APPROACHES TO IDENTIFYING, MEASURING, AND AGGREGATING ELEMENTS OF VALUE

Adrian Towse, Paul Barnsley

Office of Health Economics

Background: Two general alternative approaches, cost-effectiveness analysis and the therapeutic added value approach, link the pricing and approval of drugs to value. Value as assessed by payers is a function of: benefit less cost, willingness to pay for benefit, and how they handle uncertainty.

Methods: This study uses international examples to explore the elements of value that can be included in the assessment of health technologies, approaches to scoring the elements of value and how they can be combined to make a decision.

Results: A range of value elements, measures, and approaches to aggregation are identified across different HTA systems. We show that seemingly arbitrary differences in measurement and aggregation can lead to significantly different outcomes, and argue that the choice of values, measures, and decision-making processes should be informed by the societal values that underpin a health system.

Conclusions: We identify three areas for further research to improve both health system and industry R&D decision making: (i) whether more consistency could be achieved across health systems on the elements of value that matter; (ii) the relative merits of discrete versus continuous measures of value; and (iii) how structured decision making (to aggregate the elements of value) could or should become.

Keywords: Cost-effectiveness analysis, Cost-effectiveness threshold, Deliberative processes, Pharmaceutical pricing, Pricing and reimbursement, Therapeutic added value

Most industrialized countries have universal coverage for pharmaceuticals with modest patient co-payments. However, because such insurance makes patient demand highly price-inelastic, public and private insurers use various forms of pharmaceutical price regulation to constrain producer moral hazard. We distinguish between two major approaches that explicitly aim to measure value:

1. Cost-effectiveness analysis (CEA). Using CEA, drugs are assessed for use or for a reimbursement price by projecting the incremental health-related effects (often measured and valued using the quality-adjusted life-year (QALY) and incremental costs relative to existing treatments. Economists regard the use of CEA for drugs (which has the effect of regulating drug prices *indirectly* through a review of cost-effectiveness) as being, in theory, consistent with principles of efficient resource allocation (1). Over the past 20 years, there has been a substantial increase in the number of public and private third-party payers using formal CEA for assessing the value of drugs, vaccines, and other health technologies. Countries using this approach include Australia, New Zealand, several Canadian provinces, the United Kingdom, and Sweden.

2. Therapeutic added value (TAV). TAV assessments typically involve comparison with other, established drugs in the same class, or with other treatments used in the standard of care (SoC) with higher prices allowed or negotiated for improved health or other elements of value recognized by payers. If companies are able to charge higher prices when they can demonstrate superior effect over other relevant products, then prices are taking

account of the value generated for payers and their patients. This can be achieved by using an assessment of 'relative effectiveness' (the term used in Europe) or "comparative effectiveness," the term used in the United States. Countries using this approach include the German Arzneimittel-marktneuordnungsgesetz (AMNOG) pricing system, the current French system, and U.S. private sector payers.

Both the use of CEA and the TAV approach link price to value. Price can, therefore, be thought of as a function of the decision-maker's perception of value.

For the decision maker, we can further decompose value as additional benefit minus additional cost. These costs can be thought of as comprising additional costs associated with using the technology (excluding acquisition cost, i.e., "price") minus cost-offsets [including the costs saved by the displacement of other technologies]. In addition, decision makers weighing value are also concerned about the opportunity cost of resources. In the case of payers using CEA, this is explicit (although they may not say what opportunity cost threshold they are using). In the case of payers rewarding manufacturers with price premiums for value, it is implicit in their willingness to pay higher prices for additional value. A rule of thumb is usually used in a TAV system to estimate the price premium they are willing to pay for additional value (for example, by reference to prices sought elsewhere by the company) or a price is negotiated.

Finally, decision makers are concerned about the uncertainty of the evidence associated with their estimation of value. Substantial uncertainty is likely to lead to a lower price, delay in use of the drug pending resolution of the uncertainty with more evidence, or some form of use linked to the collection of evidence designed to resolve the elements of uncertainty

(often called coverage with evidence development or managed entry) (2). As such, the decision-makers' value determination is a function of these four elements: Benefit, Cost, the opportunity cost of funds, and uncertainty.

Moving from the concept of value to making a decision on the value of a particular drug involves three steps: First, identifying the elements of value to be included, then determining how to measure and gather evidence of each of those value elements, and finally how to aggregate the combined elements of value in arriving at a decision.

Concepts of Value

For most decision makers, the health effect is usually the single most important benefit and hence element of any assessment of value, while cost-offsets within the healthcare system are a second key benefit. Uncertainty in the measures of cost and health gain also tends to influence decision makers.

Other elements of value that are *sometimes* recognized by decision makers fall into four distinct types:

First, the "value" of the health gain to society may be higher or lower depending on who gets it. The severity of the disease is a particular factor. For example, the UK National Institute for Health and Care Excellence (NICE) applies a specific value weight when appraising end-of-life medicines. Several health systems treat drugs for orphan diseases differently (where a requirement for designation is that the degree of disease severity is high), allowing in this case higher prices and/or lower evidence standards for evidence of relative effectiveness or therapeutic added value. In the German AMNOG process, orphan drugs are automatically assumed to be innovative without a consideration of the strength of the evidence (3). In the United Kingdom, some orphan drugs were exempt from the NICE review process, but this has now changed. However, in the future, NICE will use a different process to review these drugs as compared to its conventional CEA approach (4).

Second, there may be elements of benefit to the patient that are not necessarily captured in the measure of health gain. These can include health-related quality-of-life (QoL) aspects not well captured in a generic measure of health gain such as the EQ-5D that may be important in some disease areas. In addition, traditional measures of health gain often exclude health-care-process-related aspects such as replacing an injectable with an oral formulation, or being treated with dignity and at a convenient time and location, and after only a short wait. There may also be value to the patient of information which, for example, enables lifestyle choices to be made, independent of any health effects that may arise.

Third, systems may consider other costs and benefits beyond those to patients and the health care system. Outside of health care, an economy-wide (or societal) perspective is conventionally used by economists (5), including all costs and consequences related to the initial interventions in a cost-benefit analysis. Applying such an approach would involve expanding the CEA to include some or all of unrelated medical costs, productivity effects, costs incurred outside the healthcare sector, and benefits accruing to all stakeholders in society including the patient's family. Several countries, including Norway, Sweden, and the Netherlands, already require that economic evaluations are conducted using a societal perspective (6).

Finally, innovative attributes of a technology may be deemed to have value independently of the health gain generated. Japan and Italy use a categorical rating to assess the degree of innovativeness. France and Germany use categorical rating to estimate the degree of clinical TAV. The innovation issue is arguably the most controversial between payers and the pharmaceutical industry. The argument that there is an independent value for innovation over and above the health effect, or even that innovation can be objectively defined is not readily accepted by payers, given that all new technologies are in some sense innovative (7). One articulation of the issue is to treat it as the purchase of an option on future products which are developed as a consequence of approving a product today (8). If rewarding the first-in-class product makes it more likely that another, better, product comes along sooner, then in principle this has value to payers additional to the immediate health effect that is being delivered. Using an option framework allows consideration of questions such as if and how such a value could be established.

Ideally, the set of values recognized by a nation's system of HTA should reflect that nation's societal preferences, and differences between systems should reflect genuine differences in those nations' preferences, rather than accidents of history or administrative inertia. We believe that there is value in further research to determine the appropriate level of convergence in values considered between nations.

It will often be impossible to arrive at exact estimates of the degree of value provided by a technology across each of these value elements. Where uncertainty exists, decision makers will, in practice, tend to reward products that are able to provide relatively certain, narrow estimates as to the range of possible outcomes, over those for which the value produced is less certain. In a sense, certainty of outcome is treated by decision makers as an element of value, in and of itself. This raises two questions:

1. Under what circumstances should certainty be valued, over and above the expected (average) value a product provides?

2. If uncertainty is to be taken account, how should it be factored into the decision-making process?

In relation to the first question, there is a sound theoretical case that government payers, and hence decision makers acting on their behalf, should be risk neutral across a known distribution of value: Governments make a large number of decisions, and allocate large sums of money. This means individual decisions which realize negative outcomes should tend to average

out when considered against positive realizations of uncertainty in other cases, suggesting a very low variance in overall value when measured across all government decisions. An HTA system working according to this approach would simply ignore uncertainty in elements of value (most obviously health gain and cost) and focus only on central estimates. There are at least two reasons why such a decision maker may choose not to behave in this way. First, rather than taking the available information as fixed, it may be optimal to seek additional evidence to reduce uncertainty if this can be linked to an opportunity to change the decision, and second, because better evidence may have a value, payers may wish to encourage investment in better evidence collection before the submission of a clinical or reimbursement dossier. Note that these arguments abstract away from the technical challenges associated with calculating an expected distribution of value in the presence of structural uncertainty (9).

Some health systems, including Germany's, place a very high value on the certainty of an outcome, in and of itself, and adopt apparent "risk averse" behavior based on the perceived variance of evidence rather than the most likely parameter estimates it suggests. They might believe that uncertainty as to health gain cannot be treated in the same way as financial risk aversion, for example because losses in health are in some sense disproportionately worse than equivalent gains. Finally, decision makers might believe that the data they are presented are biased as a result of the interests of the parties producing them, and that proper skepticism requires consideration of low-end estimates as well as the outcomes that are the most probable.

Proper treatment of risk may depend on the element of value to which it relates (comparable confidence intervals around, say, cost and health gain may not lead to comparable levels of concern), and on why certainty is being valued in the first place: risk aversion is a property which emerges from how the decision maker, and ultimately the society, values the attribute in question, and how that value changes at the margin (relative to the other factors that matter to the decision maker). The marginal social value of health gain and of health funding are important areas for further research.

Measuring Value Added

Both health effects and non-health effects have to be measured and supported with evidence if they are to be included in an assessment of value. This involves three steps, which we can illustrate with the case of health effects: *measurement*, through the use of QALYs, clinical outcomes, patient-reported outcomes (PROs), or disease-specific instruments; *evidence collection* by means of Randomized Controlled Trials (RCTs), observational studies, patient testimony, or clinical opinion; and *valuation*, by reference to population or patient values and the use of categories or discrete scales.

We do not discuss measurement or evidence support further in this paper. We focus on the valuing or rating of the measure

of effect given the evidence. The effects are all valued or rated, either explicitly or implicitly. The QALY uses population values of health states; hence, the need for "national" valuation sets. Populations may differ in their willingness to trade length of life and different levels of functioning that underlie health-related quality of life. With disease-specific and PRO-based measures, we are measuring quality of life as assessed by the patient. It is left to the decision maker to place a value on the measured effect.

Continuous or Discrete Measurement Scales

A key decision when considering how to measure an agreed value element is whether the scale of measurement used should be discrete or continuous.

This decision should first turn on whether the concept of value being measured is defined up to a continuous, interval scale (monetary values have a natural interpretation along a continuous scale, for instance) but also whether sufficient *evidence* is available to enable the value to be divided up along a continuous scale. In the face of uncertain evidence, using a continuous scale to measure a variable can lend false sense of precision to a given attribute. This is true even where the attribute in question naturally lends itself to a continuous interpretation: the wider economic effects of a health intervention are ultimately a continuous monetary variable. But if the HTA system lacks the evidence to estimate these monetary values with any precision, then it may be more accurate to assign these effects to one of a few discrete values than to use a precise but inaccurate estimate.

The key advantage of adopting a continuous scale is that it avoids the necessity of large "jumps" in the aggregate value of a product at the cutoff between points on the discrete scale. In the context of burden of disease, we can note that the current UK NICE "end of life" adjustment treats patients who have six months or less to live as up to 70% more valuable than those with a little over six months to live (10). In the French system, the Service Médical Rendu (SMR) rating of "major or important" innovation has two categories of disease burden "severe" and "non-severe," which determine the patient co-payment level. The granting of orphan status by a regulator is a binary decision based on disease severity and rarity. As we noted earlier, orphan drugs are often treated differently to non-orphan drugs by payers and HTA bodies assessing value.

In practice, decision makers with discretion may avoid some discontinuities in the measurement of value, but such discretion either shifts the point at which the jump in value occurs or generates an implicit scale applied for values close to the cutoff. In either case, it will usually be better to be explicit about how values close to the cutoff are to be handled.

Even where a decision has been made to use a continuous measurement scale, the choice of scale can make a substantial difference to the treatment of a product by an HTA system. We can illustrate this with a consideration of competing

measures of severity or burden of illness (BoI) effects, the widely accepted proposition that health gains accruing to the worse off have additional value relative to those accruing to the better off (11). Measuring BoI using "absolute shortfall" assumes that society views "worse off" as meaning patients who will, without additional treatment, experience a large *absolute* gap between their current prognosis and the number of remaining QALYs they would expect to enjoy were they fully healthy. Conversely, the "proportional shortfall" measure assumes that the "worst off" are those who currently expect to forego the largest *proportion* of their remaining lifespan when calculated as if they were healthy. Two recent, as yet unpublished, surveys provide estimates societal preferences defined across these measures (12;13) and suggest that the strength of social preference between the best and worst off according to each measure is relatively similar. The key difference between these two approaches to measuring BoI, however, is *which* groups of patients occupy the position of "worst off": The "absolute" approach will tend to give greater weight to younger patients, while the "relative" approach measures severity independently of age and can assign high measures of burden to elderly patients who expect to lose a high proportion of their remaining QALYs.

Decision Making: How to Aggregate the Elements of Value

Most payer HTA bodies have a committee to appraise evidence and make a decision. The mechanism by which the members of a committee combine the various forms of evidence with local context and judgments about interpretation and uncertainty to reach a "decision on value" is a deliberative process. It involves two types of challenges for decision makers.

The first challenge is appraising the evidence in circumstances where there is either uncertainty about technical information (needing scientific judgments), or issues relating to fairness and social values (needing value judgments) need to be taken into account. Culyer (14) defines scientific judgment as, "usually about an effect . . ., its size, the ways in which it can be achieved, for whom, for how long, and how much uncertainty there is about the outcomes" and value judgments as "tend[ing] to be in a different territory but . . . might be about, for example, how worthwhile a technology is, how defensible the tough bits of the decision are, how tolerant of uncertainty the committee ought to be, .. inter-personal comparisons, ..whether the [outcome measure] is a good tracker of the relative health benefits of the interventions that were compared".

The second challenge is to weight the multiple criteria relevant to the decision using a combination of deliberative processes and algorithms. At one extreme, a pure deliberative process does not use any formal structure and so is a "black box" to outsiders and potentially to committee members themselves, which may lead to a lack of consistency and a lack of clear signals as to what matters, and, at the other extreme, a pure

algorithmic approach does not need a committee but simply an administrator who puts the numbers into the formula.

Public payers are typically trying to reflect social and/or political preference and in some cases interpreting statutory or regulatory responsibilities. Where these criteria are clear, and measurement, evidence requirements, and rating is also pre-agreed, then a formulaic approach can be used for these elements of value. Even here, however, where the value judgments are "pre-set" the scientific judgments may not be, particularly around uncertainty about the evidence. And in reality, even if some criteria are clear, value judgments will be required for others.

This raises the question as to whether decision support tools can improve the transparency and effectiveness of a deliberative process used by a payer and/or HTA body. Multi-criteria decision analysis (MCDA) methods have been advocated for use in health care priority setting (15). MCDA is a methodology for appraising options on multiple (often conflicting) criteria with the goal of providing a combined appraisal that includes an overall ordering of those options. It provides a framework for explicitly trading off various objectives against each other. It is particularly useful when these objectives do not share a common unit of valuation, and when aggregating the elements of value typically involves mixing health, monetary, distributional, and political objectives.

Use of MCDA in this context could be attractive if it led to processes becoming more transparent and systematic, so improving both the signals sent to patients and drug developers, and the quality of decision making. However, it might require a greater time commitment on the part of decision makers. The burden on decision makers of using this approach would need to be proportional to the benefits of improved decision making. To date, no HTA body is using formal MCDA techniques, although the EMA has explored its use for regulatory decision making (16).

There are several difficult issues in combining criteria to support decisions: avoiding unintentional double counting (an element of value is captured under two or more different headings), the need to handle uncertainty appropriately, and the need for appropriate willingness-to-pay measures for heterogeneous elements of value (more important for CEA than TAV approaches). It may also be the case, understandably, that decision makers prefer a "black box" element to reduce legal and political challenge to their decisions. However, it may be possible to improve the *internal* clarity of what committee members are valuing without necessarily removing their collective choice as to how much of this they choose to tell the rest of the world.

As the set of value elements considered becomes larger, decision-making may become more complex and more difficult. In the absence of a complete set of MCDA weightings for trading off different criteria, one, albeit imperfect, response to this difficulty, observed in several systems including those of

Spain and Germany, is to treat several categories as explicitly secondary and to consider these less important characteristics only in situations where they are likely to alter the decision made on the basis of primary sources of value, typically, health gain and sometimes cost.

The difficulties with this approach are twofold. First, without an overall model for how "secondary" values are to be traded off against health gain, decision makers cannot be certain whether a particular case is marginal enough to warrant inclusion of secondary criteria; but once there is a proper model of the relative weighting of all elements of value, then it becomes much more straightforward to consider the full set of criteria in all cases, not just those that fall close to the line.

Second, looking at extended criteria only when they are likely to change the decision is a useful strategy when the decision is a binary "yes" or "no," or covers a relatively small number of categories, but becomes less useful as the HTA system attempts to identify increasingly fine distinctions in products' overall value. At the extreme, where the HTA system precisely specifies a price for a product, every source of value may need to be considered in every decision.

CONCLUSION

We believe that it is useful to characterize HTA systems according to what they value, how they measure that value, and how they aggregate those measures in reaching a decision.

We identify three areas for further research to improve health system and industry R&D decision making: (i) whether more consistency could be achieved across health systems on the elements of value on that matter, or whether these differences reflect genuine societal differences between nations; (ii) the need for a deeper understanding of the impact of choosing between continuous and discrete scales for assessing elements of value; and (iii) how structured decision making could or should become.

CONTACT INFORMATION

Adrian Towse (atowse@ohe.org) **MA, MPhil,** Director, **Paul Barnsley, BA (Hons),** Economist, Office of Health Economics, London, UK

CONFLICTS OF INTEREST

REFERENCES

1. Danzon PM, Towse A, Mestre-Ferrandiz JM. Value-based differential pricing: Efficient prices for drugs in a global context. NBER Working Paper w18593. 2012 December.
2. Henshall C, Schuller T, Mardhani-Bayne L. Using health technology assessment to support optimal use of technologies in current practice: The challenge of "disinvestment". *Int J Technol Assess Health Care*. 2012;28:203-210.
3. Hensche C, Sundemacher L, Busse R. Structural changes in the German pharmaceutical market: Price setting mechanisms based on the early benefit evaluation. *Health Policy*. 2013;109:263-269.
4. Anon. NICE to assess high cost drugs for rare conditions [internet] London (UK) NICE 2012 July 20. http://www.nice.org.uk/newsroom/news/NICEToAssessHighCostDrugsForRareConditions.jsp (accessed August 1, 2013).
5. Garrison LP, Mansley EC, Abbott TA, et al. Good research practices in cost-effectiveness analyses: A societal perspective: The ISPOR Drug Cost Task Force Report—Part II. *Value Health*. 2010;13:8-13.
6. Sussex J, Towse A, Devlin N. Operationalising value based pricing of medicines: A taxonomy of approaches. *Pharmacoeconomics*. 2013;13:1-10.
7. Claxton K. OFT, VBP: QED? *Health Econ*. 2007;16:545-558.
8. Cook J, Golec JH, Vernon JA, Pink GH. Real option value and path dependence in oncology innovation. *Int J Econ Bus*. 2011;18:225-238.
9. Bojke L, Claxton K, Sculpher M, Palmer S. Characterizing structural uncertainty in decision analytic models: A review and application of methods. *Value Health*. 2009;12:739-749.
10. Longson C and Littlejohns P. *Update Report on the Application of the End-of-Life. Supplementary Advice in Health Technology Appraisals*. London: NICE; 2009.
11. Shah KK. Severity of illness and priority setting in healthcare: A review of the literature. *Health Policy*. 2009;93:77-84.
12. Nord E, Johansen R. 2013. Personal communication.
13. Brazier J. 2013. Personal communication.
14. Culyer AJ. *Deliberative processes*. London: Office of Health Economics; 2009.
15. Devlin N, Sussex J. Incorporating multiple criteria in HTA: Methods and processes. London: Office of Health Economics; 2011.
16. Phillips LD, Fasolo B, Zafiropoulos N, Beyer A. Is quantitative benefit-risk modelling of drugs desirable or possible? *Drug Discov Today Technol*. 2011;8:e3-e10.