ABSTRACT
Democratic societies find it difficult to reach consensus concerning principles for healthcare distribution in the face of resource constraints. At the same time the need for legitimacy of allocation decisions has been recognised. Against this background, the National Institute for Health and Clinical Excellence (NICE) aspires to meet the principles of procedural justice, specifically the conditions of accountability for reasonableness as espoused by Daniels and Sabin, that is, publicity, relevance, revisions and appeal, and enforcement. Although NICE has adopted a highly standardised approach and continuously publishes key documents on its website, its technology appraisal programme does not fulfil the publicity condition of accountability for reasonableness. Economic models are not made sufficiently transparent to enable public scrutiny, and decision criteria other than cost-effectiveness remain enigmatic. NICE’s reliance on cost-utility analysis and “plausible” cost-per-quality-adjusted life year (QALY) benchmarks further raises serious issues with regard to the relevance condition of accountability for reasonableness. This is illustrated by counterintuitive cost-per-QALY rankings that are difficult to justify using reflective equilibrium methods, and by the current debate surrounding expensive therapies for rare diseases (“orphan” treatments). In addition, an excessive focus on QALYs may stand in the way of exploiting the best available effectiveness evidence. The NICE mechanism for revision and appeals is also more restrictive than provided in accountability for reasonableness. As to the enforcement condition, no effective quality assurance processes are in place for technology assessments, and implementation of guidance remains imperfect. NICE, despite impressive efforts, appears to have a long way to go before meeting the conditions of accountability for reasonableness.

NICE’s multiple technology appraisal process is broadly considered a role model for health technology assessments that include economic evaluation.1–3 A review team of the World Health Organization (WHO) described key principles of the NICE approach as “the use of best available evidence in decision-making, transparency, consultation, inclusion of all key stakeholders, and responsiveness to change”.4 They concluded that, “in all of these areas, it is clear that NICE is setting a new, international benchmark, for which it can and should be congratulated”.4 Further, NICE has assumed a leading role internationally by fostering methodological advances such as the use of probabilistic sensitivity analyses (designed to capture decision uncertainty) and mixed treatment comparison techniques (in order to enable indirect comparisons of technologies in the absence of head-to-head studies).

THE LOGIC OF COST-EFFECTIVENESS
The logic of cost-effectiveness, as adopted by NICE and in contrast to traditional cost-benefit analysis, does not represent an orthodox application of economic welfare theory.5–9 The development of the cost-effectiveness framework was, instead, heavily influenced by decision analysts with operations research backgrounds, who were striving to transfer methods used to optimise the efficiency of manufacturing processes to the production of health.10

NICE has chosen specifically to be prescriptive about the use of cost-utility analysis—a variant of cost-effectiveness analysis—as its reference case, with QALYs as a universal and comprehensive measure of health-related outcomes.11 Cost-utility analysis, however, is compatible with standard cost-benefit analysis only under restrictive assumptions, including a constant (context-independent) willingness-to-pay for each QALY gained.12–14

Although the use of QALYs is backed by a strong research agenda,15 important methodological issues still remain to be resolved. For example, different valuation techniques give rise to inconsistencies in utility values for similar health states, causing serious reliability problems.16–18 Another well-known but unresolved issue concerns the difference between the utility of a health state expected by healthy persons (or for that purpose, a sample of the general population, as required by NICE for reference case analysis15) and the utility of this health state actually experienced by patients, often confounded by adaptation to disability and disease. This raises further concerns about the content validity of derived QALYs.19–24 Yet a key motivation (beyond the integration in one index of multiple clinical outcomes for defined patients) for the use of QALYs has been the promise to allow meaningful comparisons across a wide range of interventions and, therefore, different patient groups.21–26 Despite the obvious relationship of both subjects, however, examination of the usefulness of QALYs as a measure of health outcomes should not be logically confused with debate about interpersonal comparisons and the appropriateness of specific aggregation rules.27

It is a fundamental and well-established principle of decision analysis that “the identification and structuring of objectives essentially frames the decision being addressed. It sets the stage for all
that follows. To be relevant, analytic decision support relies on prior clarification of the values and objectives to be pursued. Then, to a great deal, applying the logic of cost-effectiveness to inform healthcare resource allocation decisions hinges on the assumption that “the principal objective of the National Health Service (NHS) ought to be to maximise the aggregate improvement in the health status of the whole community.” While it appears trivial that healthcare services (should) produce health, it is by no means self-evident to make a quick leap from here to an assumed “principal objective” of collectively financed healthcare to simply maximise some construct (QALYs or else) of health-related consequences.

In fact, there is little if any evidence that an emphasis on maximisation (sometimes justified by an asserted “consensus in the literature” without specifying sources) is shared by the general population. On the contrary, there is a rapidly growing body of studies that collectively show that this assumption is “empirically flawed.” Controversy revolves around (but is not limited to) a higher social priority for interventions when the severity of the patient’s condition increases, with life-saving interventions most highly valued (this is sometimes referred to as “the rule of rescue”), and for people in so called double jeopardy (ie, with more than one condition causing impairment) who have less QALYs to gain from successful interventions compared to otherwise healthy individuals. To address these issues, there has been a call for more research into “empirical ethics” by leading health economists.

The QALY maximisation assumption is also critiqued from a normative perspective. Arguments prominently include the implied valuation of human life as a function of health status, as opposed to viewing the value of human life as a dimension distinct from health, that is, to assign individual life the same value independent of the presence of disorders and functional impairment. Recently, the premise that “all people are equal regardless of their QALY score” and the presumption of potentially “disastrous effects [of denial of treatment for reasons of cost-effectiveness or, more precisely, the lack hereof] on the sense of personal worth and security” of afflicted patients gave rise to a passionate debate in this journal.

In the absence of a gold standard against which to judge the criterion validity of the logic of cost-effectiveness, it has been proposed to use the so-called reflective equilibrium approach to examine the social acceptability of the resulting rankings of healthcare programmes. Central to a reflective equilibrium approach is the claim that considered moral judgments about justice in particular cases carry weight. The inconsistencies that can arise from the application of standard decision rules derived from the logic of cost-effectiveness are perhaps illustrated best using an example. Assuming the cost per QALY gained (incremental cost-effectiveness ratio, ICER) is, for example, £6000 for sildenafil in erectile dysfunction, £7000 for pharmacotherapy of children with attention deficit hyperactivity disorder, and £120 000 for beta-interferons and glatiramer in multiple sclerosis, would this ranking reflect the comparative social desirability of these interventions?

The issue of counterintuitive rankings is not a phenomenon encountered only in England and Wales but was a major obstacle faced by the protagonists of cost-effectiveness analysis for resource allocation in the Oregon Health Plan (OHP). Under the OHP programme, for example, capping teeth for exposed pulp received a better ranking than an appendectomy for acute appendicitis. While some analysts correctly pointed out that capping a tooth for 150 patients (not one!) was ranked higher than an appendectomy for one person, others insisted that the ranking failed to reflect “the powerful human proclivity to rescue endangered life.” Decomposing reflective equilibrium problems like those cited above reveals several questions, which are not addressed adequately by conventional cost-effectiveness analysis, including: (1) What priority should be given to the worst off—those with the most serious and/or immediate conditions? (2) When should small benefits to a large number of people outweigh large benefits to a small number of persons? (3) How can the conflict between fair individual chances and best aggregated outcomes be resolved?

It is conspicuous that reviews of the usefulness of such rankings (“OALY league tables”) by many health economists, while addressing a variety of technical issues in detail, have not given attention to the larger issue of the validity of the rankings. Of course, the issue of counterintuitive rankings should not be confused with the problem of distorted human judgments due to “heuristics and biases.” Moral intuitions in the sense of reflected values and beliefs—such as Rawls’ non-welfarist account of primary social goods—and Sen’s appeal to a capabilities-based account—cannot be invalidated simply on grounds of their incompatibility with competing normative claims. It has been argued by some philosophers that there may exist an irreducible pluralism at the foundations of normative ethics.

ACCOUNTABILITY FOR REASONABLENESS
Recognising both the difficulty of democratic societies to achieve consensus on distributive principles for healthcare and the need for legitimacy of allocation decisions, Norman Daniels and James Sabin proposed a framework for institutional decision-making, which they call “accountability for reasonableness.” In order to narrow the scope of controversy, accountability for reasonableness relies on “fair deliberative procedures that yield a range of acceptable answers” and consists of four conditions:

**Publicity**, that is, resource allocation decisions must be public, including the grounds for making them. Transparency should open decisions and their rationales for scrutiny by all affected, not just the members of the decision-making group.

**Relevance**, that is, “the grounds for decisions must be ones that fair-minded people can agree are relevant to meeting healthcare needs fairly under reasonable resource constraints.” Arguments should rest on scientific evidence, though not necessarily a specific kind of evidence, and appeal to the notion of “fair equality of opportunity.” Although Daniels and Sabin acknowledge that stakeholder participation may improve deliberation about complicated matters, they believe it is neither a necessary nor a sufficient condition of accountability for reasonableness.

**Revisions and appeal**, that is, there must be an institutional mechanism to engage a broader segment of society in the process, providing those affected by a decision to reopen deliberation, and to offer decision-makers an option to revise funding decisions in light of further arguments.

**Enforcement** entails some form of regulation to make sure that the first three conditions are met.

**NICE’s use of cost-effectiveness as an exemplar of a deliberative process?**
Seeking to combine legitimacy and pragmatism, and realising that utilitarianism “has next to nothing to offer in eradicating health inequalities”, NICE put aside questions of whether matters of content can be resolved solely with a reference to
“due process”\textsuperscript{50} and explicitly subscribed to the principles of accountabilty for reasonableness.\textsuperscript{59} At the same time, NICE reaffirmed its preference for cost-utility analysis with QALYs “as its principal (though not only) measure of health gain”.\textsuperscript{50}

Qualitative research may serve to illuminate the performance of NICE in relation to accountability for reasonableness. A preliminary case study of a recent NICE Technology Appraisal (No. 98;\textsuperscript{62} see http://www.nice.org.uk/) focused on the processes adopted by NICE. The case study, which was largely in agreement with the positive findings of the WHO review,\textsuperscript{4} confirmed the high (albeit not perfect) level of transparency, predictability, and the participatory nature of the NICE approach.\textsuperscript{81} However, the analysis also indicated a need for further in-depth inquiry. A subsequent, more comprehensive review focusing on the technology assessment report informing NICE Technology Appraisal No. 98\textsuperscript{62} did not confirm the expected robustness of the NICE evaluation process, revealing a striking number of limitations and anomalies.\textsuperscript{82} Collectively these left the assessment open to critique regarding all essential components of a technology review question, namely the population studied, the choice of interventions, the clinical and economic criteria used, as well as the study designs and selection criteria.\textsuperscript{83}–\textsuperscript{84} Furthermore, the structure of the economic model itself was found to be prone to distortion and bias in various ways. An unsettling number of consistency problems were identified within the assessment report.\textsuperscript{82}–\textsuperscript{85} As a consequence, the assessment did not consider fully the best available evidence and was unable to identify any differences in clinical effectiveness between the treatment options evaluated.\textsuperscript{82}–\textsuperscript{85}

A number of underlying problems were suggested as causing the observed limitations in the assessment, including, notably, an insufficient integration of clinical and economic perspectives; a high level of standardisation demanding the problem fit a preconceived solution approach, including (but not limited to) the use of QALYs as effectiveness measure; and, somewhat surprisingly, issues related to the technical quality of the assessment itself.\textsuperscript{50}

Significant gaps are observed when the NICE technology appraisal process is compared to the conditions of accountability for reasonableness.

Publicity
The overall process was well structured and followed well-defined timelines with predictable opportunities for (some) stakeholders to provide input; key documents were continuously published at the NICE website. Major limitations of transparency were related to the use of commercial-in-confidence information (a situation on which NICE has taken action meanwhile), the economic model developed by the assessment group, and decision-making criteria beyond cost-effectiveness used by the appraisal committee. For example, the detailed health state vignettes used to elicit utility estimates were not published with Technology Appraisal No. 98.\textsuperscript{82}–\textsuperscript{85} Some of the company submissions may have been biased, as the results submitted tended to favour their respective products. Subsequent in-depth review drawing on peer-reviewed publications (which had not been in the public domain at the time of assessment) identified specific sources of distortion, contributing to these inconsistencies.\textsuperscript{82}–\textsuperscript{85}

Even more importantly, NICE designate economic models as “proprietary”. This insulates a major component of their technology assessments from public scrutiny and does not meet established standards of good economic modelling practice.\textsuperscript{50}–\textsuperscript{56} Read-only copies of models are provided to consultees and commentators only upon their request in writing, with the caveat that these stakeholders “must not publish the model wholly or in part”\textsuperscript{57} and are not permitted to “re-run the model with alternative assumptions or inputs”\textsuperscript{17}. This practice also prevents academic debate and, therefore, is not conducive to the further development of health economic evaluation methods.

As admitted by NICE (cf above), quasi-utilitarian maximisation of QALY gains irrespective of their distribution does not provide for a sufficient basis for healthcare resource allocation in tune with social preferences. Thus, it is a further critical transparency issue that decision criteria other than cost-effectiveness have not (yet) been codified by NICE.\textsuperscript{4}–\textsuperscript{50} Official statements by NICE have remained vague.\textsuperscript{88}–\textsuperscript{89} Appraisal committee meeting minutes are hardly informative,\textsuperscript{81}–\textsuperscript{85} despite reasonable expectations created by NICE’s own process description, claiming, “the minutes provide an accurate record of its proceedings and discussions and also inform the public of the matters discussed at the meeting”.\textsuperscript{57}

Relevance
In the absence of codified criteria for fairness and with its heavy (albeit not exclusive) reliance on cost-effectiveness benchmarks, the NICE approach may be characterised as an “efficiency-first” strategy, with “efficiency” defined according to the logic of cost-effectiveness.\textsuperscript{11}–\textsuperscript{57} The NICE priority for “efficiency” is demonstrated by its expectation that adopting a cost-per-QALY threshold, even if interpreted somewhat flexibly,\textsuperscript{12}–\textsuperscript{59} will “maintain consistency across the many different types of healthcare technologies that NICE appraises”.\textsuperscript{1} It has been argued by observers that this approach in practice will result in the marginalisation of other factors “as outside of NICE’s terms of reference”.\textsuperscript{50} It seems indeed unlikely that the current approach will enable to adequately capture social preferences for healthcare provision.

The current debate surrounding the cost-effectiveness of expensive drugs to treat patients with rare disorders (“orphan drugs”) illustrates this issue. Given the high fixed (ie, volume-independent) and low variable cost structure of the pharmaceutical industry,\textsuperscript{81}–\textsuperscript{85} applying the logic of cost-effectiveness would inevitably deprive these patients of any chance to receive effective treatment.\textsuperscript{86}–\textsuperscript{96} The costs per QALY gained for these treatments often exceed £100 000 at current NHS acquisition prices. It is therefore impossible to justify NHS coverage of these treatments using NICE’s cost-effectiveness benchmark of “a most plausible” (maximum) cost per QALY in the range of £20 000–£30 000.\textsuperscript{88}–\textsuperscript{97}–\textsuperscript{99} However, a majority of the members of the NICE Citizens Council (cf below) believed that “the National Health Service should consider paying premium prices for drugs to treat patients with very rare diseases,” reasoning inter alia that this approach avoids breaching “the human rights of individuals [afflicted with rare disorders]” and “helps shape a more humane society”.\textsuperscript{100} Economists agree that the shadow price of (or willingness-to-pay for) a QALY may well depend on resources,\textsuperscript{100} even if that is associated with a significant loss of efficiency in terms of maximising aggregated outcomes.\textsuperscript{100}–\textsuperscript{105}

The example of orphan drugs sheds light on the end of a continuum, not a distinct well-defined category. From an economic perspective, this example illustrates the role of budgetary impact (as the opportunity costs of programmes...
depend crucially on their size in reimbursement decision-making—a role that NICE has repeatedly denied taking into consideration, despite at least some indications to the contrary. While the position taken by NICE appears questionable on both theoretical and pragmatic grounds, it is evident that explicit recognition of budgetary impact would have fatal implications for any attempt to interpret the logic of cost-effectiveness in a normative way. 

Daniels and Sabin reasoned that fair-minded people “should accept many kinds of evidence and reasons as relevant”, including “scientific evidence about effectiveness and safety”. An overly narrow focus on data which are thought to enable the computation of QALYs may result in the exclusion of relevant information and thus contribute to a situation where technology assessments, adhering to NICE reference case provisions, fail to use the best available clinical evidence. The analysis of Technology Appraisal No. 98 illustrates that this is not merely a theoretical concern. The occurrence of such situations is in stark contrast to claims by leading NICE representatives that its “guidance is based on the best available evidence”.

Revisions and appeal
NICE provisions for appeal are more restrictive than those provided for by accountability for reasonableness. Appeals are narrowly limited to specific grounds and do not permit the debate to reopen. New evidence or simply disagreement with an appraisal will “almost certainly” not be accepted. Although understandable from a pragmatic perspective, these limitations are not adequately compensated for by opportunities for (invited) consultees and commentators to provide inputs during the appraisal process. Only relatively short windows of opportunity are provided, with a massive amount of data to be reviewed under limited transparency.

Enforcement
There is no indication that NICE has implemented an effective quality assurance system for its technology assessments. Again, the case of Technology Appraisal No. 98 suggests that this is not simply a theoretical observation. This issue is exaggerated by the limited transparency of economic models, as discussed earlier. Independent analysts concluded that “absolute transparency of reporting is needed” to address the problem of poor methods in economic evaluations. Conventional peer-review processes are unlikely to be up to the task of assessing the quality of economic evaluation models. Design of effective quality assurance systems must take into account these challenges.

Further, following Hasman and Holm, proper enforcement of appraisal-based decisions should be implemented to ensure that reasoning is “decisive in priority setting and not merely a theoretical exercise”. Although NICE and the NHS have made substantial efforts to improve actual implementation of guidance, significant gaps remain in this area as well. It has been suggested that guidance may be “more likely to be adopted when there is strong professional support, a stable and convincing evidence base” and that “guidance needs to be clear and reflect the clinical context”. These conditions were arguably not fulfilled in the case of Technology Appraisal No. 98.

NICE has established a Citizens Council to provide input on the “topics it wants the council to discuss” and to ensure that its “value judgments resonate broadly with the public”, while maintaining that its guidance “is based on clinical and cost-effectiveness evidence”. The Citizens Council has shown some concern for considerations of social justice, but has broadly endorsed NICE’s approach, concluding that “cost-utility analysis is necessary but should not be the sole basis for decisions on cost-effectiveness”. What is unclear is whether the Citizens Council was confronted with the issue of counterintuitive cost-per-QALY rankings such as those cited above, that is, with the logic that the benefit of providing ten people with a utility gain of 0.1 for the rest of their life (corresponding to sildenafil treatment for men with erectile dysfunction) is indeed considered equivalent to saving the life of a single (otherwise healthy) person. With respect to NICE’s attempts to ensure stakeholder input in general, and to its Citizens Council in particular, it seems worth mentioning that Daniels and Sabin, with explicit reference to advisory bodies and commissions, believe that the absence of a democratic representational procedure “constitutes a decisive objection to claiming that consumer participation contributes to legitimacy”. They further refute the idea “that organisationally based deliberation can substitute for broader democratic processes”. Although selection of Citizens Council members might be described as only “symbolic representation” of the public, NICE should be acknowledged for their efforts to achieve a diverse cross-section of the population.

CONCLUSIONS
In conclusion, there are good reasons to be suitably impressed by the attempts of NICE to ensure rigorous systematic reviews, objective economic evaluation, stakeholder participation, and transparency of process as well as value judgments in their assessment. This notwithstanding, NICE is still in its infancy, and there remains a long way to go until it will have met its stated objective to fulfill the conditions of accountability for reasonableness.

Competing interests: None.

REFERENCES

J Med Ethics 2008;34:534–539. doi:10.1136/jme.2007.021683

537


114. Sheldon CA, Cellum N, Dawson D, et al. What’s the evidence that NICE guidance has been implemented? Results from a national evaluation using time series analysis, audit of patients’ notes, and interviews. BMJ 2004;329:999–1006.


