

# Evidence and Quality, Practicalities and Judgments: Some Experience from NICE

Anthony Culyer and Michael Rawlins

## Abstract

**The National Institute for Health and Clinical Excellence (NICE) is the principal provider of information about the evidence relating to effectiveness and cost-effectiveness in healthcare in the National Health Service of England and Wales. NICE regards quality as primarily to do with effectiveness, safety and the patient experience. In this paper we comment on the quality of evidence regarding these three and speculate about the consequences of widening the range of interventions for appraisal and taking more complete account of upstream determinants of health. We also comment on the type and quality of the evidence, as well as the way in which it is used, and the values – too often hidden – that permeate both the evidence and the way in which it is used.**

Quality, in the context of healthcare, has many dimensions, but Britain's National Health Service (NHS) recognizes three interrelated components

- effectiveness
- safety; and
- the patient experience.

The National Institute for Health and Clinical Excellence (NICE) was set up in 1999 as an independent agency within the National Health Service of England and Wales to provide an authoritative evidential base for “clinical governance,” a system-

atic way of managing and maintaining quality in hospitals and community healthcare providers. NICE's clinical guidelines, technology guidance and all its quality standards are developed by independent committees of experts including clinicians, patients, caregivers and health economists, and now includes guidance on public health interventions. The technologies considered include medicines, medical devices like hearing aids or inhalers, diagnostic techniques, surgical procedures and health promotion. All guidance is considered and approved by the NICE Guidance Executive, a committee made up of NICE executive directors, guidance centre directors and the communications director. A Citizens Council, composed of 30 members of the public, provides the NICE Board with advice that reflects the public's perspective on what are often challenging social and moral issues. NICE International offers overseas jurisdictions advice on the use of evidence and social values in health-care policy. The topics selected for NICE's investigation are determined by the Department of Health (the ministry) after widespread consultation with experts, researchers, NHS service providers and patient representatives. The board of NICE consists of executive and non-executive directors broadly representing the principal stakeholder groups in England and Wales. NICE's scope is likely to be enlarged in the future to embrace interventions in the social care sector. Here, we focus on the contribution NICE makes through its technology appraisals, clinical guidelines and public health guidance to the information available to professionals and patients. In this context, it is

the type and quality of the evidence, as well as the way in which it is used, that matters, and the values – too often hidden – that permeate both the evidence and the way in which it is used.

## Evidence

Incorporating the results of medical research into clinical practice to ensure effectiveness has become entrenched in the notion of evidence-based medicine. As this approach spreads beyond clinical medicine and into the broader domain of health policy, the concept has become subtly transformed from “evidence-based” to “evidence-informed.” Behind this note of realism lie, however, some fundamental questions. What counts as evidence? And, related to that, do some kinds of evidence (or ought they) carry more weight than others?

The *Oxford English Dictionary* defines “evidence” as “facts or testimony in support of a conclusion, statement, or belief.” This begs the question of what counts as a “fact” and gets us nowhere in answer to whether some forms of evidence carry more weight than others. There are many problems with “facts,” of which one, in the present context, is especially problematic. Statements, which everyone may agree to be factual, may be either false or true, or partially one or the other. For example, the statement “antidepressant drugs are used in alleviating the symptoms of depression in dementia” is a factual statement, but recent trials have shown they are no better than placebo (Banerjee et al. 2011). Similarly, the statement “hormone replacement therapy is used to prevent heart attacks” is a factual statement but false in terms of “usefulness” (Rawlins 2011).

### ... what is “scientific” about scientific evidence, and what differentiates it from colloquial evidence?

The kinds of falsity or truth we have in mind are empirical. Agencies like NICE need factual information in order to answer the questions with which they must wrestle in the evaluation of healthcare technologies: “Does it work?” “For whom does it work?” “Relative to what does it work better or worse?” “At what cost does it work?” and “is the expected health gain worth the extra cost?” There are, however, other matters of concern to such agencies. These include, “How confident can we be in the asserted facts?” “How relevant are the known facts to the appraisal of the intervention under investigation and its comparators?” “How complete is the factual information that is available?” and “How – as well as by whom – is the factual evidence contested?”

When non-scientists in the clinical, management or policy worlds are asked what they consider to be evidence, they typically come up with a complex mixture of both scientifically general and locally idiosyncratic types of information – so-called collo-

quial evidence (Culyer and Lomas 2006; Lomas et al. 2005). Clinical or program effectiveness data compete with assertion (sometimes “expert” assertion), cost-effectiveness algorithms sit alongside political acceptability, and data on public or patient attitudes are combined with vivid recollections of personal encounters. The colloquial concept of evidence is broader than the more restricted scientific view and is generally regarded by many scientists as of poor quality. This raises the question of what is “scientific” about scientific evidence, and what differentiates it from colloquial evidence.

The things that are “scientific” about scientific evidence seem to be threefold. First, a formalized hypothesis or theory is being tested. Second, recognized and replicable methods are used to assemble evidence (as, for example, in controlled experiments such as clinical trials). Third, recognized and replicable methods are used to analyze and interpret the evidence (for example, using multivariate regression, propensity scoring or grounded theory). It is not the questions about which evidence is sought that give scientific evidence its distinctive character (Culyer 1981). What makes evidence scientific is the manner in which the questions are answered, not the objects studied or questions asked.

Within this more restricted scientific view of evidence, there are two distinctive manners of study relevant to healthcare decision making. One, relating mostly to testing hypotheses about the efficacy of interventions, uses methods that try to exclude contextual “contaminants,” such as the natural variability in the skills and attitudes of doctors, the symptom presentation of patients, or the organization and funding of service delivery, as well as the more usual “confounders” of epidemiology that blur the line of causality between intervention and outcome. This type of science typically employs, for example, randomized controlled trials (RCTs) to uncover, as far as is epistemologically possible, “context-free” knowledge. The other approach, more common in the social sciences and in the environments in which decisions will be implemented, uses methods that explicitly describe and evaluate the contextual factors that might influence the practical impact of an intervention once it is deployed. This type of science employs a wide variety of methods to make judgments about the likely effectiveness of an intervention “in the field.” This science – for it can be no less scientific in its principles and methods than the experimental approach to evidence gathering – is designed to provide “context-sensitive” results that appraise the facilitating or attenuating circumstances surrounding a particular decision. In context-free science, the emphasis is on what epidemiologists term “internal validity,” meaning the degree of certainty with which the outcome of a trial can be attributed to an intervention rather than to some other variable. In context-sensitive science, the focus is usually on the variables for which the first approach controls, and the emphasis is on “external validity”: the degree

of certainty with which a causal relationship can be generalized to settings other than those of the study. In epidemiology the former is commonly referred to as “efficacy” (the extent to which an intervention produces a beneficial effect under ideal conditions) and is in contrast to “effectiveness” (the extent to which a specific intervention, when used under ordinary circumstances, does what it is intended to do) (Cochrane Collaboration 2012). Context-free evidence is plainly less generalizable and less able to support decision making in contexts that do not approximate that of the original trial. Hence there is a need for supplementary context-sensitive evidence.

### **Hierarchies of Evidence?**

Should the three types of evidence – context-free scientific evidence, context-sensitive scientific evidence and colloquial evidence – be ranked in a quality hierarchy? At one level, the answer might be yes. When they are available, both kinds of scientific evidence must be ranked above the colloquial as far as dependability is concerned. But the science is not always good or complete. Weak evidence sometimes requires use of either inappropriate comparators or indirect comparisons, and estimates of effect have to be derived from observational studies rather than RCTs (Chalkidou et al. 2008). Colloquial evidence comes into its own when scientific evidence is not available or is incomplete in particular and relevant respects (which it frequently is) with regard to context-sensitive matters, on which there is typically much less scientific research than on context-free matters. So colloquial evidence comes into play in a significant fashion when the issue is not whether, say, a medical procedure works in general (as might be demonstrated in US trials), but whether it is likely to work in Canada or Wales, or in community hospitals. If it is believed to work in such places, does it work well enough to warrant public funding? If it seemed to work well over the five-year period of a trial, can it be expected to continue to be beneficial over patients’ expected remaining lifetimes? Or if it were introduced this year, could local services cope with the expected demand? And so on. Evidence that addresses one set of questions is not usefully, or generally, ranked in terms of quality, with evidence addressing another set of questions. Indeed, if colloquial evidence is all there is on one aspect of the performance of an intervention, then the quality of the scientific evidence, relevant though it may be to other aspects of performance, is actually relatively very poor with regard to that aspect.

Contextual facts are matters about which scientific evidence could be collected, but rarely is. If the guidance derived from a deliberative decision-making process is to be as helpful and comprehensive as possible, then colloquial evidence has two essential functions. It provides the relevant context for the context-free science, and it fills in gaps in the knowledge base – gaps that could be filled by scientific evidence but that often

have not been. The issue confronting any decision maker within a deliberative process is thus not so much how to balance the three types of evidence or to assess the weight to place on each, but rather to allow each to perform its appropriate task:

- Scientific context-free evidence is evidence about general potential
- Scientific context-sensitive evidence is evidence about likely realistic scenarios
- Colloquial evidence helps to provide a context for otherwise context-free evidence and to supply the best evidence short of scientific evidence when there is neither context-free nor context-sensitive evidence

This list is not a hierarchy and, as in the evaluation and appraisal of evidence to inform clinical decision making (Rawlins 2011), there is likewise no place for using hierarchies of evidence to inform healthcare policy.

### **Quality beyond Effectiveness**

Decisions are informed not only by evidence about effectiveness or cost-effectiveness, whatever its kind. Values are also all-pervading (Rawlins and Culyer 2004) and range from judgments about the suitability of outcome measures, the weighting of different aspects of a healthy life on the benefit side, to the public and private expenditure consequences on the cost side; from the likely consequences of a decision for distributive justice, and how that is weighed in the balance, to the overall affordability of an intervention compared with the alternatives and the acceptability of the processes through which care is delivered to clients. NICE has sought to resolve issues of these kinds through highly consultative and deliberative decision-making procedures, which include an exercise in “direct democracy” in the form of a Citizens’ Council (Culyer 2005, 2006; Rawlins 2005).

Jurisdictions that are wrestling with issues of quality in healthcare will almost certainly take effectiveness, in the sense of expected impact on people’s health, as the main point of departure. It plainly makes little sense to speak of high-quality healthcare that had a negative impact (iatrogenesis) or a negligible impact (“flat-of-the-curve” medicine). An important role for NICE-type agencies is precisely to address this aspect and, indeed, to generalize it so that no care is excluded from the “insured bundle” that is more effective than care that is included in it (i.e., cost-effectiveness). Moreover, if this aspect of the quality of care is to be treated adequately, the means used by such agencies must themselves be of high quality, which is why NICE strove from the beginning to enlist the active support of the best people in populating its advisory committees and its specially sponsored research groups in universities – and never relying only on the evidence supplied by manufacturers. Quality of this sort comes at a cost – of resources and of time.

But the quality agenda inevitably needs extension beyond effectiveness. One obvious extension relates to the equity of the distribution of healthcare benefits or of health itself. NICE does not have a definitive answer to how this is best done. Indeed, it seems likely that “definitive” answers do not exist and that at least part of the best solution to this element of the quality agenda lies in establishing processes through which concerns about equity can be articulated and embodied – together with their appropriate evidential base – in the advisory processes leading to clinical guidelines and advice on the use of technologies. To this end NICE and the NHS’s National Institute for Health Research have commissioned research that it is hoped will enable an appropriate extension of the usual limitations of cost-effectiveness methodologies (Asaria et al. 2012).

A further extension that also seems inevitable is to apply the evaluative quality principles used by NICE beyond the well-trodden territory of pharmaceuticals into the appraisal of other technologies such as medical devices and diagnostics, beyond these into the evaluation of public health, and eventually into the appraisal of “technologies” relating to the many environmental and “upstream” determinants of health. It is at this point that the limitations of characteristic political structures become sharply clear and why we have only ministries of healthcare rather than ministries of health. It is not merely that we lack the ability to coordinate a comprehensive health policy for quality but that we have only the rudiments of an understanding of the quantitative impact of such health-affecting phenomena and lack even the rudiments of a set of methodologies for evaluating the levers that might be pulled and the ways in which their pulling might integrate with the usual business of healthcare.

Yet another extension is into the patient experience as each patient is in receipt of care. These process aspects of the benefits and harms of healthcare, their measurement and how they might be integrated into more complete appraisals have scarcely been addressed by scholars, let alone implemented by agencies such as NICE.

NICE adopts a diversity of approaches. The scope of its appraisal is constantly widening, as is its evidential base. It certainly does not abandon RCTs in favour of observational studies, nor would it wish to discourage investigators of all kinds from developing and improving their methods. Rather, it seeks to find ways of extending the evidence base, quantitatively and qualitatively, to a wider set of factors that affect health and its distribution. Above all, NICE recognizes that facts, especially facts about “quality,” never “speak for themselves,” needing interpretation, contextualisation and evaluation; that values are all-pervading but may not command universal assent; and that decision-making processes need to be open, consultative and deliberative. Implicit in all these is that what are always required are the exercise of judgment and being able to account honestly for its exercise (Rawlins 2008). **HQ**

## About the Authors

**Anthony J. Culyer** has the Ontario Research Chair in Health Policy and System Design at the University of Toronto and is a professor of economics at the University of York, England. He was formerly the vice-chair of the National Institute for Health and Clinical Excellence in London, England.

**Sir Michael Rawlins** is the founding chair of the National Institute for Health & Clinical Excellence, president of the Royal College of Physicians of London, and professor emeritus at the University of Newcastle. He was formerly Ruth and Lionel Jacobson Professor of Clinical Pharmacology at the University of Newcastle.

## References

- Asaria, M., S. Griffin, R. Cookson, K. Claxton, A.J. Culyer, N. Rice and M. Sculpher. 2012. *Measuring Health Inequality in the Context of Cost-effectiveness Analysis: Don't Concentrate on the Concentration Index!* Paper Presented at the Health Economists' Study Group meeting June 2012, University of Oxford.
- Banerjee, S., J. Hellier, M. Dewey, R. Romeo, C. Ballard, R. Baldwin et al. 2011. “Sertraline or Mirtazapine for Depression in Dementia (HTA-SADD): A Randomised, Multicentre, Double-Blind, Placebo-Controlled Trial. *The Lancet*. doi: 10.1016/S0140-6736(11)60830-1.
- Chalkidou, K, A.J. Culyer, B. Naidoo and P. Littlejohns. 2008. “Cost-effective Public Health Guidance: Asking Questions from the Decision-Maker's Viewpoint.” *Health Economics* 17: 441–8.
- Cochrane Collaborationm. 2012. *Glossary of Terms in the Cochrane Collaboration*. Retrieved October 23, 2012. <[www.cochrane.org/glossary/5](http://www.cochrane.org/glossary/5)>.
- Culyer, A.J. 1981. “Economics, Social Policy and Social Administration: The Interplay between Topics and Disciplines.” *Journal of Social Policy* 10: 311–29.
- Culyer, A.J. 2005. “Involving Stakeholders in Healthcare Decisions – The Experience of the National Institute for Health and Clinical Excellence (NICE) in England and Wales.” *Healthcare Quarterly* 8(3): 56–60. Retrieved 23 October 2012. <<http://www.longwoods.com/content/17155>>.
- Culyer, A.J. 2006. “NICE's Use of Cost-effectiveness as an Exemplar of a Deliberative Process.” *Health Economics, Policy and Law* 1: 299–318.
- Culyer, A.J. and J. Lomas. 2006. “Deliberative Processes and Evidence-Informed Decision-Making in Health Care – Do They Work and How Might We Know?” *Evidence and Policy* 2(3): 357–71.
- Lomas, J., A.J. Culyer, C. McCutcheon, L. McAuley and S. Law. 2005. *Conceptualizing and Combining Evidence for Health System Guidance*. Ottawa: Canadian Health Services Research Foundation.
- Rawlins, M.D. 2005 “Pharmacopolitics and Deliberative Democracy.” *Clinical Medicine* 5(5): 471–5.
- Rawlins, M.D. 2008. “De Testimonio. On the Evidence for Decisions about the Use of Therapeutic Interventions.” *Clinical Medicine* 8(6): 579–88.
- Rawlins, M.D. 2011. *Therapeutics, Evidence and Decision-making*. London: Hodder.
- Rawlins, M.D. and A.J. Culyer. 2004 “National Institute for Clinical Excellence and Its Value Judgements.” *British Medical Journal* 329(7459): 224–7.