concerns not only who is invited, but what role they play and whether their contribution can actually bear upon the decision.

The three-domain impact framework for patient involvement (Gousset et al. 2024) strengthens this point. They propose evaluating patient involvement in terms of its likely impact on the HTA result or recommendation, its impact on the HTA agency, and its impact on patient participants. That is a useful correction to tokenism. Patient involvement is not good merely because patients were present. It is good if it improves relevance, transparency, expectation setting, institutional learning, participant confidence or the quality of the recommendation. A deliberative process that includes patients but cannot identify what their involvement changed may need to ask whether it involved them well.

Representational quality also requires willingness of members to share values openly. Much disagreement in HTA is not merely about evidence. It concerns what counts as equitable, what counts as clinical significance, what counts as a relevant benefit, what weight should be given to severity, rarity, unmet need, end-of-life claims, disability, innovation or anonymous opportunity cost. If these values remain implicit, deliberation may become a technical discussion concealing a normative dispute. If they are made explicit, they can be examined, challenged and weighed.

Inclusivity of stakeholder consultation also matters, but again consultation must not be confused with deliberation. Opportunities for all affected parties to be heard can widen the evidential and normative base of deliberation. They can help identify omissions, misunderstandings and implementation problems. But representational quality depends on more than hearing. It depends on whether the process has a coherent account of who needs to be represented, why, in what capacity, with what support, and at what stage. It also depends on how partiality, conflicts of interest, professional loyalties and strategic behaviour are managed.

Where feasible, involvement in the design of the deliberative process itself may also support ownership. People are more likely to accept the authority of a process whose rules they understand, whose purpose they recognise, and whose design they have had some opportunity to influence.

7.4 Accountability and learning

The fourth set of characteristics concerns accountability and learning. A deliberative process should include an opportunity for appeal where the process is suspected of being flawed or the guidance that comes out appears unreasonable. Appeal is not an embarrassment to deliberation. It is part of the institutional architecture that makes deliberation credible. A process without appeal may appear final but not fair. A process with appeal acknowledges that judgement can fail, evidence can be missed, criteria can be misapplied, reasons can be unclear, and affected parties may sometimes have grounds for challenge.

Transparency is also essential, though it must be handled intelligently. The process should record what evidence was considered, what criteria were used, what disagreements emerged, what reasons were accepted, what uncertainties remain, and why the recommendation was made. This does not mean that every sentence uttered must be published. It means that the reasons for the decision should be recoverable and scrutinised. Without such records, future decision makers cannot learn from their predecessors. Nor can affected parties understand why the decision was reached. The audience for accountability is not limited to those directly affected by a particular recommendation. It also includes those with a legitimate interest in the design, consistency and fairness of the process itself, including future patients, clinicians, commissioners, researchers and public institutions.

This is where casuistry, in its literal sense, becomes important. Institutions may need to become more casuistical, not by abandoning principles, but by developing case studies over time of what they actually decided and why. A body that records reasons well accumulates institutional memory. Similar cases can then be compared. Apparent inconsistencies can be explained or corrected. Difficult trade-offs can be examined as precedents rather than rediscovered each time as surprises. This would require HTA bodies to record their decisions differently from the way they often do now, but it would greatly strengthen accountability and learning. In the absence of a complete general theory or fully settled evidence about how deliberative processes should work, the best approach may be to accumulate experience systematically, compare it across contexts, and infer general principles from that experience (Culyer, 2020). The point is not merely to defend deliberation, but to make deliberative practice itself an object of inquiry.

Monitoring and evaluation are therefore not optional. Oortwijn et al. (2022) place monitoring and evaluation as one of the six phases of a model deliberative process, asking whether the desired change from implementing a deliberative process has been established, how it should be measured, how it should be assessed, and what indicators should be used. This maps closely onto the argument of this paper. If deliberation is said to improve decisions, institutions must ask whether it has done so. Prospectively, the questions are: will deliberation improve evidential comprehension, contextual fit, implementability or acceptability? Is it likely to help participants revise their views, make reasons clearer, improve trust, identify knowledge gaps or lead to better guidance? Retrospectively, the questions are: did it do these things in the cases where it was used, and if not, why not? These are not simple questions, but they are the right ones.

Simangolwa, Mbonigaba and Govender's (2024) review of the weaknesses in many current systems identified defects that were not merely defects. They are failures of accountability and learning. If roles are unclear, no one knows who is responsible for what. If criteria are weak or implicit, reasons cannot be tested. If stakeholders are excluded, legitimacy suffers. If dissatisfaction is not examined, institutions learn too little from their own failures.

Gousset et al.'s (2024) framework encourages institutions to evaluate the impact of involvement rather than assume it to be positive. It asks what patient involvement changed in the

recommendation, the HTA agency and patient participants. Reporting of this kind helps prevent involvement from becoming symbolic. It also helps build institutional memory. Over time, an HTA agency can learn which forms of involvement improve relevance, transparency, process design and decision quality, and which forms impose burdens without sufficient benefit.

Accountability and learning also require attention to the alternatives. When we say that deliberative processes are more likely to do something, the comparison matters. More likely than what? The status quo? A narrow expert committee? A fixed cost-effectiveness threshold? A written consultation? A cheaper process relying mainly on industry supplied evidence? It would be useful to specify such alternatives and, where possible, test aspects of them. The costs of running decision making organisations can be high, and it is reasonable to ask whether the process is itself cost-effective.

Taken together, evidence quality, process quality, representational quality, and accountability and learning provide a practical framework for judging the quality of deliberative processes. The framework does not guarantee better decisions. No framework can. But it identifies the conditions under which better decisions become more likely. A deliberative process with poor evidence, bad chairing, confused roles, narrow membership, rushed timelines, no opportunity for reflection, no appeal, weak records and no learning mechanism is unlikely to deliver on the promise of deliberation. A process with high quality evidence, clear scoping, good chairing, appropriate expertise, inclusive representation, openness about values, workable timelines, transparent reasons, opportunities for appeal and a commitment to learning stands a better chance.

The conclusion is deliberately modest. Deliberation is not a magic property that attaches to any committee with stakeholders in the room. It is a demanding institutional practice. It has to be designed, chaired, supported, recorded, challenged and evaluated. Its participants must recognise that they are not merely ambassadors of sectional interest, but participants in a judgement directed towards the whole. Its institutions must be humble enough to learn.

8 CONCLUSION AND FUTURE AGENDA

We began with a contrast between the deliberative process and the algorithm. That contrast should not be misunderstood. We are not against algorithms. Indeed, much of modern HTA would be impossible, or at least much worse, without formal modelling, cost-effectiveness analysis, QALYs, systematic reviews, thresholds, modifiers and explicit criteria. These tools discipline intuition. They expose opportunity costs. They make some trade offs visible that would otherwise remain hidden. But their proper role is to be an aid to thought, not a substitute for it.

The argument of this paper has therefore not been that HTA needs less analysis. It needs analysis of the best available kind. The argument has been that analysis becomes guidance only through judgement. HTA decisions are interpretive, value-laden and institutionally situated. They require the interpretation of heterogeneous evidence, the articulation of values, the weighing of uncertainty, the assessment of context, and the justification of trade-offs to those affected by them. In such circumstances, deliberation is not a decorative democratic supplement to a basically technical exercise. It is one of the institutional practices through which evidence and judgement are brought into a form capable of public decision.

This is why consultation is not enough, commenting is not enough, and algorithm is not enough. Consultation may widen the evidential base. Commenting may correct errors and expose objections. Algorithms may structure the comparison of options. But a deliberative process can, unlike these other procedures, be intrinsic to the actual decision-making process. It is the setting in which evidence, reasons and values may be tested against one another for the purpose of reaching a recommendation. This is what makes deliberation critically different.

We have suggested that deliberative processes in HTA should be judged not by their ceremonial fullness, nor by the number of people invited to speak, nor by the warmth of their atmosphere, but by whether they lead to better decisions. The four tests proposed here are that decisions should be more comprehensively evidence-informed, better matched to the context of application, more efficiently implementable, and more widely acceptable to those affected by them, whether positively or negatively.

The answer will not always be favourable. Deliberation may fail. It may be captured by dominant voices. It may give the articulate more influence than the silent. It may become slow, expensive, obscure or performative. It may manufacture the appearance of agreement while leaving the real grounds of disagreement untouched. It may consult without listening, include without empowering, and deliberate without changing anything. These are not small risks. They are among the reasons why deliberative processes should not be endowed with overblown claims of virtue.

The defence of deliberation must therefore remain conditional. Deliberation is likely to be useful when evidence is heterogeneous, incomplete, uncertain or context dependent; when methods are contested; when outcome measures carry hidden values; when equity and fairness are in issue; when implementability and operational feasibility matter; when professional ownership is needed; when public credibility is at stake; and when political trust is fragile. Where these conditions do not arise, or where the cost of deliberation is disproportionate to the decision at hand, a simpler process may be perfectly adequate.

8.1 Six tasks for the future

The future agenda follows directly from this modest defence of deliberation. The first task is comparative. When we say that deliberative processes are more likely to improve decisions, the obvious question is: more likely than what? The relevant comparator might be the status quo, a narrow expert committee, a fixed threshold rule, a written consultation, a less costly process relying mainly on industry supplied evidence, or a more formal multi criteria decision framework. These alternatives should be specified rather than left implicit. Only then can deliberation be assessed against plausible institutional substitutes.

The second task is to ask whether deliberative processes are themselves cost-effective. This is a hard question, but a reasonable one. The costs of running decision making organisations can be high. They include staff time, expert review, patient and public involvement, administrative support, documentation, meetings, appeals, publication and delay. A good deliberative process may well justify these costs if it improves evidential interpretation, contextual fit, implementation and legitimacy. But that claim ought, where possible, to be examined rather than assumed. Haby et al. (2025) frame evidence informed public health policymaking as a structured process concerned not only with evidence, but with fairness, feasibility, transparency, accountability and political context. Their framework points towards the kind of institutional evaluation that deliberative HTA now needs.

The third task is to build a more systematic body of cases. This is casuistry in the best sense of the term: not the abandonment of principle, but the disciplined accumulation of cases, reasons, uncertainties, objections and outcomes. There is a great deal to be said for developing a database of typical scenarios that deal with recurrently conflicted issues, together with the factors that decision makers considered relevant and the reasons they gave. This would not mean replacing judgement with precedent. It would mean making judgement more cumulative. HTA bodies could record not only what they decided, but why they decided it, what was uncertain, what values were in tension, what objections were raised, and how those objections were answered. Over time, such a database could help committees reason by comparison, notice inconsistency, refine criteria and learn from hard cases. This would be especially useful in cases involving severity, rarity, end of life treatment, disability, innovation, uncertainty, equity, unmet need and displacement of resources. It would also require HTA bodies to record their reasons more fully and more consistently than many currently do.

The fourth task is to examine stakeholder and patient involvement with more precision. Carter et al. (2025) show that stakeholder engagement along HTA pathways varies significantly in form, timing and purpose. This matters because engagement before scoping, during evidence appraisal, at committee deliberation, through written comment, or after draft recommendation may have very different effects. Future work should ask not only whether stakeholders were involved, but what role they had, what their involvement changed, and whether the process could explain why a contribution did or did not alter the recommendation.

The fifth task is to make the anonymous opportunity cost victim more visible. HTA often hears from identifiable patients and groups who stand to benefit from a favourable recommendation. That is proper and important. But in a fixed budget system, resources devoted to one group are not available for others. Who are these other people? They are often anonymous. We do not know their names, conditions or stories. But it does not follow that they are less worthy, less needy or less deserving of consideration. There is a case for developing some convention for representing the anonymous opportunity cost victim within deliberative processes.

The sixth task is learning by doing. There is still no complete theory of evidence informed deliberative processes, and the empirical evidence on how best to design them remains limited. There is a need to accumulate experience systematically across countries and institutions, so that general principles may gradually be inferred from practice (Culyer, 2016a, 2020, 2022). Oortwijn et al. (2022) have already provided guidance and a practical checklist for designing and implementing deliberative processes in HTA. The next step is to use, test, revise and compare such guidance in real institutional settings.

None of this changes the central conclusion. Algorithms remain essential, but they remain aids to thought. Evidence remains essential, but facts do not speak for themselves. Consultation and commenting remain valuable, but they are not deliberation. HTA needs deliberation when decisions require the interpretation of heterogeneous evidence, the explicit weighing of values, the management of uncertainty, the assessment of context, and the public justification of trade-offs. A good deliberative process does not make difficult decisions easy. It makes them more reasoned, more accountable, more contestable and more publicly defensible.

The final thought is therefore a practical one. The future of deliberative HTA should not lie in praising deliberation, but in improving it. That means better evidence, better scoping, better chairing, better representation, better records, better appeals, better learning and better tests of whether the process has improved, or is likely to improve, the decision. If deliberation is to deserve its place in HTA, it must earn it by helping institutions make decisions that are more comprehensively evidence informed, better matched to context, more efficiently implementable, and more widely acceptable to those affected. That is a demanding standard, but not an impossible one.

REFERENCES

Ako-Arrey, D. E., Brouwers, M. C., Lavis, J. N., Giacomini, M. K., & AGREE-HS Team. (2016). Health systems guidance appraisal—A critical interpretive synthesis. *Implementation Science: IS*, *11*, 9. https://doi.org/10.1186/s13012-016-0373-y

Brouwer, W. B. F., Culyer, A. J., Van Exel, N. J. A., & Rutten, F. F. H. (2008). Welfarism vs. Extra-welfarism. *Journal of Health Economics*, *27*(2), 325–338. https://doi.org/10.1016/j.jhealeco.2007.07.003

Burke, E. (1958). The Correspondence of Edmund Burke. In *The Correspondence of Edmund Burke. Vol. II (July 1768–June 1774). Edited By Lucy S. Sutherland. (Chicago: University of Chicago Press. 1960. Pp. Xxiii, 567. No price.* (Vol. 2). University of Chicago Press. https://doi.org/10.1017/S0003055400302581 (Original work published 1774, Adam and Charles Black)

Carpini, M. X. D., Cook, F. L., & Jacobs, L. R. (2004). When the People Speak: Deliberative Democracy and Public Consultation. *Annual Review of Political Science*, *7*(1), 315–344. https://doi.org/10.1146/annurev.polisci.7.121003.091630

Carter, D., Laka, M., Gao, Y., Choi, O., Tamblyn, D., & Merlin, T. (2025). Engaging stakeholders along health technology assessment pathways: A scoping review of international practice. *International Journal of Technology Assessment in Health Care*, *41*(1), e69. https://doi.org/10.1017/S0266462325100494

Chalkidou, K., Li, R., Culyer, A. J., Glassman, A., Hofman, K. J., & Teerawattananon, Y. (2016). Health Technology Assessment: Global Advocacy and Local Realities Comment on 'Priority Setting for Universal Health Coverage: We Need Evidence-Informed Deliberative Processes, Not Just More Evidence on Cost-Effectiveness'. *International Journal of Health Policy and Management*, *6*(4), 233–236. https://doi.org/10.15171/ijhpm.2016.118

Chambers, S. (2003). Deliberative Democratic Theory. *Annual Review of Political Science*, *6*(1), 307–326. https://doi.org/10.1146/annurev.polisci.6.121901.085538

Charlton, V., DiStefano, M., Mitchell, P., Morrell, L., Rand, L., Badano, G., Baker, R., Calnan, M., Chalkidou, K., Culyer, A., Howdon, D., Hughes, D., Lomas, J., Max, C., McCabe, C., O'Mahony, J. F., Paulden, M., Pemberton-Whiteley, Z., Rid, A., … Wester, G. (2024). We need to talk about values: A proposed framework for the articulation of normative reasoning in health technology assessment. *Health Economics, Policy and Law*, *19*(2), 153–173. https://doi.org/10.1017/S1744133123000038

Culyer, A. J. (2006). NICE's use of cost effectiveness as an exemplar of a deliberative process. *Health Economics, Policy and Law*, *1*(3), 299–318. https://doi.org/10.1017/S1744133106004026

Culyer, A. J. (2009, June). *Deliberative processes in decisions about health care technologies: Combining different types of evidence, values, algorithms and people* [Office of Health Economics]. Office of Health Economics. Deliberative Processes in Decisions about Health Care Technologies. https://www.ohe.org/publications/deliberative-processes-decisions-about-health-care-technologies/

Culyer, A. J. (2016a). Cost-effectiveness thresholds in health care: A bookshelf guide to their meaning and use. *Health Economics, Policy and Law*, *11*(4), 415–432. https://doi.org/10.1017/S1744133116000049

Culyer, A. J. (2016b). HTA – Algorithm or Process? Comment on 'Expanded HTA: Enhancing Fairness and Legitimacy'. *International Journal of Health Policy and Management*, *5*(8), 501–505. https://doi.org/10.15171/ijhpm.2016.59

Culyer, A. J. (2020). Use of Evidence-Informed Deliberative Processes—Learning by Doing Comment on 'Use of Evidence-informed Deliberative Processes by Health Technology Assessment Agencies Around the Globe'. *International Journal of Health Policy and Management*, *9*(6), 263–265. https://doi.org/10.15171/ijhpm.2019.116

Culyer, A. J. (2022). Reinforcing Science and Policy, With Suggestions for Future Research Comment on 'Evidence-Informed Deliberative Processes for Health Benefit Package Design – Part II: A Practical Guide'. *International Journal of Health Policy and Management*, 1. https://doi.org/10.34172/ijhpm.2022.7398

Culyer, A. J., & Lomas, J. (2006). Deliberative processes and evidence-informed decision making in healthcare: Do they work and how might we know? *Evidence & Policy*, *2*(3), 357–371. https://doi.org/10.1332/174426406778023658

Culyer, A. J., & Wagstaff, A. (1993). Equity and equality in health and health care. *Journal of Health Economics*, *12*(4), 431–457. https://doi.org/10.1016/0167-6296(93)90004-X

Daniels, N., & Sabin, J. (1997). Limits to Health Care: Fair Procedures, Democratic Deliberation, and the Legitimacy Problem for Insurers. *Philosophy & Public Affairs*, *26*(4), 303–350. https://doi.org/10.1111/j.1088-4963.1997.tb00082.x

Fishkin, J. S. (2011). *When the People Speak: Deliberative Democracy and Public Consultation*. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199604432.001.0001

Fox, D. M., & Leichter, H. M. (1991). Rationing Care in Oregon: The New Accountability. *Health Affairs*, *10*(2), 7–27. https://doi.org/10.1377/hlthaff.10.2.7

Gouseti, A., James, F., Fallin, L., & Burden, K. (2025). The ethics of using AI in K-12 education: A systematic literature review. *Technology, Pedagogy and Education*, *34*(2), 161–182. https://doi.org/10.1080/1475939X.2024.2428601

Gousset, V. L., Silveira Silva, A., Holtorf, A.-P., Toledo-Chávarri, A., & Single, A. (2024). The three-domain impact framework for characterizing impact of patient involvement in health technology assessment. *International Journal of Technology Assessment in Health Care*, *40*(1), e52. https://doi.org/10.1017/S0266462324000400

Gutmann, A., & Thompson, D. (2004). *Why Deliberative Democracy?* Princeton University Press. https://doi.org/10.1515/9781400826339

Habermas, J., Rehg, W., & Habermas, J. (1998). *Between facts and norms: Contributions to a discourse theory of law and democracy* (1 MIT Press paperback ed., 4. printing). MIT Press.

Haby, M. M., Reveiz, L., Thomas, R., & Jordan, H. (2025). An integrated framework to guide evidence-informed public health policymaking. *Journal of Public Health Policy*, *46*(1), 193–210. https://doi.org/10.1057/s41271-024-00535-9

Küçükuncular, A. (2026). Refined utilitarianism in practice: Reinterpreting the ethical foundations of the NHS. *Philosophy, Ethics, and Humanities in Medicine*, *21*(1), 3. https://doi.org/10.1186/s13010-025-00206-x

Oortwijn, W., Husereau, D., Abelson, J., Barasa, E., Bayani, D. (Dana), Canuto Santos, V., Culyer, A., Facey, K., Grainger, D., Kieslich, K., Ollendorf, D., Pichon-Riviere, A., Sandman, L., Strammiello, V., & Teerawattananon, Y. (2022). Designing and Implementing Deliberative Processes for Health Technology Assessment: A Good Practices Report of a Joint HTAi/ISPOR Task Force. *Value in Health*, *25*(6), 869–886. https://doi.org/10.1016/j.jval.2022.03.018

Oortwijn, W., Jansen, M., & Baltussen, R. (2022). Evidence-Informed Deliberative Processes for Health Benefit Package Design – Part II: A Practical Guide. *International Journal of Health Policy and Management*, 1. https://doi.org/10.34172/ijhpm.2021.159

Oregon Health Services Commission. (1991). *Prioritization of Health Services: A Report to the Governor and Legislature*. Oregon Health Services Commission. https://www.oregon.gov/oha/hpa/dsi-herc/pages/prioritized-list.aspx

Rawlins, M. D., & Culyer, A. J. (2004). National Institute for Clinical Excellence and its value judgments. *BMJ*, *329*(7459), 224–227. https://doi.org/10.1136/bmj.329.7459.224

Simangolwa, W. M., Mbonigaba, J., & Govender, K. (2024). Health technology assessment for sexual reproductive health and rights benefits package design in sub-Saharan Africa: A scoping review of evidence-informed deliberative processes. *PLOS ONE*, *19*(6), e0306042. https://doi.org/10.1371/journal.pone.0306042

Part of