

# How Good Is Good Enough? Standards in Policy Decisions to Cover New Health Technologies

Comment savoir si c'est suffisamment bon?  
Normes relatives aux décisions stratégiques qui  
portent sur les nouvelles technologies de la santé



by MITA GIACOMINI, PHD

*Professor, Department of Clinical Epidemiology and Biostatistics*

*Centre for Health Economics and Policy Analysis*

*McMaster University*

*Hamilton, ON*

## Abstract

Health technology coverage decisions require reasonable criteria, for example, the requirement that a technology be effective, efficient, legitimate in purpose, acceptable in its effects, safe and so on. The leap from such criteria to decisions requires not only evidence, but also *standards*. Decision-makers must specify their values, which apply in general, regarding what is “good enough” before they can judge any technology in particular. This paper will do the following: (1) describe the key analytic tasks involved in defining coverage criteria and their standards, (2) identify some of the policy applica-

tions of *explicit* standards to coverage decisions and (3) review the policy uses of such standards, including some challenges they pose. The problem of identifying cost-effectiveness standards will be used to illustrate key issues. It is argued that a precedent-based understanding of standards is relevant in the Canadian policy context, where fairness is crucial. Studies of actual decision-making that seek standards inductively have been misguided in their focus on central tendencies to the neglect of outliers (precedents), while deductive analyses and rules of thumb have been ungrounded in prevailing values.

## Résumé

Les décisions relatives à la protection des technologies de la santé exigent des critères raisonnables, par exemple, qu'une technologie soit efficace, efficiente, légitime dans ses fins, acceptable dans ses effets, sécuritaire, et ainsi de suite. De franchir le pas entre ces critères et la prise de décision requiert non seulement des preuves, mais aussi des normes. Les décideurs doivent préciser leurs valeurs – qui s'appliquent de façon générale – sur ce qui est « suffisamment bon », avant de pouvoir évaluer une technologie en particulier. Dans cet article : (1) on décrit les principales tâches d'analyse nécessaires afin de définir les critères quantitatifs et de protection et leurs *normes*, (2) on identifie certaines applications stratégiques des normes *explicites* pour les décisions relatives à la protection, et (3) on examine l'utilisation stratégique de telles normes, de même que certains des défis qu'elles posent. Le problème de l'identification de normes économiques sera utilisé pour illustrer des enjeux majeurs. On avance qu'une compréhension des normes fondée sur les précédents est pertinente dans le contexte des politiques canadiennes, où l'équité est essentielle. Des études de prises de décision réelles qui cherchent des normes de façon inductive ont fait fausse route en insistant sur les tendances centrales et en négligeant les aberrations (précédents), alors que les analyses déductives et les règles empiriques n'étaient pas fondées dans les valeurs prédominantes.

---

**F**AIR PUBLIC POLICY DECISIONS REQUIRE REASONABLE PROCESSES AND CRITERIA. Many bodies charged with making decisions on health technology coverage now strive for more systematic, evidence-based and transparent bases for their recommendations. Common criteria for judging new technologies include, for example, effectiveness, safety and efficiency. A fuller set of criteria normally includes both quantitative considerations of how well a technology performs and categorical considerations regarding the appropriateness of its purposes and effects. To formulate an evaluative judgment, decision-makers must collect and interpret evidence regarding

each criterion. The leap from evidence to decision requires standards. That is, beyond the knowledge of how “good” a given technology is, evaluators require pre-formed ideas about how good would be “good enough” and what kinds of technologies would be the “good” ones. This paper outlines key analytic tasks involved in applying criteria and evidence to coverage decisions in any context where a systematic, evidence-based approach is pursued. Particular attention is given to the challenge of defining standards – the underappreciated values that link evidence to decisions.

## Criteria, Evidence and Standards Are Different Things

It is important to distinguish among criteria, evidence and standards in evidence-based decision-making. A *criterion* is a general principle (e.g., effectiveness) by which we value any health technology. *Evidence* is evaluative information that tells us how good or fitting a particular technology is, in relation to a given criterion (e.g., research evidence of effectiveness). *Standards* are values that indicate how good would be good enough to qualify for coverage (e.g., how effective is effective enough). The nature, development and application of standards has received comparatively little policy analytic attention.

*Quantitative* evaluation criteria are measured and expressed in numerical terms.

The most familiar of these are effectiveness and efficiency; others include safety, efficacy, budget impact, likely demand and disease burden. Because quantitative evidence is expressed as a matter of degree, quantitative standards take the form of *thresholds* that distinguish adequate technologies from inadequate ones – for example, a relative risk of  $<0.5$  or  $>2.0$  as a compelling effect size for any intervention (GRADE Working Group 2004). Applying such standards to decisions is straightforward: if the technology’s performance is above a threshold level, it passes that criterion and may qualify for coverage.

*Categorical* criteria are those that require more descriptive information. An example is the purpose of a technology: is it preventive or curative, for lifestyle or life-saving? Does it provide information or intervention? Does it target special needs of the poor, elderly or children? Some categories (e.g., whether a physician or hospital service, whether a drug or device) are pragmatically driven by the institutional organization and funding of healthcare (Giacomini 1999). Many other types of categorical criteria may apply, for example, whether the technology affects others besides the patient, or whether it requires adjunct technology. Such distinctions can matter for ethical, political and social reasons, and often help answer fundamental policy questions such as the “medical necessity” of a service for coverage under Canadian medicare. Categorical standards call for *categorical priorities*, not thresholds. To construct these, technology types are sorted into higher- and lower-priority commitments, or acceptable and unacceptable types. Decision-makers classify a given technology using inductive judgments of how well it fits into a qualifying priority category.

## Standards Are Always Used, Whether They Are Apparent or Not

Both quantitative thresholds and categorical standards share key features. First, *standards apply in general*, across all technologies that are candidates for coverage within the relevant policy mandate. Whether a standard is actually followed in decision-making, and the extent to which a given standard is used to justify a given decision, are separate issues. Second, each evaluative criterion entails its own standard. If six criteria are applied, there will be at least six distinct standards that pertain to a decision about a given technology. A standard for one criterion could be conditional on standards for other criteria. Finally, all coverage decision-making involves the use of standards – whether implicit or explicit, consistent or capricious. Explicit, consistent and transparent standards are an important feature of accountability. However, decision-makers may be reluctant to articulate and apply standards transparently when prevailing standards are tacit or do not rest on a clear understanding of consensual values.

Coverage standards remain implicit and intuitive in most Canadian health technology assessment and coverage decision-making. Some advisory committees explicate their criteria for their decision-making, and tremendous strides have been made in the use of evidence. However, few committees can yet articulate their standards. Fugitive standards operate nevertheless, as decisions are made – we can presume that “good enough” judgments underlie coverage recommendations, and they are not completely arbitrary. Unfortunately, these tacit standards may fluctuate with the vagaries of institutional memory, membership and politics of advisory committees. The next stage in the development of rational, evidence-based coverage decisions should involve the critique and improvement of our fugitive standards.

## Explicit Standards Support Fairness

Explicit standards offer several advantages. The first is consistency and fairness. Standards serve the equity imperative to “treat like technologies alike.” To the extent that we judge health technologies equally, we also give their human stakeholders and beneficiaries fairer treatment. Standards resonate with the rule of precedent in common law. Decisions that exceed established standards set new precedents and imply new standards for future decisions. In practice, decision-makers often forge standards not from abstract principles, but from analogical comparisons to past coverage decisions that serve as implicit precedents for acceptability (Giacomini 2005). Transparent criteria and standards give concrete meaning to the values governing the health system, and make it easier to hold decision-makers accountable to them. When decisions based on prevailing standards seem nonsensical, the standards – and underlying values – can be re-examined. Explicit attention to standards also expedites decision-making because policy makers need not deliberate “what’s good enough” each time they face a specific

case. This is especially important for committees of diverse and fluctuating membership, where repetitive conflict among individuals' tacit standards can cost time and focus.

Explicit standards also shift moral burden from the shoulders of advisory committees who routinely make discrete coverage recommendations to those who would periodically set the standards, in general. Ideally, standards should be set outside the pressing context of decision-making, and by a legitimate body constituted for the purpose of values clarification and interpretation (Giacomini 2005). Even so, the coverage decision-making process must provide some feedback and input to the standard-setting

---

**When it becomes clear “how good is good enough,” innovators can make technologies “good enough” – or more perversely, seem to be.**

process, especially as new technologies challenge pre-existing ideas about what is acceptable or valuable. In case-by-case decisions, the task of applying explicit criteria and standards requires decision-makers to face and reconcile diverse criteria into a summative judgment. If a decision seems to violate

one standard (e.g., a cost-effectiveness threshold), this calls for explanation in terms of another criterion and its standard (e.g., a worthy medical purpose or a needy target population). Arguments from analogy to other technologies and precedents help to highlight true evaluation criteria, and to move deliberations from less relevant criteria to more relevant ones (Giacomini 2005). As a classic example, some suggest that Viagra® is far more cost-effective than renal dialysis (J. Smith, Health Management Research Centre, University of Birmingham, personal communication 2003) – yet insurers balk at covering Viagra® (Titlow et al. 2000). Many would reject dialysis as a *relevant* precedent for comparison. This thinking reveals that the crucial criterion is perhaps not cost-effectiveness, but rather, categorical differences between the two technologies' purposes.

Explicit coverage standards may affect the development of health technologies. When it becomes clear “how good is good enough,” innovators can make technologies “good enough” – or more perversely, seem to be. For categorical criteria, this may entail clearer articulation of a technology's uses and effects – reframing clinical endpoints, target populations and rationales. To meet a quantitative threshold – for example, for effectiveness – developers may design the technology for greater success, or enhance apparent effectiveness by refining patient selection or presuming adjunct resources such as supportive care. Cost standards create pressures to lower prices, but also to offload adjunct costs to other payers. Thresholds for cost-effectiveness may send signals to increase effectiveness or to lower prices. They may also lead developers to raise

the price of a new, effective technology to achieve a cost-effectiveness ratio just beneath threshold – raising both proprietary profits as well as health system costs.

## Illustration: The Search for a Standard of Cost-Effectiveness

One concerted effort to establish coverage standards has been the quest for a cost-effectiveness threshold for publicly insured health services. This case study illustrates the gap between our compelling need for standards and our incapacity to specify and apply them systematically. To establish a standard, scholars have proposed rules of thumb, imputed thresholds from actual decisions, or imported dollar values for human life from outside the health sector. Table 1 summarizes such estimates of a dollar-per-QALY threshold. A more ad hoc approach has been to identify individual covered technologies – the cervical Pap test, beta-interferon, mammography, Viagra® and others – as precedents for acceptable cost-effectiveness. References to allegedly precedent-setting technologies are found throughout the cost-effectiveness literature in healthcare, as well as in published opinions, news media and court records (Giacomini, 2005).

One threshold deserves special attention: the \$50,000 quality-adjusted life-year (QALY) figure. This popular rule of thumb is often cited as the accepted ceiling for fundable health services, with little justification, in US and Canadian cost-effectiveness research. Ubel (1999) notes that this standard originated in 1982, based on the estimated cost-effectiveness of renal dialysis, which has special significance in US health policy because a federal entitlement program for end-stage renal disease guarantees its public funding. Thus, it is considered an important precedent for US government willingness to pay. Ubel notes two important misconceptions. First, the precedent should probably be viewed as a floor, not a ceiling: by covering renal dialysis, the United States made a commitment to technologies costing *at least* \$50,000 per QALY, but we do not know if a higher cost per QALY would have changed the decision. A case in which a technology has been rejected for coverage because of unacceptable cost-effectiveness gives a more precise estimate of a precedent threshold. Second, the figure of exactly \$50,000 per QALY has persisted in policy and research literature since 1982, remarkably with no adjustment for inflation (Ubel 1999). It has crossed the border into Canada without adjustment for currency or inflation; cost-effectiveness evaluations from the United States and Canada still cite the \$50,000/QALY threshold. The present-day Canadian value of the 1982 US figure is approximately Cdn\$114,487/QALY.

Studies that impute cost-effectiveness thresholds from observed, usual patterns of policy decisions should not neglect outliers in their search for central tendencies. Exceptions can set precedents and become new standards in the minds of stakeholders. Outliers tell us how far decision-makers are willing to go – and in so doing, they locate the real thresholds. Rational arguments from fairness and other criteria, if loud enough, may succeed in holding decision-makers to extremes. For example, a study

asking “does NICE have a threshold?” (Towse and Pritchard 2002) neglected some outliers to induce that NICE’s threshold must be roughly £30,000 per QALY. Table 2 lists all the NICE decisions concerning technologies less cost-effective than this ostensible threshold. Three such technologies were recommended: riluzole, trastu-

TABLE 1. Some possible standards for cost-effectiveness of health technologies

Jurisdiction and origin	Reference: First author, year	Original value/ QALY*	2004 Cdn\$†
<b>Canada</b>			
Rule of thumb, intuitive	Laupacis 1992	1992 Cdn\$100,000	\$124,600
Rule of thumb, from US	Ubel 2003	1982 US\$50,000	114,487
<b>United Kingdom</b>			
National Institute for Clinical Excellence (NICE), mention in orlistat guidance	NICE 2001	2001 £30,000	63,191
NICE, imputed, 1999–2002 recommendations	Towse 2002	2002 £30,000	62,317
Value of life, unspecified method, road accident fatalities	Loomes 2002	2002 £30,000	62,317
<b>Australia</b>			
Pharmaceutical Benefits Advisory Committee, imputed, drug coverage recommendations, 1991–1996	George 2001	1999 Au\$76,000	77,848
<b>New Zealand</b>			
Pharmaceutical Management Agency, imputed from drug coverage recommendations, 1998–2001	Pritchard 2002	2002 NZ\$20,000	17,648
<b>United States</b>			
Value of life, median, 19 empirical WTP job risk studies	Hirth 2000	1997 US\$428,286	600,102
Value of life, median, 35 empirical WTP studies	Hirth 2000	1997 US\$265,345	371,794
Rule of thumb, proposed interim	Ubel 2003	2003 US\$200,000	254,702
Value of life, median, 8 empirical WTP contingent evaluation studies	Hirth 2000	1997 US\$161,305	226,016
Value of life, median, 8 empirical WTP safety studies	Hirth 2000	1997 US\$93,402	130,872
Rule of thumb, US standard, original year	Ubel 2003	1982 US\$50,000	114,487
Value of life, median, 6 human capital studies	Hirth 2000	1997 US\$24,777	34,717

WTP = willingness to pay

\*Where original values were expressed as ranges, the top of the range is given.

†2004 Cdn\$ based on Canadian currency values for original year based on purchasing power parity ratios, updated to 2004 values using the Canadian Consumer Price Index.

zamab/paclitaxel and etanercept/infliximab. Per QALY, these cost up to £43,500, £37,500 and £35,000, respectively. The least cost-effective technology reviewed was beta-interferon, at up to £104,000 per QALY; it was not recommended. Viewing this pattern with an eye to precedence and thus a focus on the outliers, the actual NICE threshold appears to lie somewhere between £43,500/QALY and £104,000/QALY, not at £30,000/QALY.

TABLE 2. NICE recommendations concerning technologies costing over £30,000 per QALY

	Recommendation	2002 £	2004 Cdn\$ per QALY
Beta-interferon and glatiramer acetate for MS	Reject	£104,000	\$216,032
Laparoscopic surgery for inguinal hernia	Restrict	50,000	103,861
Riluzole for motor neurone disease	Accept	43,500	90,359
Zanamivir (Relenza®) – all adults	Reject	38,000	78,935
Trastuzamab for metastatic HER2 breast cancer	Accept	37,500	77,896
Etanercept and infliximab for rheumatoid arthritis	Accept	35,000	72,703
Temozolomide for brain cancer – GBM	Restrict	35,000	72,703*
Temozolomide for brain cancer – AA	Restrict	35,000	72,703*
Topotecan for advanced ovarian cancer (per year of response)	Restrict	32,500	67,510**
Zanamivir (Relenza®) – at-risk adults	Reject	31,500	65,433
Cox-2 selective inhibitors	Reject	30,000	62,317

\*Per life-year gained (LYG), not QALY

\*\*Per year of response

Source: Adapted from Towse et al. 2002, appendix

Such inductive searches for standards can mislead for several reasons. Despite the appeal of a strict cut-off, cost-effectiveness thresholds appear malleable. Experience shows that even where there is an apparent threshold, “political” exceptions are made, as for example in the case of the New Zealand decision to cover beta-interferon (Pritchard 2002), or the UK decision to cover Relenza® (Smith 2000) contrary to negative, cost-effectiveness–based recommendations. However, dismissing such exceptions as “politics” neglects the fact that criteria other than efficiency may legitimately and rationally mitigate a cost-effectiveness threshold. Recommendations may be misattributed to one criterion (cost-effectiveness) without accounting for other criteria and

their associated standards. The upper limit of £104,000/QALY in this NICE example assumes that the reason for rejecting beta-interferon was based significantly on low cost-effectiveness. If the decision were based primarily on another criterion, then the cost-effectiveness ceiling was in fact not tested in this set of cases, and the inductive threshold may be higher. Indeed, many call for additional values to supplement cost-effectiveness information (despite methodological controversies about what the QALY does and does not capture), e.g., “perceived need in the community” and “seriousness of the intended indication” (George et al. 2001), equity (Pearson and Rawlins 2005) or

life-threatening conditions (Neumann et al. 2005). Cost-effectiveness thresholds are commonly mistaken for affordability thresholds – but a “good enough price” per QALY says little about whether a budget can afford the QALY that a technology “sells,” or the real sacrifices required to afford it (Birch

---

**Standards operate whether acknowledged or not, but they are fairest when predetermined, explicit and consistently applied.**

and Gafni 2006). More fundamentally, to search for a cut-off point presumes that a point exists. Some suggest that the relationship between incremental cost-effectiveness values and probability of rejection is “S”-shaped (Rawlins and Culyer 2004), with reluctance to approve rising gradually with the cost per QALY. To the extent that individual decisions are understood as precedents, extreme cases will steadily pull standards upwards. Finally, the necessary evidence is often missing or biased, and available evidence is sensitive to value-laden assumptions. Indeed, 13 of 54 NICE decisions were made in the absence of cost-effectiveness information (Towse and Pritchard 2002).

## Conclusions

We require standards to make coverage decisions that are consistent, principled and evidence-based. Standards operate whether acknowledged or not, but they are fairest when predetermined, explicit and consistently applied. Because we use multiple criteria to assess technologies for coverage, we need multiple standards – at least one for each criterion – and we need to understand better how these standards interact with one another in the formulation of recommendations and decisions. Quantifiable criteria require standards in the form of thresholds, representing, for example, categorically impressive effect sizes or the limit of our willingness to pay for any new service and its benefits. Categorical criteria require standards in the form of prioritized categories

of service, representing, for example, special health problems or clinical goals that have priority for public funding. Standards intended as hurdles for coverage may evolve into goals for research and development, organization, marketing or targeting of services. Policy signals about what is “good enough” can have both positive and perverse effects on technological innovation.

The example of cost-effectiveness thresholds offers important lessons for policy making. Current methods for articulating such thresholds are intuitive and ad hoc. Simple, round figures such as \$50,000 or £30,000 per QALY persist, despite inadequate justification and changes of inflation or currency. Induced thresholds from actual decisions could be misleading: “usual practice” does not point to real limits, limits may not yet have been tested in past cases and the role of other criteria (effectiveness, affordability, priorities among categorical purposes and populations and so forth) must be understood and interpreted. Standards for criteria other than cost-effectiveness are less well examined. The identification and application of standards should become a focus for more accountable and deliberative methods in decision-making related to health technology assessment and coverage (Abelson et al. 2007).

Correspondence may be directed to: Mita Giacomini, PhD, McMaster University, HSC-3H1C, 1200 Main Street West, Hamilton, ON L8N 3Z5; tel.: 905-525-9140 X22879; e-mail: giacomin@mcmaster.ca.

#### ACKNOWLEDGMENTS

Earlier versions of this paper were presented to the Ontario Health Technology Assessment Committee, the Canadian Agency for Drugs and Technology in Health Invitational Symposium and the Cancer Care Ontario Systemic Therapy Search Conference. I am grateful for the feedback received from participants in these meetings. I also thank Jeremiah Hurley, three anonymous reviewers and the editors for their helpful suggestions.

#### REFERENCES

- Abelson, J., M. Giacomini, P. Lehoux and F.P. Gauvin. 2007. “Bringing ‘the Public’ into Health Technology Assessment and Coverage Policy Decisions: From Principles to Practice.” *Health Policy* 82(1): 37–50. Epub 2006 Sep 22.
- Birch, S. and A. Gafni. 2006. “Information Created to Evade Reality (ICER): Things We Should Not Look to for Answers.” *Pharmacoeconomics* 24(11): 1121–31.
- George, B., A. Harris and A. Mitchell. 2001. “Cost-Effectiveness Analysis and the Consistency of Decision-Making: Evidence from Pharmaceutical Reimbursement in Australia (1991 to 1996).” *Pharmacoeconomics* 19(11): 1103–9.

- Giacomini, M. 1999. "The 'Which' Hunt: Assembling Health Technologies for Assessment and Rationing." *Journal of Health Politics, Policy, and Law* 24(4): 715–58.
- Giacomini, M. 2005. "One of These Things Is Not Like the Others: The Idea of Precedence in Health Technology Assessment and Coverage Decisions." *Milbank Quarterly* 83(2): 193–223.
- GRADE Working Group, D. Atkins, D. Best, P. Briss, M. Eccles, Y. Falck-Ytter et al. 2004. "Grading Quality of Evidence and Strength of Recommendations." *British Medical Journal* 328(7454): 1490.
- Hirth, R., M. Chernew, E. Miller, A. Fendrick and W. Weissert. 2000. "Willingness to Pay for a Quality-Adjusted Life Year: In Search of a Standard." *Medical Decision-Making* 20(3): 332–42.
- Laupacis, A., D. Feeny, A.S. Detsky and P.X. Tugwell. 1992. "How Attractive Does a New Technology Have to Be to Warrant Adoption and Utilization? Tentative Guidelines for Using Clinical and Economic Evaluations." *Canadian Medical Association Journal* 146(4): 473–81.
- Loomes, G. 2002. "Valuing Life Years and QALYs: 'Transferability' and 'Convertability' of Values across the UK Public Sector." In A. Towse, C. Pritchard and N. Devlin, *Cost Effectiveness Thresholds* (pp. 46–55). London: King's Fund.
- National Institute for Clinical Excellence (NICE). 2001. "Technology Appraisal Guidance No. 22: Guidance on the Use of Orlistat for the Treatment of Obesity in Adults." London: Author.
- Neumann, P.J., A.B. Rosen and M.C. Weinstein. 2005. "Medicare and Cost-Effectiveness Analysis." *New England Journal of Medicine* 353(14): 1516–22.
- Pearson, S.D. and M.D. Rawlins. 2005. "Quality, Innovation, and Value for Money: NICE and the British National Health Service." *Journal of the American Medical Association* 294(20): 2618–22.
- Pritchard, C. 2002. "Overseas Approaches to Decision-Making." In A. Towse, C. Pritchard and N. Devlin, *Cost Effectiveness Thresholds* (pp. 56–68). London: King's Fund.
- Rawlins, M.D. and A.J. Culyer. 2004. "National Institute for Clinical Excellence and Its Value Judgments." *British Medical Journal* 329: 224–27.
- Smith, R. 2000. "The Failings of NICE." *British Medical Journal* 321: 1363–64.
- Titlow, K., L. Randel, C.M. Clancy and E.J. Emanuel. 2000. "Drug Coverage Decisions: The Role of Dollars and Values." *Health Affairs* 19(2): 240–47.
- Towse, A. and C. Pritchard. 2002. "Does NICE Have a Threshold? An External Review." In A. Towse, C. Pritchard and N. Devlin, *Cost Effectiveness Thresholds* (pp. 25–30). London: King's Fund.
- Towse, A., C. Pritchard and N. Devlin. 2002. *Cost-Effectiveness Thresholds: Economic and Ethical Issues*. London: King's Fund.
- Ubel, P.A. 1999. "How Stable Are People's Preferences for Giving Priority to Severely Ill Patients?" *Social Science and Medicine* 49(7): 895–903.
- Ubel, P.A. 2003. "What Is the Price of Life and Why Doesn't It Increase at the Rate of Inflation?" *Archives of Internal Medicine* 163: 1637–41.