International Journal of Technology Assessment in Health Care, 30:2 (2014), 233–238.

doi:10.1017/S0266462314000154

EVIDENCE, VALUES, AND DECISION MAKING

Michael D. Rawlins
Royal Society of Medicine

Background: The evidence supporting the use of new, or established, interventions may be derived from either (or both) experimental or observational study designs. Although a rigorous examination of the evidence base for clinical and cost-effectiveness is essential, it is never sufficient, and those undertaking a health technology assessment (HTA) also have to exercise judgments.
Methods: The basis for this discussion is largely from the author's experience as chairman of the national Institute for Health and Clinical Excellence (NICE).
Results: The judgments necessary for HTA to make are twofold. Scientific judgments relate to the interpretation of the science. Social value judgments are concerned with the ethical principles, preferences, culture, and aspirations of society.
Conclusions: How scientific and social value judgments might be most appropriately captured is a challenge for all HTA agencies. Although competent HTA bodies should be able to exercise scientific judgments they have no legitimacy to impose their own social values. These must ultimately be informed by the general public.

Keywords: clinical effectiveness, cost effectiveness, scientific judgements, social value judgements

It is a gloomy—but inescapable—fact that no country seeking to provide universal healthcare can meet all the health needs of its citizens. None—in either developed or developing countries—has sufficient resources. Nevertheless, countries generally spend what they can afford and there is a close relationship, among both developed and developing (Figure 1) countries, between their gross domestic products (GDP) and their expenditures on healthcare.

Because of the inevitable budgetary constraints on healthcare expenditure decision makers, in all countries, have to decide on their priorities. The role of health technology assessment (HTA) agencies, and others that engage in or sponsor HTA, is to provide decision makers with the evidence upon which these priorities can be informed in an open, and transparent, manner. The evidence for both clinical and cost-effectiveness is therefore a critical component of HTA.

Clinical Effectiveness

The evidence for assessing the clinical effectiveness of new or existing interventions may be derived from either experimental or observational studies (1;2).

Experimental Studies

These are represented by randomized controlled trials in which two (or more) interventions are allocated randomly to groups of patients. The outcomes between the groups are then compared; and if the difference between them reaches "statistical significance" (usually p less than .05) one treatment is considered to be superior.

Randomized controlled trials have three great advantages over observational designs (1). First, the technique minimizes bias—especially selection bias—because every patient has an equal chance of receiving each treatment. Second, it minimizes the malign influence of confounding factors because these are likely to be evenly distributed between the groups. And, third, provided the groups are of an appropriate size, it minimizes the play of chance. Randomized controlled trials therefore have an established place in the assessment of the effectiveness of new and established interventions; and they will continue to have an important place in the future especially if they take advantage of Bayesian methods in their design and analysis (2).

But randomized controlled trials also have limitations (1;2). First, the design and analysis is traditionally based on the "null hypothesis" which assumes, at the outset, that there is no difference between the effectiveness of the two treatments. Only if the probability of the observed difference is less than 1 in 20 (i.e., $p < .05$) is the null hypothesis rejected. The difficulty with this approach is that the null hypothesis is only of interest if the intention is to investigate whether one treatment is superior to another. Although statisticians have developed equivalence, inferiority and futility designs to overcome the restrictions of the null hypothesis, these are clumsy in both their design and are unsuited to the tasks demanded of them (2). Added to which the p-value is too often erroneously believed by many nonstatisticians to distinguish between truth and falsehood.

Second, randomized controlled are undertaken in homogenous patient populations, with few or no co-morbidities, and for relatively short periods of time. Moreover the young, the frail elderly and ethnic minorities are often under-represented. The generalizability (external validity) of the results, to the real world of routine clinical care and particularly patients with co-morbidities, will inevitably be uncertain (2).

Third, randomized controlled trials have become outrageously expensive (2) due, in no small part, to the

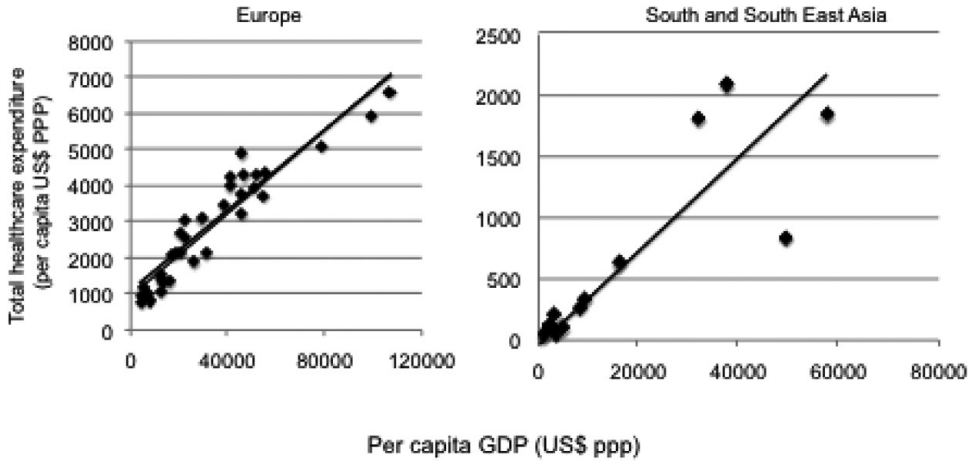

Figure 1. Relationship between individual countries' per capita gross domestic products and health expenditure in US$ at purchasing power parity (PPP) for Europe ($r = 0.92$) and South East Asia ($r = 0.86$).

requirements of drug regulatory authorities. This is part of the reason why new pharmaceutical products have become so expensive. And finally, as discussed later, randomized controlled trials are sometimes unnecessary.

Observational Studies

There are a variety of observational designs that have substantial utility in the assessment of the effectiveness of therapeutic interventions (1). Of these, two—historical controlled trials and case-controlled designs—have been particularly important in evaluating the effectiveness of interventions.

Historical controlled trials compare the effects of an intervention with what is reliably known about the natural history of the condition or from a cohort of patients that have previously been assembled and appropriately observed (2). The use of many therapeutic interventions—of unquestioned effectiveness—is based on historical controlled trials. Some are shown in Table 1. Colchicine, the active ingredient of the plant colchicum, has been used since ancient Egyptian times for the treatment of acute gout; but was only licensed by the U.S. Food and Drug Administration, in 2009, after the results of a placebo controlled trial "confirmed" its efficacy (3). The effectiveness of ganciclovir, in preventing blindness due to cytomegalovirus retinitis in patients with HIV/AIDs, was accepted in the United Kingdom on the basis of evidence from historical controls; but the U.S. Food and Drug Administration required a placebo controlled trial (2). In both instances, patients in the placebo arms of these trials were required to suffer solely in the cause of regulatory science.

In a case-control study exposure to the drug of "interest" is compared among people with and without a suspected iatrogenic disorder (1). An increased exposure in those taking the product suggests a causal association. The technique has been used to great effect in the identification of the harms associated with many pharmaceutical products and a few examples are shown in Table 2. The technique was also used to demonstrate the absence of harms associated with measles immunization (4).

Table 1. Some Interventions Whose Effectiveness Is Based on the Results of Historical Controlled Trials[1,2]

Intervention (year of introduction)	Indication
Colchicine (1830)	Acute gout
Thyroxine (1891)	Myxoedema
Streptomycin (1948)	Tuberculous meningitis
Defibrillation (1948)	Ventricular fibrillation
Ganglion blockers (1959)	Malignant hypertension
Estrogen + progestogen (1960)	Oral contraception
N-acetylcysteine (1979)	Paracetamol poisoning
Ganciclovir (1986)	CMV retinitis
Imiglucerase (1990)	Gaucher's disease
Laser therapy (2000)	Port wine stains
Imatinib (2001)	Chronic myeloid leukaemia

Observational designs too have potential defects (1). The opportunities for bias and confounding are substantially greater than with randomized control trials but these are reduced if the effect size is substantial (5). Observational techniques, however, offer a degree of generalizability that is unattainable with experimental designs (1); and, generally, they are much less expensive to undertake.

Overall, therefore, it should be obvious that both experimental and observational techniques have an important role to play in the assessment of the effectiveness of therapeutic interventions. Yet many—indeed too many—agencies have constructed and used so-called "hierarchies" of evidence in their assessment of therapeutic interventions (1;2).

https://doi.org/10.1017/S0266462314000154 Published online by Cambridge University Press

Table 2. Evidence of Harms Based on Case-Control Studies[1,2]

Intervention	Harms
Combined oral contraceptives	Venous thromboembolism
Diethylstilbestrol in pregnancy	Vaginal carcinoma (in offspring)
Non-steroidal anti-inflammatory drugs	Upper gastrointestinal bleeding
Aspirin in children	Reye's syndrome
Hormone replacement therapy	Venous thromboembolism
Hormone replacement therapy	Breast cancer
Anticonvulsants	Stevens-Johnson syndrome
Olanzapine	Diabetes mellitus
Fluoroquinolone antibiotics	Ruptured Achilles tendon
Biphosphonates	Atypical femoral fracture

Hierarchies of Evidence

The hierarchy of evidence shown in Supplementary Table 1, which can be viewed online at http://dx.doi.org/10.1017/S0266462314000154, is just one of over sixty that have been published over the past 30 years (2). They all place randomized controlled trials at their summits with observational techniques at the foothills. They have been widely used not only by HTA agencies, but also by public health bodies and guideline developers, to assess the quality of the available evidence and inform the strength of any recommendations that they may make.

The use of such hierarchies is not only intellectually (and scientifically) unsound but is also unsafe. Confidence in the effectiveness of penicillin in the treatment of lobar pneumonia (level 2 according to the schema in Supplementary Table 1) is no less than confidence of the effectiveness of statins for the prevention of cardiovascular events (level 1 in Supplementary Table 1). Moreover, these hierarchies are also incapable of handling mixed study designs that are being increasingly used in decision-analytic modeling (2).

The fundamental flaw with the development and use of hierarchies of evidence is that they fail to recognize that it is not the method that matters, but whether the particular method is appropriate to answer the particular question. In many instances questions about the effectiveness of a pharmaceutical product can only be answered by randomized controlled trials (2). But this is not invariably the case; and practitioners of HTA harm themselves, their discipline—not to mention the patients they seek to serve—by slavish adherence to hierarchies of evidence.

Cost-Effectiveness

In the economic evaluation of healthcare interventions, there are four ingredients that have to be taken into account (1): The economic perspective that is to used, The type of economic analysis to be adopted, The threshold distinguishing cost-effective from cost-ineffective interventions, and The approach that is to be taken in resolving the problem of "distributive justice."

Economic Perspective

An economic analysis of the cost-effectiveness of a therapeutic intervention will incorporate the health gains it brings about together with the costs and savings to the healthcare system itself. There are, though, secondary potential costs and savings to families and carers, to the public services (such as sickness benefits or disability payments). Taking a wider perspective, encompassing societal costs and benefits (such as effects on employment or consequences for future healthcare consumption), are described by economists as "productivity" gains and losses (1).

There are, at least superficially, attractions in taking a broader economic perspective because it places the healthcare costs and benefits in a broader economic context. There are, though, disadvantages. For example, taking account of the effects of employment has—in the current economic climate—a net zero sum effect from the perspective of society as a whole. With unemployment rates of around 10 percent to 12 percent across the European Union someone currently unemployed will replace someone who loses their job due to ill health.

The economic perspective to be used in any economic evaluation of healthcare therefore has to be carefully considered. In particular, the inclusion of productivity gains may disenfranchise those members of society who are what the economists describe as "economically inactive" and include children, adolescents and the elderly. Ultimately it is a political, rather than a strictly scientific, decision that has to be made.

Economic Analysis

The favored approach to estimating cost-effectiveness, by many HTA agencies, is the use of cost utility analysis. In this the direct and indirect costs (and savings) are estimated but, in doing so, the scope will inevitably be predicated by the economic perspective that is to be used.

In cost utility analysis, the health-related benefits are expressed from the degree of improvement in the health-related quality of life. The degree of improvement is multiplied by the number of years for which the improvement is maintained to yield the quality-adjusted life-year gained (the "QALY"). The increased net costs are then divided by the QALY gained to yield the "incremental cost-effectiveness ratio (the "ICER").

Cost utility analysis has a major role to play in priority setting. It allows an explicit examination of the consequences of adopting one intervention, for one condition, compared with another intervention for a completely different condition.

Cost-Effectiveness Thresholds

Estimating the incremental cost-effectiveness ratio—the ICER—is one thing, but identifying a threshold distinguishing a cost-effective, from a cost-ineffective, intervention is another. Indeed, there is no consensus among health economists how this might most appropriately be accomplished. Several approaches have been suggested (1).

1. If the ICERs of all the interventions used by a healthcare system are ranked in ascending order the point at which healthcare resources are exhausted might be a basis for defining the threshold. Apart from being a daunting task, such a rigid threshold would reflect a utilitarian approach (see later) that many may find offensive.
2. A threshold could be devised on the basis of the public's "willingness to pay" for incremental healthcare gains. The technique has, for example, been used in transport economics but attempts to undertake this in health economics has yielded very variable results. Furthermore, its applicability for a publicly funded healthcare system is questionable.
3. Another approach, and one favored by some health economists, especially in the United Kingdom, is to examine current patterns of expenditure (so-called "program budgets") and impute the point at which benefits are foregone if a new intervention is adopted. This must necessarily assume that all current healthcare expenditure is cost-effective.
4. The World Health Organization (WHO) has suggested that the threshold should be based on a country's GDP (6). Indeed, WHO advocates that an ICER below a country's per capita GDP should invariably be considered to be cost-effective; and one that is three times greater that per capita GDP should always be considered to be cost ineffective. Although there is logic to relating the threshold to a country's GDP, I am unaware of an empirical basis for the quantitative suggestions of WHO.
5. Finally, there is the approach adopted by the UK's National Institute for Health and Care Excellence (NICE) which is based, essentially, on the collective judgment of the UK's health economic community.

Perhaps fortuitously, the methods described in 3) and 4) above give results that are not too dissimilar to the ICER threshold range used by NICE.

Distributive Justice

Distributive justice is the term used by political and moral philosophers in discussing what is right or just in the allocation of goods within society (7). Three theories of distributive justice—libertarianism, utilitarianism, and egalitarianism—each have a particular resonance to the allocation of resources in healthcare (7).

Libertarianism is based on the premise that individuals should be able, and expected, to finance the healthcare needs of themselves and their families; and that market forces should enable them to do so at a reasonable price. Even in countries where this approach has been most widely adopted, such as the United States, arrangements are in place to protect the less affluent members of society.

Utilitarianism requires that healthcare expenditure is used to maximize the health of the population as a whole. While emphasizing the importance of efficiency in healthcare it does little for either minorities with rare diseases or in reducing health inequalities associated with race, gender, or socioeconomic status.

Egalitarianism tries to distribute healthcare resources so as to allow each individual a fair share of the available opportunities. It seeks to ensure an adequate, but not necessarily a maximum, level of healthcare but raises difficult questions about what is fair—and what is bad luck (8).

The tensions between utilitarianism and egalitarianism can be overstated (8). Many utilitarians accept that social values should be incorporated into their approach to distributive justice; and qualified egalitarians accept the concept of opportunity costs. There is, though, no formal synthesis of these two approaches to distributive justice. At NICE, an attempt has been made to capture the social values of the general public through its Citizens Council.

Capturing Social Values

Social values take account of the ethical principles, preferences, culture, and aspirations that should underpin the nature and extent of the care provided by a healthcare system (7). These include the answer to questions such as whether special priority should be given to children and young people? Whether a healthcare system should be prepared to pay premium prices for drugs to treat very rare serious diseases. Who, though, should decide what social values should HTA agencies and decision makers adopt in prioritizing healthcare?

It might be argued that it should be the role of either Parliament or the government of the day to determine the social values adopted in the provision of healthcare (7). Parliament after all makes laws, raises taxes, and decides on how tax revenues should be spent. Whether it has any special legitimacy to make social value judgments for a healthcare system, even if (as in the United Kingdom) it is funded from general taxation, is unclear. Experience in the United Kingdom, at least, suggests that politicians find it extraordinarily difficult to make such decisions in the face of electoral pressures.

Public meetings are a time-honored way in the British National Health Service to sound out public opinion (7). They provide little opportunity, however, for reflection or deliberation and attendees are usually dominated by those with a vested interest in the particular issue under discussion (7).

Opinion polls and surveys, when conducted competently, can elicit the public's immediate preferences on particular issues. Responses, however, may be colored by inaccurate media activity, and there is no opportunity for discussion or considered thought (7). Replies, moreover, are exquisitely sensitive to the precise manner in which the question is phrased or framed. Polls and surveys do not provide the public's considered conclusions, based on deliberation, about the complexities surrounding priority setting in healthcare (7).

A better understanding of the reasons underlying the public's immediate preferences might be elicited using focus groups which can provide some insight into why the public believes as it does (7). Focus groups, however, are an extension of polling, and the time available (usually not more than 2 to 4 hours) does not allow much opportunity for discussion and deliberation.

NICE's Citizens Council

In an attempt to capture the social values of the British public NICE, in 2002, established a Citizens Council (9). The concept of the Council is based on citizens juries (7) and comprises thirty members drawn from the British public. Members were

https://doi.org/10.1017/S0266462314000154 Published online by Cambridge University Press

Table 3. Some Topics Considered by NICE's Citizens Council

Year	Topic
2002	Clinical need
2003	Age and cost-effectiveness
2004	Ultra-orphan drugs and cost-effectiveness
2005	Mandatory public health measures
2006	Use of the rule of rescue
2007	Patient safety and cost-effectiveness
2008	Departing from the ICER threshold
2009	Innovation
2010	Health improvement and financial incentives
2011	Discounting costs and benefits
2012	Social care values

(and still are) recruited by advertisements in the media. Applications are encouraged from anyone, provided that they are not involved, professionally, with healthcare or the healthcare industries. Those appointed reflect the demographic structure of England and Wales, with respect to gender, age, ethnic background, socio-economic status, and disability. Their ages have ranged from 18 to 76 years, and members have included a London cab driver, a scaffolder, a single parent, and a retired airline pilot.

The Council meets for two and a half days, once or twice a year, and considers a particular question that is of concern to the Institute. Some examples of the topics considered are shown on Table 3. Meetings start with an explanation of the question and why the answer is important. Members then have an opportunity to listen, and cross-examine, external experts who provide evidence and views supporting both sides of the question that has posed. For example, in 2011, the Council was asked *"Under what circumstances are incentives to promote individual behavior change an acceptable way of promoting the health of the public?"* The reason for asking the question was that there is evidence, including from randomized controlled trials, that offering material or financial rewards to individuals to lead healthier life styles (such as abstaining from smoking or elicit drug use) resulted in significant positive results. NICE, though, was unsure whether the use of public funds would meet with general approval. The Council, in its report (10), accepted that that there was strong scientific evidence to support the use of such incentives, but that members were nevertheless uncomfortable with using public funds to "bribe" people in this manner.

Decision Making

When making decisions about the findings arising from HTA, judgments have to be made (11). These fall into two groups: scientific judgments; and social value judgments.

The evidence underpinning any HTA is always incomplete (11) and scientific judgments are needed if patients are to be provided with appropriate care. These judgments include whether the totality of the evidence for an intervention's effectiveness is sufficiently compelling. They often also encompass issues such as whether, despite the homogeneity of the patients included in the clinical studies, the evidence is generalizable to routine clinical practice or whether changes in the quality of life be adequately captured especially in children where quality of life is notoriously difficult to assess. Or whether, on grounds of clinical or cost-effectiveness, it is appropriate that use be restricted to one or more subgroups of patients. These are scientific judgments that HTA agencies, or their advisory committees, should be able to make.

Social judgments relate to the social, rather than the clinical, sciences (11). Should an additional year of a child's life be valued differently from that of its parents or grandparents? Are the last few months of life, for people with fatal conditions, so precious that additional healthcare resources should be available to them by increasing the ICER threshold? These judgments are not ones that HTA agencies, or their advisory committees, have any legitimacy to make.

NICE, therefore, has prepared a guideline (12) on social value judgments for use by its staff and its guidance-producing advisory bodies. This guideline brings together the reports of its NICE's Citizens Council together with relevant legislation especially the Equality Act 2010. NICE's program directors are required to confirm that all NICE guidance conforms to the tenets of the guideline or, if there are any departures from it, that the reasons are compelling and explicit. It is important, though, to appreciate that NICE's social values are specific to the United Kingdom and should not be extrapolated more widely.

In the light of the social values developed by the Citizens Council, and enshrined in NICE's social values guideline, there are occasions when the Institute's advisory bodies have exceeded the ICER threshold of £20,000 to £30,000 per QALY gained. Examples (13) include the use of riluzole (with an ICER of £38,000 to £42,000 per QALY gained) to delay the need for tracheostomy in patients with motor neuron disease (amyotrphic lateral sclerosis); or the use of insulin pumps, in children, where the increment in the quality of life was deemed to be unreliable, but nevertheless substantial, and the overall budgetary impact likely to be modest.

CONCLUSIONS

Judgments—both scientific and social—play a significant role in making decisions about healthcare priorities. Such judgments should of course be informed by the evidence but reliance on hierarchies should play no part. As the nineteenth century English poet, artist, and philosopher—William Blake—put it (14): *"God forbid that truth be confined to mathematical demonstration".*

https://doi.org/10.1017/S0266462314000154 Published online by Cambridge University Press

SUPPLEMENTARY MATERIAL

Supplementary Table 1: http://dx.doi.org/10.1017/S0266462314000154

CONTACT INFORMATION

Michael D. Rawlins, MD, FMedSci (michael.rawlins@nice.org.uk), President, Royal Society of Medicine, 1 Wimpole Street, London W1G 0AE, England

CONFLICTS OF INTEREST

No conflicts of interest.

REFERENCES

1. Rawlins MD. *Therapeutics, evidence and decision-making*. London: Hodder; 2011.
2. Rawlins MD. De Testimonio: On the evidence for decisions about the use of therapeutic interventions. *Lancet*. 2008; 372:2152-2161.
3. Kesselheim AS, Solomon DH. Incentives for drug development — The curious case of colchicine. *N Engl J Med*. 2010;362:2045-2047.
4. Smeeth L, Cook C, Fombonne E, et al. MMR vaccination and pervasive developmental disorders: A case-control study. *Lancet*. 2004;364:963-969.
5. Glasziou P, Chalmers I, Rawlins MD, McCullough P. When are randomized trials unnecessary? Picking signal from noise. *BMJ*. 2007;334:349-351.
6. WHO Commission on Macroeconomics and Health. *Macroeconomics and health: Investing in health for economic development*. Geneva: World Health Organization; 2001.
7. Rawlins MD. Pharmacopolitics and deliberative democracy. *Clin Med*. 2005;5:471-475.
8. Beauchamp TL, Childress JF. *Principles of biomedical ethics.* Oxford: Oxford University Press; 2001.
9. Littlejohns P, Rawlins MD, eds. *Patients, the public and priorities in healthcare*. Oxford: Radcliffe Press; 2009.
10. National Institute for Heath and Care Excellence. *Citizens Council Report. The use of incentives to improve health*. London: National Institute for Health and Care Excellence; 2011. www.nice.org.uk/aboutnice/howwework/citizenscouncil/reports.jsp?domedia=1&mid=10860C6D-19B9-E0B5-D46FA5607685FB05 (accessed April 16, 2014).
11. Rawlins MD, Culyer AJ. National Institute for Clinical Excellence and its value judgments. *BMJ*. 2004;329:224-227.
12. National Institute for Heath and Care Excellence. *Social value judgements: Principles for the development of NICE guidance*. London: National Institute for Health and Care Excellence; 2008. http://www.nice.org.uk/media/C18/30/SVJ2PUBLICATION2008.pdf (accessed April 16, 2014).
13. Rawlins M, Barnett D, Stevens A. Pharmacoeconomics: NICE's approach to decision-making. *Br J Clin Pharmacol*. 2010;70:346-349.
14. Gilchrist A. *The life of William Blake*. London: Bodley Head; 1907.
15. Harbour R, Miller JA. A new system for grading recommendations in evidence based guidelines. *BMJ*. 2001;323:334-336.