The Clinical Utility Index as a Practical Multiattribute Approach to Drug Development Decisions

B Poland¹, FL Hodge¹, A Khan², RT Clemen³, JA Wagner², K Dykstra¹ and R Krishna²

We identify some innovative approaches to predicting overall patient benefit from investigational drugs to support development decisions. We then illustrate calculation of a probabilistic clinical utility index (CUI), an implementation of multiattribute utility that focuses on clinical attributes. We recommend use of the CUI for the support of early drug development decisions because of its practicality, reasonable accuracy, and transparency to decision makers, at stages in which financial factors that may dominate later-phase decisions are less critical.

Pharmaceutical development teams face complex decisions that require tradeoffs among the desirable and undesirable attributes of their investigational new drugs. For example, which would patients prefer: a highly effective compound with poor tolerability or a less effective but well-tolerated compound? How does one compound’s various side effects—some rare but serious, some common but mild—compare with those of another (an internal backup or a competitor)? Should the dose be twice daily to smooth out drug exposure or once daily to improve ease of use (especially if the doses of competitors or concomitant medications are already once daily)? Higher doses generally increase efficacy but worsen tolerability and safety; given this, which is the dose that would trade off these attributes optimally? Also, how should development planning account for varying degrees of uncertainty in all the attributes? We review a simple, practical version of a multiattribute utility approach to such questions—the clinical utility index (CUI)—in the context of drug development in pharmaceutical companies, particularly during early clinical development.

A variety of methods have been proposed for quantifying the benefit vs. harm or risk to patients of treatments in order to support drug development decisions. Any of these methods can (and usually should) be made probabilistic: uncertainty in the inputs can be modeled with probability distributions to produce probability distributions for the output, so that, for example, we compute the probability that a drug’s net benefit will exceed that of a competitor drug. Multiattribute utility theory is a theory for making decisions among alternatives with multiple relevant attributes, without reducing the attributes to a common measure, such as money, before calculating utility.¹ ² Another approach uses financial valuation, determining monetary value and cost over a period of time (often summarizing expected net present value of future cash flows to the pharmaceutical company). Alternatively, public health measures could be considered, such as the number needed to treat and the number needed to harm, which are the average numbers of patients that must be treated until the treatment succeeds or causes a harmful effect, respectively. Another health economic metric, quality-adjusted life years, reflects both benefits and harms or risks of treatment in terms of equivalent healthy life years gained. The Supplementary Data online compare these approaches.

The CUI³–⁶ is a multiattribute utility that is normally calculated with an additive utility function and assessed with swing weighting⁷ for practicality. Mathematically, it is expressed as:

$$\text{CUI} = \sum w_i U_i(x_i)$$

with $i$ indexed over $m$ drug attributes, $w_i$ corresponding to the overall importance weight of the $i$th attribute, and $U_i(x_i)$ being a utility function for that attribute (see Supplementary Data online for details). The attributes are typically limited to clinical attributes of the product profile; those directly related to market value or development cost or time are omitted. Therefore, the CUI is not intended for choices among compounds for which these differ markedly, either with certainty or in terms of probabilities. Also, the CUI is a product-level measure, not a patient-level one, in that it does not attempt to capture the variability of individual patient preferences. Because of the high levels of

¹Pharsight Corporation, Strategic Consulting Services, Mountain View, California, USA; ²Merck & Co., Whitehouse Station, New Jersey, USA; ³Duke University Fuqua School of Business, Durham, North Carolina, USA. Correspondence: R Krishna (rajesh_krishna@merck.com)

Received 4 February 2009; accepted 6 April 2009; advance online publication 20 May 2009. doi:10.1038/clpt.2009.71
uncertainty in product profiles during development, probability distributions are assessed for the attributes, which are then translated to a probability distribution for the CUI itself.

**AN EXAMPLE OF A CUI**

As an illustration of the practical application of multiattribute utility theory, consider a go/no-go decision for a drug approaching large-scale, expensive confirmatory trials. With competitors already in the market, would this drug be commercially viable at any dose? The development team identified 10 key efficacy and safety attributes and agreed that these made essentially additive contributions to utility. Two or three response categories were defined for each attribute, relative to the key competitor or placebo. For example, the first attribute's outcomes were divided into three categories: worse, equivalent, or better relative to the competitor's primary end point, efficacy, as shown in Table 1. The corresponding individual utility function, scaled from 0 for the worst case to 1 for the best case, required just one utility assessment, namely, for the middle case. This utility represented the probability of the best vs. worst case that makes the decision maker indifferent between this gamble and the middle case for sure (Figure 1; this utility concept and standard gamble have a long history in von Neumann–Morgenstern utility functions, and they have been used in health economics). For attributes that are assigned only two levels, such as a safety attribute measured by the presence or absence of serious adverse events, no individual utility function assessments were needed.

Next, attribute weights were assessed in terms of the impact on utility of swinging each attribute from its worst to best case, relative to swinging the attribute with the greatest impact (see Supplementary Data online). These weights were rescaled to sum to 1, completing the utility function assessment. Weights were elicited from the project team in this example, but could also emerge from the results of a formal conjoint or discrete choice analysis, testing the characteristics of a hypothetical marketable therapy for a given indication.

In parallel, the team built drug–disease models describing exposure–response for each attribute of their compound based on the available studies through phase IIA, as well as for the competitor, based on published studies. For example, the primary efficacy end point was described by a log-linear exposure–response model with variability across patients: higher doses tended to have greater effect but at a decreasing rate. Simulations of a large patient population (incorporating model parameter uncertainty) produced response predictions, and these were accumulated into probabilities for the response categories. For the primary efficacy attribute, for example, at a particular dose, 40% of the simulated population mean values fell into the “worse than competitor” category, 45% into the “equivalent” category, and the remaining 15% into the “better” category. As the dose

---

**Table 1 Clinical utility construct**

<table>
<thead>
<tr>
<th>Attribute, $i$</th>
<th>Attribute weight, $w_i$</th>
<th>Response level, $x_{ij}$</th>
<th>Response utility, $u_{ij}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Efficacy 1</td>
<td>0.27</td>
<td>Worse than comparator</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Equivalent to comparator</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Better than comparator</td>
<td>1.0</td>
</tr>
<tr>
<td>2. Tolerability effect 1</td>
<td>0.27</td>
<td>Worse than comparator</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Same or better than comparator</td>
<td>1.0</td>
</tr>
<tr>
<td>3. Class effect 2</td>
<td>0.14</td>
<td>Worse than comparator</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Same or better than comparator</td>
<td>1.0</td>
</tr>
<tr>
<td>10. Food effect on pharmacokinetics</td>
<td>0.01</td>
<td>Present</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Absent</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**Figure 1** Assessing utilities and weights in the go/no-go decision example. (a) The utility of the middle efficacy case is the probability $P$ that would make one indifferent between the gamble for best-case vs. worst-case efficacy and the middle efficacy with certainty. (b) The weight for efficacy in an additive utility function is the probability $P$ that would make one indifferent between the gamble and certain outcome shown.
was increased, the proportions moved toward the “better” category—but, of course, side effects increased as well.

Using the exposure–response models to extrapolate doses below and above previously studied doses, the team found a clear relationship between the CUI and dose but found no feasible dose at which the expected CUI of their compound reached that of the key competitor, and likewise the probability that the CUI would reach that of the competitor was very small at any dose. Sensitivity analysis confirmed that this was because of the key side effect, which was likely to occur more frequently than for the competitor drug. Even if efficacy were definitely better than the competitor drug, the expected CUI would not have become competitive at any dose. Development of the compound was therefore terminated. Senior management was satisfied with the decision, as it saved the huge cost of late-phase trials involving likely inferior treatment of many patients. This decision allowed attention to be focused to another project with a greater chance of success.

DISCUSSION

The CUI is a tool that drug development teams can use to quantitatively determine optimal tradeoffs among key drug attributes and to clarify the evaluation of a potential new medicine’s expected clinical properties. In our experience, drug development teams that create a CUI emerge with a more transparent picture of the probability of success in the target indication than they had before undertaking the exercise. In the example cited in the previous section, the CUI helped make ambiguous, potentially contradictory decision criteria clearer, where a go/no-go decision needed to be made in the face of incomplete exposure–response exploration. In another case, the use of a CUI allowed the identification of biopharmaceutical characteristics that mitigated tolerability issues while increasing the duration of therapeutic effect; this led to a viable development effort. In a recently published example, the CUI methodology helped the development organization halt expensive, parallel development of a backup compound.

The example just described also raises various caveats and suggests potential enhancements. First, CUI is better suited for a no-go than a go decision: if the expected CUI, based on efficacy and side effects, had been higher than that of the competition, other factors might still favor a no-go decision, such as development cost and time, and even opportunity cost of allocating resources to this compound vs. others in development. Therefore, a competitive product profile is necessary but not sufficient for a go decision. Sometimes a CUI analysis is sufficient to clarify a decision, as in this case; otherwise, the CUI can be a prelude to a more comprehensive analysis.

Another caveat is that results may be sensitive to subjective inputs. However, this is an advantage in that many drug development decisions require subjective judgment, whether acknowledged or not; the CUI methodology documents these assessments in a systematic way and puts decision criteria into a transparent format that can be subjected to sensitivity analysis. In the case described, the development team determined that the weight of the key side effect attribute would need to be cut in half before the compound (at the highest dose) would become comparable to the competitor with respect to the CUI. Such sensitivity tests are important for gaining confidence in the results. If greater sensitivity had been found, it might have been worthwhile to invest in market research (e.g., a conjoint analysis) to increase confidence in how patients or health-care providers would trade off attributes. Indeed, as a program progresses and the development organization gains experience in a therapeutic area, market research that can support the CUI assessments is often done regardless.

A third caveat is that the analysis might deserve refinement, e.g., more categories for each attribute or continuous values (as in the drug–disease modeling output). This is conceptually straightforward but did not seem important in this instance because the decision did not appear to be a close call. Moreover, discrete categories are helpful when competitors or placebo treatment (which can also be considered a competing treatment) define sharp nonlinearities in market share and hence utility from the developer’s perspective. Market share can be expected to jump when attribute levels exceed those of all competitors and plummet when they are all worse, because one would not knowingly buy a product dominated by another in all attributes, even if its net benefit is almost as great.

These competitive effects are illustrated in Figure 2, which assumes only two attributes and shows indifference curves

**Figure 2** A two-attribute example showing indifference curves and competitive effects. More is better on both scales, but at a decreasing rate. (a) Indifference curves starting at 0 utility and two competitors, A and B. Broken lines swing utility from 0 (worst case) to 0.5 (best when swinging one attribute at a time). (b) Likely competitor effects on market share—not reflected in utility function.
that a linear weighting of safety with other attributes cannot be effective in making decisions because of its practicality and transparency, although it is particularly useful for supporting early drug development in different ways: in terms of utility, financial value, number of patients treated, and therefore probably most of the market share, all else being equal. Competitor A may still be commercially viable, given that patient needs and preferences vary. However, products that fall in the lower left shaded areas are likely to have no market share, because there is a competitor that is better in both attributes; products that fall in the upper right shaded areas dominate one or both competitors and are likely to take away their market shares. Thus, competitors can introduce steep nonlinearities in market shares that the CUI does not attempt to capture. When the CUI is very uncertain, the expected CUI can average over dominated or dominating possibilities whose value to the drug developer (as opposed to the patient) in the face of competition is not properly represented by the CUI. If this is a concern, we may need to replace the CUI with a table relating possible product profiles to market share (and perhaps price), or relate CUI to market share with a market model, and then use market share or a full-fledged financial valuation to support drug development decisions.

The CUIs of key competitors can provide a useful replacement for the concept of target product profile for a compound or class of compounds. The target typically specifies performance levels for various attributes, whereas the CUI more realistically trades off attributes; for example, safety or tolerability deficiencies might be acceptable in return for best-in-class efficacy. In this manner, the target could become not a profile but a CUI, set to equal or exceed that of a top competitor.

Can a CUI capture safety issues adequately? It might seem that a linear weighting of safety with other attributes cannot rate an unsafe drug properly. But the individual utility functions for safety attributes (like any attributes) can be as nonlinear as desired and can be weighted so that a small risk of serious harm reduces the CUI as much as very high efficacy increases it. This will tend to favor safer but less effective competitors (including the option of no treatment). Considering the CUIs of these alternatives, we have not found difficulties with the assumption of additive utilities, for well-chosen attributes in the context of early-phase development decisions.

In conclusion, multiattribute utility theory, financial value, number needed to treat and number needed to harm, and quality-adjusted life year approaches all have useful roles in the quantification of net patient benefit. Despite differences from the factors traditionally emphasized, all the approaches are essentially multiattribute in nature, while measuring the attributes in different ways: in terms of utility, financial value, number of patients, or equivalent healthy years of life. The CUI method is particularly useful for supporting early drug development decisions because of its practicality and transparency, although it must be used with care when omitted factors (e.g., probability of development success, development time and cost, or market factors) differ across alternatives being compared. In late-stage development, financial measures gain more importance, as these factors become quantifiable in the context of reduced uncertainty. In either approach, drug–disease models based on available data can support a probabilistic description of the product profile, so that, as development progresses, updates can flow through to the metrics, generally tightening their distributions as uncertainties are resolved. At the end of development, even regulatory approval processes could be made more transparent and structured with such tools, as has been argued for benefit–risk analysis more generally.

Finally, combinations of the approaches, e.g., linking CUI to a market model or pricing out utility in terms of quality-adjusted life years, have potential to better support clinical development.

SUPPLEMENTARY MATERIAL is linked to the online version of the paper at http://www.nature.com/cpt

ACKNOWLEDGMENT
We acknowledge the input of Itay Perlstein, Joanna Peng, and Jeremy Hing (Merck & Co., Inc.).

CONFLICT OF INTEREST
B.P., F.L.H., and K.D. are employees of Pharsight Corporation and are paid consultants to Merck & Co., Inc. A.K., J.A.W., and R.K. are employees of Merck & Co., Inc., and may hold stock or stock options in the company. R.T.C. is employed by Duke University and consults with Pharsight Corporation. This work is sponsored by Merck & Co., Inc.

© 2009 American Society for Clinical Pharmacology and Therapeutics