Three Sets of Case Studies Suggest Logic and Consistency Challenges with Value Frameworks

Joshua T. Cohen, PhD*, Jordan E. Anderson, BA, Peter J. Neumann, ScD
Center for the Evaluation of Value and Risk in Health, Tufts Medical Center, Boston, MA, USA

ABSTRACT

Objective: To assess the logic and consistency of three prominent value frameworks. Methods: We reviewed the value frameworks from three organizations: the Memorial Sloan Kettering Cancer Center (DrugAbacus), the American Society of Clinical Oncologists, and the Institute for Clinical and Economic Review. For each framework, we developed case studies to explore the degree to which the frameworks have face validity in the sense that they are consistent with four important principles: value should be proportional to a therapy’s benefit; components of value should matter to framework users (patients and payers); attribute weights should reflect user preferences; and value estimates used to inform therapy prices should reflect per-person benefit. Results: All three frameworks can aid decision making by elucidating factors not explicitly addressed by conventional evaluation techniques (in particular, cost-effectiveness analyses). Our case studies identified four challenges: 1) value is not always proportional to benefit; 2) value reflects factors that may not be relevant to framework users (patients or payers); 3) attribute weights do not necessarily reflect user preferences or relate to value in ways that are transparent; and 4) value does not reflect per-person benefit. Conclusions: Although the value frameworks we reviewed capture value in a way that is important to various audiences, they are not always logical or consistent. Because these frameworks may have a growing influence on therapy access, it is imperative that analytic challenges be further explored. Keywords: cost-effectiveness analysis, healthcare costs, oncology treatments, value frameworks.

Introduction

Any algorithm used to evaluate the value of medical therapies has limitations. A single number or set of numbers cannot capture all information pertaining to the myriad benefits, risks, and costs of a wide range of treatments.

Nevertheless, algorithms used in “value frameworks,” such as those being developed and promulgated by a range of professional societies and other organizations in the United States, can be designed to better reflect outcomes of interest to stakeholders, and to account for the preferences of the patients and other agents such as payers. This study uses concrete examples—case studies—to explore the extent to which three well-known value frameworks achieve these goals.

These and other frameworks are a response to the important needs of payers, clinicians, and patients to systematically evaluate and in some cases compare therapies. The frameworks seek to expand upon traditional evaluation methodologies (e.g., cost per quality-adjusted life-year [QALY] ratios) by more explicitly accounting for the preferences of framework “users,” and by reporting results in a user-friendly manner. Also, it bears emphasizing that developing value frameworks is challenging. It is easier to criticize frameworks than to construct them. Nonetheless, it is important to explore whether notable frameworks have face validity, that is, do they align with an externally logical and credible characterization of value? To address this question, we constructed a series of case studies.

Methods

Value Frameworks

We focused on three value frameworks: DrugAbacus (from the Memorial Sloan Kettering Cancer Center), the American Society of Clinical Oncologists (ASCO) framework, and the Institute for Clinical and Economic Review (ICER) framework.

DrugAbacus [1] aims to “determine appropriate prices for cancer drugs based on what experts tend to list as possible components of a drug’s value.” It derives an “appropriate” price for a drug on the basis of its incremental survival benefit and the value of each added survival year, as designated by the user (Fig. 1). The algorithm underlying DrugAbacus (as of August 2016) then scales this price by a series of factors, each of which reflects a characteristic, as specified by the framework’s authors (e.g., the drug’s toxicity, novelty, and cost of development) and a “preference weight” selected by the user. Each factor’s preference weight represents that factor’s maximum impact on price. For example,
DrugAbacus price = Added survival * Value of each year * Adjustment factors

DrugAbacus Assumption | User-specified value | Depends on user preference and DrugAbacus assumption

<table>
<thead>
<tr>
<th>Adjustment factor</th>
<th>User preferences</th>
<th>Preference scale</th>
<th>DrugAbacus assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toxicity Discount</td>
<td>User-specified max discount</td>
<td>30%</td>
<td>DrugAbacus toxicity assumption: Based on prevalence of Grade 3 and 4 toxicity effects and on the probability that patient discontinues treatment because of toxicity.</td>
</tr>
<tr>
<td>Cost of Development</td>
<td>User-specified max premium for high development costs</td>
<td>3x</td>
<td>DrugAbacus development cost assumption: Based on number of subjects in drug’s FDA approval trials: “Development cost is low”: No premium. “Development cost is high”: Full premium.</td>
</tr>
<tr>
<td>Rarity</td>
<td>User-specified max premium for drugs treating &quot;rare&quot; conditions</td>
<td>3x</td>
<td>DrugAbacus rarity assumption: Based on incidence of treated disease: “Treated disease is not rare”: No premium. “Treated disease is very rare”: Full premium.</td>
</tr>
<tr>
<td>Population Health Burden</td>
<td>User-specified max premium for drugs treating “high burden conditions”</td>
<td>3x</td>
<td>DrugAbacus population burden assumption: Based on life years (lost in population to the treated disease: “Treated disease is low burden”: No premium. “Treated disease is high burden”: Full premium.</td>
</tr>
<tr>
<td>Unmet Need</td>
<td>User-specified premium for drugs treating conditions not addressed by other drugs</td>
<td>3x</td>
<td>DrugAbacus unmet need assumption: Based on how many treatments are recommended for drug’s target condition by the National Comprehensive Care Network: Minimal unmet need (many alternatives): No premium. High unmet need (few alternatives): Full premium.</td>
</tr>
<tr>
<td>Prognosis</td>
<td>User-specified premium for drugs treating severe conditions</td>
<td>3x</td>
<td>DrugAbacus prognosis assumptions: Based on median survival in the absence of therapy (from FDA label): Low severity (long median survival in the absence of therapy): No premium. High severity (short median survival in the absence of therapy): Full premium.</td>
</tr>
</tbody>
</table>

**Fig. 1** – DrugAbacus framework. FDA, Food and Drug Administration.

A “novelty” preference weight of 2.5 means that DrugAbacus inflates the price of the most novel drugs (drugs with a novel mechanism of action) by a factor of 2.5. DrugAbacus does not scale the price of drugs with minimal or no novelty (i.e., next-in-class drug prices are multiplied by 1); drugs with intermediate novelty (known target but novel delivery) are scaled by an intermediate value—in this case, the average of 1 and 2.5, or a factor of 1.75.

The ASCO framework has a stated goal of aiding patient and clinician shared decision making. Its June 2016 revised framework [2,3] characterizes drug value in terms of points awarded on the basis of a drug’s clinical benefits, toxicity, and “bonus” considerations (see Fig. 2). How it awards points depends on the type of data available. Clinical benefit points reflect overall survival (OS) if that information is available, progression-free survival (PFS) as a second choice, and response rate otherwise. Toxicity points correspond to the new drug’s potential harms, taking into account the number of distinct toxic symptoms, their severity, and prevalence. Finally, the bonus category adds points for increased long-term (“tail of the curve”) survival, improved cancer-related symptoms, improved quality of life, and extended time between treatments.

The ICER framework [4] envisions payers as a key audience. It assesses value using a two-part approach. First, the ICER identifies a “care value” benchmark representing the highest price consistent with a cost-effectiveness ratio less than (more favorable than) a cost-per-QALY threshold of either $100,000 or $150,000 per QALY. Which threshold it uses (the smaller, more stringent value or the higher, less stringent value) depends on a qualitative assessment of contextual factors, such as the supporting comparative effectiveness evidence, the drug’s other benefits and disadvantages (e.g., adherence levels), and other considerations (e.g., ethical, legal, or other issues). Second, the ICER identifies a “health system value” benchmark consistent with “sustainable” health care budget growth. In practice, the ICER identifies the highest drug price consistent with aggregate annual population spending of no more than $904 million. (The ICER explains that the $904 million spending limit per new drug is consistent with the annual drug spending growth of no more than 1% higher than a historical annual gross domestic product growth rate of 2.75%, and that this rate corresponds to societal willingness to pay for Medicare spending growth, as expressed in the Accountable Care Act [5].)
Data

Because this study sought to explore the properties of the frameworks in actual practice, we developed case studies to illustrate each framework’s salient features. The degree to which the case studies mirror the experience of real-world therapies is not entirely clear, although we believe that the issues we highlight help flesh out the important potential consequences of using frameworks given their “inner workings.” Here, we explain the origin of our case studies and how we identified or computed framework-derived values for each therapy considered.

DrugAbacus recommends prices for a predesignated set of 52 cancer drugs approved by the Food and Drug Administration (FDA) between 2001 and 2015 [1]. Our review identifies a subset of these drugs to illustrate the DrugAbacus properties. The DrugAbacus Website provides a “recommended” and actual price for each of the 52 medications. Setting the toxicity discount to 0 and all other preference weights to 1.0 effectively “neutralized” their impact on the DrugAbacus price, making it possible to back-calculate the DrugAbacus-assumed number of life-years gained by each drug. We then explored how user-designated preferences can influence the DrugAbacus-recommended prices, focusing on a subset of 8 of the 52 drugs ranked in the top 100 by US sales [6].

The ASCO framework is also designed to evaluate cancer treatments. Although it is not restricted to a predesignated list, we considered a subset of the same 52 FDA-approved drugs that DrugAbacus addresses, plus Zevalin (ibritumomab tiuxetan), which the FDA approved for the treatment of non-Hodgkin lymphoma in 2002. We characterized survival, toxicity, and other clinical characteristics using results reported in each drug’s FDA label. Although other data could be used, these results serve as a plausible basis for exploring the ASCO framework’s properties.

The ICER framework is not restricted to cancer treatments or pharmaceuticals. To explore its properties, we reviewed seven therapies to which the current version of the framework has been applied. We did not include ICER’s evaluation of acute myeloid leukemia therapies because that assessment examined five therapy combinations, which could have thus complicated the comparison of this example with the other seven assessments. For each therapy, we back-calculated savings \( \Delta S \) and an incremental QALY gain \( \Delta QALY \) from the equation

\[
P_t = V_t \times \frac{QALY}{\Delta} + \Delta S,\]

where \( V \) is the assumed value of a QALY \((V_1 = \$100,000 \text{ and } V_2 = \$150,000)\), and \( P_t (P_1 \text{ and } P_2) \) represents the prices for the therapy warranted by (i.e., consistent with) those QALY values. Because there are two equations—one for each QALY value—we could solve for these two unknowns. Note that the proper way would be to use incremental price (relative to the

---

**Fig. 2 – The ASCO framework. ASCO, American Society of Clinical Oncologists. *This description is not complete. Please refer to (Schnipper et al).**
therapy being replaced), but because we do not readily have access to these values, we use absolute price as a proxy. This assumption is reasonable if the price of the therapy being replaced is small relative to the price of the drug being evaluated.

**Value Criteria**

We define a therapy’s value as the per-person price corresponding to the therapy’s average per-person benefits. We evaluate each framework in terms of its adherence to four principles, which we posit reflect logical, reasonable, and credible tenets.

First, value should be proportional to benefit, all else being equal. For example, extending life by 8 months has twice the value of extending life by 4 months. Proportionality is consistent with both cost-effectiveness and cost-benefit analyses, methods widely used to assess the value of health-promoting measures. This principle is also consistent with the idea that “a life-year is a life-year,” regardless of how it is accrued or who receives it. Hence, two therapies that confer the same survival duration gain should have the same value.

Second, value should reflect attributes that matter to users (patients or payers), that is, value should reflect elements patients or payers would in principle be willing to pay for. For example, survival gains, quality of life, health care cost offsets, and productivity arguably satisfy this criterion. Nevertheless, a drug’s research and development (R&D) costs, for example, do not. Higher research costs facilitate discovery of treatments with these qualities, and may be something that the government appropriatively subsidizes (e.g., because of positive externalities), but, per se, higher R&D costs are not part of value.

Third, the relative weight placed on therapy attributes should reflect user preferences, and should do so transparently. Specifically, in cases in which frameworks allow users to designate weights quantifying their preferences, the relationship between the weights, the attributes, and the value should be transparent. For example, if a user can place a weight on survival, it must be clear how that weight will influence estimated value in the context of a specific survival gain. Otherwise, user weight assignments will not be meaningful.

Finally, when used to inform per-person prices for a drug or therapy, value should reflect per-person benefit. That is, treatment value for one person should correspond to the benefit accrued when one person is treated; it should not depend on the size of the treatable population.

**Results**

We describe the most salient instances in which each framework was inconsistent with a principle.

**DrugAbacus**

Table 1 lists the medications and associated data used in our DrugAbacus case studies.

Value attributes are not always patient-centered and/or do not always reflect per-person benefit: A distinguishing feature of the DrugAbacus framework is its inclusion of adjustment factors that depend on user-assigned preference weights. Some of these factors are indeed patient-centered (e.g., the toxicity discount factor), but four are not (although users may set values to 1 indicating that these factors are not important) including the 1) developmental cost premium, 2) rarity premium (not pertinent to per-person value), 3) population burden premium (not pertinent to per-person value), and 4) novelty premium (although it is potentially a proxy for incremental benefit, the unmet need premium already addresses this attribute).

**ASCO Framework**

The ASCO framework quantifies drug value in terms of points awarded for clinical benefits, toxicity characteristics, and other bonus categories. Table 2 presents information used in the ASCO framework for seven cancer drugs. Case studies compare awarded points across drugs, illustrating three limitations.

Value is not proportional to benefit (1): For three drugs—Abraxane, Kadryla, and Ixempra—we consider the clinical benefit points awarded and compare those point scores with the corresponding survival benefits. Recall from Figure 2 that the ASCO framework awards clinical benefit points on the basis of OS if that information is available, as it is for these three drugs.

The OS hazard ratio (HR) for Abraxane is 0.72, implying that Abraxane’s hazard rate is 28% less than its comparator’s. That reduction yields a score of $100 \times 28\% = 28$ points. Kadryla’s clinical benefit score is 32 points, whereas Ixempra’s is 10 points.

Now consider how those point scores compare with the median OS for these three drugs. Abraxane and Kadryla have similar clinical benefit point scores (28 and 32 points, respectively) that differ by approximately 15%, but the incremental median OS for these two drugs (at 1.8 and 5.8 months, respectively) differ by a factor of 3. In contrast, Abraxane and Ixempra have very different clinical benefit point scores (28 and 10 points, respectively) but the same incremental median OS (1.8 months).

Value is not proportional to benefit (2): Recall from Figure 2 that when drugs lack OS data, the ASCO framework uses PFS information to compute the clinical benefit score. Now consider Gilotrif and Velcade. Velcade lacks OS information for multiple myeloma and mantle cell lymphoma, and so the ASCO
The framework uses its PFS HR of 0.61 (hazard reduction of 39%) to compute a clinical benefit score of 80 (39% × 0.61 = 31 points).

If Gilotrif had only PFS information, its similar PFS HR of 0.58 (42% reduction) would yield a similar clinical benefit score of 34 points. The median PFS extension for these two drugs is likewise similar (4.2 months for Gilotrif vs. 4.3 months for Velcade). But in addition to the PFS data, Gilotrif has OS data, which indicate an OS HR of 0.91. Because the ASCO framework ignores PFS data, it can compute an OS HR; its score for Gilotrif is based on only its OS HR (100 × 91% = 91 points). In short, Gilotrif’s clinical benefit score (91 points) is much lower than Velcade’s (31 points) even though both confer similar PFS benefits, but only Gilotrif has a documented (small) OS benefit.

Attribute weights do not necessarily reflect user preferences: Consider Kadcyla, Provenge, and Velcade. The ASCO framework awards 32 points to Kadcyla because it multiplies Kadcyla’s 32% reduction in the mortality hazard by 100. It awards 31 points to Velcade because it multiplies Velcade’s 39% reduction in the PFS hazard by 80. It penalizes Provenge with a loss of 6 points because it multiplies its 28% increase in weighted toxicity by 20. Although each of these score components can ostensibly be compared (they are all expressed in “points”), there is no basis for their comparison. There is no reason why a 32% reduction in mortality hazard (32 points) has approximately the same value as a 39% reduction in PFS (31 points). Nor is there a reason why these changes each have roughly 5 times the value of Provenge’s 28% reduction in the ASCO toxicity score.

We recognize that our ASCO case studies compare drugs with different applications and that users might more typically compare point totals for drugs within the same indication. The larger idea, however, is that the points awarded by the ASCO framework, which convey a drug’s value, are problematic along the lines discussed.

### ICER Framework

Table 3 lists seven therapies evaluated to date by the ICER under its present framework (as of August 2016). For each therapy, the

<table>
<thead>
<tr>
<th>Drug</th>
<th>Unadjusted price per month ($)</th>
<th>Developmental cost premium</th>
<th>Rarity premium</th>
<th>Population burden premium</th>
<th>Novelty premium</th>
<th>Adjusted price per month ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treanda</td>
<td>9,025</td>
<td>1.3</td>
<td>2.6</td>
<td>1.8</td>
<td>1.0</td>
<td>54,919</td>
</tr>
<tr>
<td>Erbitux</td>
<td>8,063</td>
<td>1.6</td>
<td>1.2</td>
<td>2.6</td>
<td>2.0</td>
<td>80,594</td>
</tr>
<tr>
<td>Spryce</td>
<td>4,721</td>
<td>1.4</td>
<td>2.8</td>
<td>1.8</td>
<td>1.0</td>
<td>33,343</td>
</tr>
<tr>
<td>Velcade</td>
<td>3,613</td>
<td>1.2</td>
<td>2.4</td>
<td>1.4</td>
<td>3.0</td>
<td>43,691</td>
</tr>
<tr>
<td>Tarceva</td>
<td>3,488</td>
<td>1.7</td>
<td>1.2</td>
<td>3.0</td>
<td>1.0</td>
<td>21,336</td>
</tr>
<tr>
<td>Alimta</td>
<td>2,713</td>
<td>1.4</td>
<td>3.0</td>
<td>1.2</td>
<td>1.0</td>
<td>13,655</td>
</tr>
<tr>
<td>Gleevec</td>
<td>2,708</td>
<td>2.0</td>
<td>2.8</td>
<td>1.8</td>
<td>3.0</td>
<td>81,900</td>
</tr>
<tr>
<td>Afinitor</td>
<td>2,571</td>
<td>1.4</td>
<td>1.6</td>
<td>1.4</td>
<td>1.0</td>
<td>8,064</td>
</tr>
<tr>
<td>Avastin</td>
<td>2,475</td>
<td>1.8</td>
<td>1.2</td>
<td>2.6</td>
<td>3.0</td>
<td>41,669</td>
</tr>
</tbody>
</table>

* We use the brand names for the DrugAbacus example for consistency; the online DrugAbacus tool lists the drugs alphabetically by brand name.

† This represents the unadjusted price of the drug per month that corresponds to the value of its conferred survival gain, valuing 1 life-year at $50,000.

‡ Same as the unadjusted price†, but also taking into account the developmental cost premium, rarity premium, population burden premium, and novelty premium, assuming that the user places a weight of 3.0 on each of those factors.
The ICER framework recommends a price low enough to satisfy two conditions: 1) the cost-effectiveness ratio cannot exceed either $100,000 or $150,000 per life-year saved (prices in the second and third columns from the left) and 2) the annual budget impact cannot exceed $904 million for drugs, or $603 million for devices (prices in the last column).

Value is not proportional to benefit: The ICER’s cost-effectiveness criterion ensures that value (the cost-effectiveness-driven price) is proportional to incremental survival, after accounting for savings (i.e., cost offsets) or additional treatment-associated costs. Consider PCSK9 inhibitors. The ICER estimated that for each treated patient, this medication achieves annual medical cost savings of $742 (excluding the cost of the drug itself), and that if a QALY is worth $100,000, the appropriate price for this medication is $5,404. That implies that the annual value of the health benefits is $5404 – $742 = $4,662. This result is consistent with the quality-adjusted survival benefit (0.047 years) and a value of $100,000 for each added year of quality-adjusted survival (because 0.047 QALY × $100,000 per QALY = $4,700).

The problem is that the ICER framework’s budget constraint requirement can break that proportional relationship. Again, consider PCSK9 inhibitors. As indicated in the first row of Table 3, the budget constraint imposes a more stringent (lower) price on PCSK9 inhibitors than cost-effectiveness considerations—$2,177 rather than $5,404. Subtracting the savings of $742 implies that the clinical benefit (0.047 QALY) is worth $1,435, or that each QALY is worth $30,800. Using the budget constraint-driven prices for Ernesto and the CardioMEMS system yields QALY values of $46,800 and $34,200, respectively.

### Discussion

Each framework we reviewed captures an aspect of value that is important to various audiences.

DrugAbacus allows users to build preferences for different drug attributes into the tool’s value calculation. The ASCO framework focuses on attributes of cancer treatment generally regarded to be most salient (survival, toxicity, etc.). The ICER framework accounts for cost-effectiveness and budget impacts. As we have argued previously [7], these and other value frameworks should be welcomed in a health system sorely in need of approaches to achieve more health for the spending. Moreover, many aspects of the frameworks we reviewed work well and even the problems we identified do not necessarily apply to all therapies.

We intend our case studies to be illustrative. We did not compile an exhaustive list of all possibly desirable principles, nor does absence of a critique imply that a framework is sensible on all other grounds. Nevertheless, as our case studies underscore, the frameworks may not always capture value in ways that seem logical or are consistent with user preferences. DrugAbacus includes attributes that may be important, but are not properly termed “value” attributes—or do not inform value estimates pertinent to a therapy’s per-person price. We recognize that these attributes may be important to society. For example, drugs treating rare conditions may warrant government subsidies. R&D costs may deserve government support because market failures create positive externalities that lead R&D to be underpriced from a societal perspective. Our main point is that these other attributes do not pertain to value, per se, and should not be used to establish “value-based” therapy prices.

In some cases, the DrugAbacus attributes lack transparency. DrugAbacus could be improved by including in the value estimation equation only those attributes related to clinical outcomes valued by patients or payers. For each drug, it should also explicitly report its drug attribute assumptions (efficacy years of added survival, the novelty score, etc.) and their basis so that users can understand their meaning and how preference weights relate to value estimates.

Our case studies highlight two important limitations of the ASCO framework. First, the ASCO framework awards clinical benefit points on the basis of relative improvements in either survival or PFS. Declining marginal utility could arguably justify this scheme (because incremental survival gains are worth more when baseline survival is shorter). But the approach is at odds with well-accepted cost-effectiveness and cost-benefit methods (which award value on the basis of absolute gains in proportion to the clinical benefit’s magnitude). Moreover, even if value should reflect declining marginal utility, the ASCO framework has not justified its particular quantitative formulation or based it on empirical evidence. This limitation can readily be addressed by changing its point scoring formula. For example, points could
be awarded for each incremental month of survival. Alternatively, scoring could reflect baseline survival (presumably awarding more points in cases in which baseline survival is the shortest). But scoring should not necessarily be tied to the idea that value is proportional to percent survival gain, as is the case in the present version of the ASCO framework.

Second, the point scores the ASCO framework awards for various benefits (extended survival, reduced toxicity, etc.) are arbitrarily derived, on the basis of expert opinion rather than empirical estimates of user preferences. ASCO should consider conducting user preference elicitation studies to measure the trade-offs stakeholders are willing to make among the benefits incorporated into the ASCO framework.

Finally, the budget constraint imposed by the ICER framework may be relevant to payers, but it is not a legitimate attribute of value. In our view, the ICER should continue to report budget impact, but its value-based price should not reflect that impact.

Source of financial support: This study was supported by the National Pharmaceutical Council and the Pharmaceutical Researchers and Manufacturers of America. The publication of the study results was not contingent on the sponsors’ approval or censorship of the manuscript.

REFERENCES