A Framework to Guide Payer Assessment of the Value of Medical Treatments

Steven D. Pearson, M.D., M.Sc.

The Institute for Clinical and Economic Review
A Framework to Guide Payer Assessment of the Value of Medical Treatments

Introduction

This paper presents a conceptual “value framework” as a guide for payers and other organizations in the United States involved in the assessment of the value of medical treatments, including drugs, medical devices, and procedures. The overall goal of the framework is to improve the reliability and transparency of payer value assessments, an outcome that will improve the dialogue on innovation and value between payers and manufacturers while also enhancing the legitimacy of coverage and reimbursement decisions with patients, clinicians, and the public.

To serve this goal the framework identifies the most important factors involved in a value assessment and groups them into five conceptual elements. The framework reviews the options for measuring and integrating each element of value into an overall assessment, guided by the insights of a multi-stakeholder workgroup that includes patient advocates, life science companies, pharmaceutical benefit managers, employers, and insurers. Benefitting from this input, the design of the value framework has been fully informed by the realities of the current US policy landscape. But the framework seeks to do more than just describe current assessment strategies: it presents a new, structured approach that can help payers and other policy makers improve the reliability, transparency, and legitimacy of value assessments across the health care system.

Elements in a Payer Assessment of Value

As shown on Figure 1 on the following page, this approach begins by grouping the factors involved in a payer value assessment into one of five conceptual elements: 1) Comparative clinical effectiveness; 2) Additional benefits for patients, family, delivery system, and society; 3) Contextual considerations related to the illness and treatment; 4) Comparative costs; and 5) Contextual considerations related to costs and delivery system or market factors.
Figure 1. Five elements in a payer assessment of value. Contextual considerations related to the illness and therapy have a broad impact on the judgments involved in all other elements of a value assessment.

Value Element 1.
Comparative Clinical Effectiveness: Net Health Benefits and Level of Certainty

Recommendations:

1. **Payers should adopt an explicit, transparent method for determining the comparative clinical effectiveness of a treatment.** This determination should include two components: a judgment of the comparative net health benefit of a treatment; and the level of certainty in this judgment that can be provided by the available evidence. The online ICER Evidence Matrix provides a template for this purpose and is freely available to all payer organizations.

2. **At the beginning of any evidence review on comparative clinical effectiveness, payers should use an explicit template for identifying the key health outcomes, including risks and benefits, that will be considered in the determination of comparative net health benefit.**
3. **Payers should use an explicit, transparent approach to judging the level of certainty they have in the magnitude of the comparative net health benefit of a treatment.**

4. **Payers should combine their judgments of comparative net health benefit and level of certainty into a single overall judgment of comparative clinical effectiveness using a categorical approach such as that provided by the ICER Evidence Matrix. This single rating can serve as the guiding step in an overall value assessment process.**

The first element of value assessment, comparative clinical effectiveness, is the cornerstone of the overall process. Payers perceive the value of a treatment as being related not simply to its “safety and effectiveness” versus a placebo but to its comparative clinical effectiveness versus other available care options. Basic information on safety and effectiveness is obviously important, and this represents the evidentiary standard used by the Food and Drug Administration (FDA), but payers think of value in comparative terms, and useful dialogue among payers and other stakeholders should be anchored in this reality.

Whether done implicitly or through an explicit rating method, payers’ judgment of comparative clinical effectiveness requires the assessment of two interconnected factors: the magnitude of the comparative “net” health benefit of a treatment versus other options; and the level of certainty in this judgment that can be provided by the available evidence. What is meant by the comparative “net” health benefit of a treatment? This involves comparing the overall balance of the health risks and benefits of a treatment with the balance of risks and benefits of one or more alternative treatment options.

Here it is important to note how critical it is for there to be a clear definition of the relevant health outcomes – including both risks and benefits -- that will be considered by the payer as part of the judgment of comparative net health benefit. The definition of the outcomes to be considered should occur during the earliest phase of an evidence assessment, often as part of determining the entire scope of the review using the “PICO” framework: PICO is an acronym for the target Population, the Intervention(s) of interest, the Comparator intervention(s), and the key Outcomes.

Payers will usually define the key outcomes of any assessment of comparative net health benefit as “patient-centered” outcomes with meaningful clinical significance for patients, such as overall or disease-specific mortality, quality of life, and functional outcomes. Quite often, however, clinical trials use short-term or surrogate outcomes that are easier to measure but which may lack evidence of their validity for representing or predicting true patient-centered outcomes. Payer assessments of comparative clinical effectiveness are often strongly influenced by the degree to which the evidence base includes what payers believe are clinically relevant outcomes, and therefore in value assessment documents payers should be as explicit as possible in describing and
justifying their selection of the outcomes they will use to judge the evidence on comparative net health benefit.

Payers also need to decide and make clear whether there are any potential risks or benefits that may not be fully captured in the clinical trial evidence but that will be included as part of the assessment of comparative clinical effectiveness. The value framework includes potential “additional benefits” beyond the health benefits considered part of comparative clinical effectiveness as a second, separate element of value assessment, but here it should be noted that some payers may decide to include certain benefits such as convenience, different routes of administration, etc. as part of their assessment of comparative clinical effectiveness, particularly if there is evidence linking these benefits back to improvements in clinical outcomes. Conversely, payers may decide that issues such as these will be considered separately from the core assessment of comparative net health benefit. Being explicit about which types of risks and benefits will be considered as part of an overall value assessment, and about when and how they will be considered, should be an explicit part of a payer’s approach to value assessment.

Once the key outcomes to be included in the evaluation of comparative net health benefit have been delineated, and an evidence review performed to gather the data from the published medical literature, some process must be used to synthesize the various risks and benefits of a treatment and its comparator(s) in order to compare them. There are different ways this can be done. In the rare occasion when a treatment and its comparator(s) have the exact same clinical risks and benefits, and the only difference is the rate of these outcomes, then a quantitative comparison is relatively straightforward. However, competing treatments usually have not only differing rates of the same risks and benefits but different types of risks and benefits. For example, one effective drug for rheumatoid arthritis might cause fatigue and present a risk of serious infection, whereas its comparator is somewhat less effective at improving joint symptoms, causes less fatigue, does not pose an infection risk, but does carry a small risk of serious liver disease. How can the overall comparative “net” health benefit of these two drugs be compared?

Some payer organizations will choose to leave risks and benefits disaggregated throughout the evaluation and compare the differences on each between a treatment and its comparator(s) separately. Separate comparison of each risk and benefit is useful given that some outcomes will be considered to be more important than others, and examining the data on all risks and benefits separately allows for payers to provide a more transparent explanation of how they weight various differences, even when this weighting remains qualitative instead of quantitative.

But payers can also consider explicitly combining data on risks and benefits into summary measures to aid in determining comparative net health benefit. By far the
most widely used aggregate measure of health outcomes is the quality-adjusted life year (QALY), a measure of health gain that applies patient preferences, or “utilities,” for different health states to combine the impact of all risks and benefits into a single number on a continuous scale. The QALY gained from a treatment represents the length of life gained, if any, compared to some other option, adjusted for the relative impact of the difference in quality of life produced by the two treatment’s different side effect profiles.

Whereas many international payer agencies have adopted the QALY as a universal metric of health outcomes by which to analyze comparative net health benefit across different types of medical interventions, very few payers in the US use the QALY in a systematic way. In part this is because of methodological concerns about whether the QALY adequately reflects the preferences of patients for different types of health outcomes. There are long-standing concerns that QALYs fail to capture important societal values favoring health benefits for patients with the most severe illnesses. And QALYs usually must be estimated from published literature through analyses that can be complex, time consuming, and ultimately lacking in the degree of transparency that is one of the most important goals of a value framework. The methodological concerns are most relevant when QALYs are used as part of analyses comparing the incremental cost-effectiveness of treatments for different conditions. When used as a summary measure of health outcomes to compare treatments for the same condition, there are far fewer methodological concerns, and there exists no superior overall measure of comparative net health benefit.

When all risks and benefits are left disaggregated, as is usually the case, some method of defining the overall comparative net health benefit is needed. And even if risks and benefits are combined and compared using the QALY, the relative importance of the magnitude of the difference in QALYs must be judged. The value framework therefore recommends that all payers adopt an explicit categorical approach to these judgments of comparative net health benefit in order to improve the consistency and transparency of the value assessment process. Categorical approaches have been used for many years by many international agencies, including HAS in France and IQWiG in Germany, where the comparative net health benefit of a treatment is analyzed and then assigned to one of several categories of “added clinical benefit.” In Germany, for example, a treatment’s added clinical benefit is categorized as either “major,” “considerable,” “minor,” or “none,” with pricing protocols linked to each category.

In the US the most prominent categorical rating system is the ICER Evidence Matrix, a tool that is now being promulgated under the auspices of the CER Collaborative (cercollaborative.org), an initiative spearheaded by the Academy of Managed Care Pharmacy (AMCP), the International Society for Pharmacoeconomics and Outcomes Research (ISPOR), and the National Pharmaceutical Council (NPC). The ICER Evidence Matrix evidence categorizes comparative net health benefit as “superior,”
“incremental,” “comparable,” and “inferior.” The online version of the ICER Evidence Matrix includes a user’s guide that can help payers use the matrix to achieve greater rigor, consistency, and transparency in making categorical judgments of comparative net health benefit.

Whatever methods are used to analyze and summarize the magnitude of comparative net health benefit, there is a second, inter-related judgment that is integral to the overall process of evaluating comparative clinical effectiveness: a judgment of the level of certainty that the evidence is able to provide. It is often possible for the existing evidence to suggest that a treatment has a “superior” or an “incremental” comparative net health benefit, but if the level of certainty in this judgment is low due to deficiencies in the evidence, then the overall judgment of comparative clinical effectiveness will suffer accordingly, as will the ultimate assessment of treatment value.

How do payers assess the level of certainty they can have in their estimate of the comparative net health benefit of a treatment? Many payers have adopted a version of the methods for judging evidence developed by the United States Preventive Services Task Force (USPSTF), the Agency for Healthcare Research and Quality (AHRQ), and the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group. These methods are also incorporated into the ICER Evidence Matrix. While each organization has some minor differences in its approach and nomenclature for rating the strength of evidence, there is a fundamental similarity: the body of evidence should be evaluated, in its entirety, along a series of criteria or “domains.” The most important domains to be considered are virtually identical across all methods:

- **Bias:**
  - The risk of bias from the study designs within the body of evidence
- **Applicability:**
  - The generalizability of the patients and clinicians to “real-world” populations (e.g., differences in patient populations between traditional and pragmatic clinical trial designs, inclusion of real-world database studies in body of evidence)
- **Consistency:**
  - The degree to which different studies have similar key findings
- **Directness:**
  - The relative directness vs. indirectness of the comparison of drug vs. comparator possible from the available evidence; and/or
  - The relative directness with which measured outcomes in the studies, e.g. surrogate outcomes, reflect true patient-centered outcomes
- **Precision:**
  - How precise the results on key outcomes are in available evidence
When the evidence base for a treatment scores well on these domains, payers will have a higher level of certainty in whatever their judgment is of the treatment’s comparative net health benefit. The ICER Matrix guides the user to examine these domains and then categorize certainty into one of three levels: “High”, “Moderate”, and “Low”. As with judgments regarding the boundaries of the categories for comparative net health benefit, there are no exact quantitative methods for defining the boundaries between these categories of certainty. However, there are typical characteristics of bodies of evidence that are representative of each level of certainty, and the online user’s guide provides benchmarks and other information to help payers achieve consistency in the rating of level of certainty within their organization.

When a payer has made both judgments -- of the estimate of comparative net health benefit, and the level of certainty provided by the available evidence -- it is our recommendation that they be combined into a single rating of comparative clinical effectiveness. The USPSTF, GRADE, and ICER Evidence Matrix methods all follow this approach and arrive at a single rating that combines judgments of net health benefit and certainty. The component judgments remain explicit and visible to all stakeholders. The judgment of comparative net health benefit, with its own internal combination of risks and benefits, can be seen separately within the ICER Evidence Matrix; equally, the rating for level of certainty remains transparent. But combining them in a single rating of comparative clinical effectiveness represents a comprehensive judgment that can serve as a clear first step in an overall value assessment.

Value Element 2.
Additional Benefits for Patient, Family, Delivery System, or Society

Recommendations:

1. Payers should design an explicit template of potential additional benefits of treatment that may not be included in ratings of comparative clinical effectiveness but which are relevant to overall considerations of value.

2. Payers should include in their documentation a description of whether any additional benefits were considered and if so, what judgments were made about their relative importance to an overall rating of value.

Although comparative clinical effectiveness is the first step and remains central to an overall value assessment, the value framework includes a second element that payers can use to capture potential “non-clinical” additional benefits that may be meaningful to patients or other stakeholders but which have not been thought reasonable to include explicitly as part of comparative clinical effectiveness. What kind of benefits might fit
into this category? Some treatments might help patients by allowing them to be treated at home instead of in a health facility, or by allowing a reduction in the need for care from friends or family. Additional benefits might accrue to delivery systems if, for example, the treatment has practical advantages related to its storage or administration. And some treatments might offer additional benefits to society by reducing disparities in care or by reducing barriers to care so that the overall number of individuals who benefit from treatment can be increased. Figure 2 below provides a list of potential additional benefits that could be considered as part of value assessment.

Figure 2. Potential additional benefits of a treatment for patients, family, delivery system or society.

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>Additional Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient, Family, or Society</td>
<td>Benefits of treatment that extend beyond patient-specific health improvement, e.g. reduction in care needed from friends and family, earlier ability to return to work</td>
</tr>
<tr>
<td></td>
<td>Allows expansion of eligible population for treatment</td>
</tr>
<tr>
<td></td>
<td>Removes or reduces barriers to treatment through new route or delivery mechanism</td>
</tr>
<tr>
<td>Providers/Delivery System</td>
<td>Presence of quality target(s) for which the treatment will improve performance</td>
</tr>
<tr>
<td></td>
<td>Other practical advantages related to preparation/storage/delivery of the treatment</td>
</tr>
</tbody>
</table>

Placeholder for description of all the additional benefits listed in Figure 2 above.

Placeholder for discussion of importance of using an explicit template to record whether these additional benefits are relevant and whether any other additional benefits have been considered during the value assessment process. For transparency, payers need to describe not just which additional benefits are considered, if any, but how they were factored into the value assessment.
Value Element 3. Contextual considerations: Illness and therapy

Recommendations:

1. *Payers should use a list of key contextual considerations regarding the illness and therapy to assure that key contextual issues are considered and to document the role of contextual considerations in value assessment.*

Also critical to the transparency of any value assessment is a clear understanding of the role of contextual considerations regarding the illness and the treatment that may have a strong influence over the entire process. This third element of payer value assessment could, in some ways, equally be listed as the first, or the last. As shown in Figure 3 below, it represents the key contextual considerations regarding the illness and therapy that color the interpretation of evidence on all potential risks benefits, whether clinical or other. Perhaps most important among these considerations is whether the treatment is the only viable treatment for a condition or whether there are reasonable alternatives available. The lack of an existing treatment, especially for a serious condition, strongly influences the way that payers assess the evidence on “comparative” clinical effectiveness, increasing the likelihood that a judgment will be made that a new treatment provides a positive net health benefit compared to “usual care.”

Figure 3. Contextual considerations regarding the illness and therapy that are most consequential for payer value assessment.

<table>
<thead>
<tr>
<th>Contextual considerations</th>
<th>Contextual considerations favoring coverage/preferred status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other treatment options</td>
<td>No other acceptable treatments exist</td>
</tr>
<tr>
<td>Heterogeneity of Treatment Effect</td>
<td>New/different mechanism of action in setting of significant heterogeneity of treatment effect</td>
</tr>
<tr>
<td>Professional standards</td>
<td>Consensus among professional statements on appropriate use</td>
</tr>
<tr>
<td>Societal values regarding the illness/condition</td>
<td>High severity and/or priority condition</td>
</tr>
<tr>
<td></td>
<td>Vulnerable population (e.g. children)</td>
</tr>
</tbody>
</table>
Placeholder for description of all contextual considerations in Figure 3.

The list of potential contextual considerations in a payer value assessment can seem almost limitless, so the value framework provides a way to categorize the most important while emphasizing that using a method to document the role of contextual considerations is of paramount importance.

Value Element 4. Comparative costs

Recommendations:

1. In order for payers to assess the value of treatments accurately they should consider total health system costs in their entirety, even if some of these costs fall outside the scope of their own financial interests.

2. Payers should seek some collaborative mechanism to commit to accommodating comparative cost estimates with time horizons of at least 5-10 years.

3. Payers to have a transparent process for documenting whether which costs they are considering, and whether broader societal costs are playing a role in their assessment of value.

Cost is often the first thing that comes to mind when thinking conceptually about the payer approach to value assessment, but as has been noted, the payer perspective on value begins exactly where it does for patients and clinicians: with a judgment of comparative clinical effectiveness. For many US payers, in fact, value assessment is so dominated by the judgment of the first three elements of value -- comparative clinical effectiveness, potential additional benefits, and contextual considerations -- that if comparative clinical effectiveness is judged to be “superior” compared to other available options, then (positive) coverage and reimbursement decisions are often essentially hard-wired even if the comparative costs are substantially higher.

But costs do play an essential role in value assessments, even when the comparative clinical effectiveness is judged to be “superior.” The most important conceptual point regarding cost consideration as an element of value assessment is that, for payers, what matters are the comparative “total” health system costs of a treatment. Costs are always considered in relation to some other course of treatment, be it a competing new agent or just “usual care.” And “total” health system costs matter because payers are ultimately responsible for health system costs beyond just the price, or acquisition cost, of the treatment being evaluated.
A payer value framework, therefore, emphasizes the need payers have for information on health system costs for entire pathways of care associated with a treatment and its comparator(s). In addition to the treatment cost, these other costs include “downstream” costs associated with different care pathways: physician office visits, tests, other treatments, hospitalizations. These broader health system costs will be assessed by payers from their specific financial perspective; that is, what the payer organization itself must pay for all of these services.

In some situations, a payer’s financial structure can present challenges for capturing the true “total” health system costs of a treatment. Acquisition costs of new treatments are likely to hit one budget silo (e.g. the pharmaceutical budget) whereas potential cost savings or additional costs related to the entire pathway of care may accrue to a different part of the organization (e.g. the medical budget), or may fall entirely out of the scope of their budget, as in the case of pharmaceutical benefit management organizations. For some payers this distinction may mean that total health system costs are not recognized or considered in the same way as costs within a single budgetary silo. In order for payers to assess the value of treatments accurately, however, they need to consider total health system costs in their entirety, even if some of these costs fall outside the scope of their own budget. This recommendation can require payers to potentially act against their own financial self-interest, which is never straightforward. But if payers exercise selective consideration of certain kinds of health system costs in the way value assessment of treatments is performed, then the incentives for life science companies to seek true total health system cost control will be vitiated, and patients may be denied preferred access to treatments that serve their longer term interests.

Another issue concerning the scope of costs to be considered is that of the time horizon. When performing value assessments of treatments, payers in the US have usually assessed potential costs and cost-offsets over relatively short 1-3 year time frames. One reason to use a short time frame is that payers often believe that cost projections beyond 1-3 years are difficult to perform and are so unreliable that they cannot be useful as planning tools. The process of health care budgeting within payer organizations also emphasizes short-term results, creating a difficult market environment in which to consider long-term cost implications. Nonetheless, when payers use short time horizons for considerations of costs they undervalue treatments that provide long-term cost offsets, whether those offsets be through reductions in clinician visits, hospitalizations, ancillary test or treatment use. Manufacturers seeking to produce treatments that will be judged as “high value” by payers face incentives to focus on short-term outcomes and may perceive greater risk in bringing to market treatments with substantial short-term costs that are highly likely to produce long-term clinical and economic benefits.
Given the tension between payer incentives to focus on short-term time horizons and the desirable attributes of a longer-term perspective on “value” for patients and other stakeholders, our first recommendation is for greater payer transparency. Payers should disclose the time horizon that they use in considering both clinical and economic outcomes. Second, we recommend that payers adopt strategies that will improve their ability to use longer-range budget impact estimates. To accomplish this, payers could develop the internal capability to design more robust budget impact simulations. Payers could also consider contracting with outside organizations to develop these models and/or evaluate models presented by manufacturers. Through these actions, over time, payers should gain the ability to extend the time horizon for consideration of economic impact when a treatment has the potential to provide cost offsets in future years. Since more certainty about payers’ time horizon for costs will be important to send the correct signals regarding “value” to manufacturers, we also recommend that payers seek some collaborative mechanism to commit now to accommodating comparative cost estimates with time horizons of 5-10 years.

There is another important issue related to the scope of costs that payers consider as part of value assessment: should employee/patient productivity or other societal costs be considered by payers? For example, what if a certain new treatment adds significant costs to the health system but results in more rapid return to work for patients following an injury? And there are even broader considerations possible of societal costs. For example, treatments for opiate dependence may add costs to the pharmaceutical budget of a payer but lead to drastically reduced law enforcement costs and other social safety net spending. Treatments of severe mental illness may increase health system costs but reduce the societal costs of substance abuse, violence, and homelessness. Should payers consider these kinds of cost implications when evaluating the value of medical treatments?

In the United States there is no single national payer assigned the mandate to include in value assessment a formal consideration of the broader societal costs of illness and health care. Yet many payers do, at least in principle, and on occasion in explicit practice. [What would be a good example?] We do not recommend that all payers in the US should routinely attempt to include – quantitatively -- the full societal perspective in value assessment. But it is reasonable for payers to have a transparent process for documenting whether which costs they are considering, and whether broader societal costs are playing a role in their assessment of value. Our recommendation therefore is that payers include an explicit template as part of value assessment that requires them to document how they defined the scope of comparative costs considered in the value assessment. Requirements for transparency may over time lead to broader inclusion of societal costs in payer value assessment, but whatever the outcome, transparency on costs will naturally lead to greater societal dialogue about the degree to which payer organizations should bear responsibility for considering the broader societal costs associated with health care.
Value Element 5. Contextual Considerations: Cost, Delivery System, and Market

Not clear yet whether this set of contextual considerations will remain an element of value or whether they will be described as being outside value assessment but critical for value-based decision-making.

A fifth element in value assessment is the set of contextual considerations related to costs, especially the specific distribution of total health system costs related to a specific treatment within payer organizations and between these organizations and other stakeholders, including patients. For example, it can be particularly relevant to evaluations of total health system cost if costs for treatments are accrued within one budget silo (e.g. pharmacy) while their cost-offsets occur in a different part of the organization’s finances (e.g. hospital costs).

Figure 4. Contextual considerations related to cost and delivery system or market factors.

<table>
<thead>
<tr>
<th>Contextual considerations</th>
<th>Contextual considerations favoring coverage/preferred status</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Special cost issues</strong></td>
<td><strong>Organizational cost considerations favor payer organizational needs</strong></td>
</tr>
<tr>
<td><em>(e.g. who incurs costs/reaps potential cost-offsets inside payer and between payer and providers; who gets rebates; and policy options available for cost-sharing)</em></td>
<td></td>
</tr>
<tr>
<td><strong>Ability to manage utilization</strong></td>
<td><strong>Good ability to identify and manage Rx to subpopulations for which Rx is most cost-effective; low risk of misuse/overuse</strong></td>
</tr>
<tr>
<td><strong>Political considerations</strong></td>
<td><strong>Risk of pushback from patients, clinicians, provider groups if non-covered/non-preferred</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Risk of overturn on appeal or other legal risks given precedents either internal or external (e.g. other payer coverage decisions)</strong></td>
</tr>
<tr>
<td><strong>Special competitive issues</strong></td>
<td><strong>Competitive market factors (e.g. ???)</strong></td>
</tr>
</tbody>
</table>
Putting the Elements Together: Procedures for Rating Value

This section is (particularly) loosely and incompletely drafted given the importance of further feedback on the basic approach that will be sought at the June 25 meeting.

Even if all the elements in a value assessment have been clearly identified, what payers in the US have lacked most is a consistent method by which to integrate these elements in an explicit, transparent rating of the value of treatments.

Cost-effectiveness: cost-utility thresholds, cost consequences, and the efficiency frontier
The first method for rating value that should be considered is incremental cost-effectiveness analysis. Comparing a treatment’s incremental costs with its incremental benefits on continuous scales, which produces an incremental cost-effectiveness ratio (ICER), is the most widely accepted core methodology for measuring the value of any health care intervention. Most discussions of cost-effectiveness analysis as a potential method for US payers focus on whether they should follow the example of some international technology assessment agencies and adopt a cost-effectiveness standard, or threshold, of a dollar figure at which the incremental benefit, expressed in QALYs, of a treatment is judged to be worth the incremental cost. For example, in England, the National Institute for Health and Care Excellence (NICE) has for many years used a cost-effectiveness threshold of approximately £20,000–£30,000 per QALY. This threshold implies that if a treatment provides an additional QALY compared to alternatives at a cost lower than this range it will be judged to represent adequate “value for money” in the English National Health Service.

The current version of the value framework includes cost-effectiveness thresholds of US dollars per additional QALY as one possible component of a multi-criteria approach to rating value. What should the cost per QALY threshold be for payers in the US? The World Health Organization (WHO) has promulgated benchmarks for incremental cost-effectiveness that are pegged to a nation’s Gross Domestic Product (GDP). These cost per additional QALY thresholds are as follows: “low value” is represented by a treatment with a cost per additional QALY greater than 3-times GDP, or approximately $150,000 in the US; “high value” is any treatment with a cost per QALY less than one-times GDP, equal to approximately $50,000 in the US. The theoretical backing for linking thresholds to multiples of GDP is not viewed as strong by many methodologists but no better approach has gained wide acceptance, and the WHO thresholds have been widely adopted by international agencies and even by some US-based payers. [probably need to describe Premera’s tiered formularies here] One US physician specialty society, the American College of Cardiology, has also accepted the WHO
thresholds recently as its sole metric for value within its clinical guidelines program.  [should AMCP dossier use of cost-effectiveness be mentioned here?]

Cost-effectiveness analyses do not always need to depend on QALYs as the metric for health gain. If a treatment and its comparator have a single key outcome measure on which they differ, then it is possible to simplify the cost-effectiveness analysis to compare the cost per key outcome between the two. This approach is known as “cost consequences.” The easiest example of this method is when three is a single outcome that differs between two treatments and the cost per key outcome difference is dramatic. For example, at the time the evidence was being evaluated on the use of Intensity Modulated Radiation Therapy (IMRT) for the treatment of localized prostate cancer, one evidence review concluded that the primary difference in comparative clinical effectiveness of IMRT versus the previous standard was in the rate of diarrhea following treatment due to excessive bowel irradiation. The analysis estimated that the incremental cost for IMRT to prevent one case of mild diarrhea was approximately $300,000. This incremental cost for the marginal clinical benefit was easily judged as not representing a high value.

Cost consequences can be also be used when there is a standard of care treatment to which two competing new treatments are being compared. For example, if there are two new drugs, A and B, to prevent strokes in patients with atrial fibrillation, and these drugs have similar rates of adverse events and only differ in their ability to prevent strokes, then it would be possible to calculate a cost per stroke prevented versus the existing standard of care for both A and B. Drug A would be considered to be more cost-effective than Drug B if its cost per stroke prevented was lower.

If, however, as is often the case, there is just a single new treatment to be compared to the standard of care, and it has an additional cost per additional key outcome, there is no explicit threshold for a cost that would be considered “cost-effective,” or, as NICE terms it, “value for money.” For example, how would a payer determine whether a drug was cost-effective if, compared to the existing standard of care, it prevented one additional stroke at an incremental cost of an additional $100,000? Other than translating this result into cost per QALYs and applying some standard threshold, the only other option is to look retrospectively back to the introduction of the current standard of care and determine what its historical cost per key outcome was versus the predecessor standard of care at the time it was introduced. Once this historical cost per stroke prevented is established, then it can be assumed to represent a “reasonable” or at least an “historically acceptable” cost per stroke prevented, and this figure can be used in comparison with the new drug today to see if the new drug is at or below this standard. In cost-effectiveness methodology using this approach to find a threshold, which can be used with single specific outcomes or with QALYs, is known as the “efficiency frontier.”
Using efficiency frontiers has some intuitive appeal, and it remains the approach used by the German pricing agency to set prices once the category of added clinical benefit is determined. But efficiency frontiers suffer from being rooted in the idiosyncratic history of innovation and pricing in particular clinical areas. It may be, for example, that through the accidents of history we have paid much higher margins for relatively small clinical gains in cardiology than we have in neurology; or for devices compared to drugs. Is it reasonable to accept the historical trajectory of innovation and pricing in different clinical areas or for different types of treatments as a “reasonable” standard?

And so there are advantages and disadvantages to each of the general approaches to using cost-effectiveness: 1) with a GDP-linked threshold for the cost per QALY; 2) comparing the cost per single key clinical outcomes; and 3) with the cost per QALY or cost per a single key clinical outcome compared to an historical efficiency frontier. As part of the value framework we recommend that cost-effectiveness be included as one of the necessary metrics to judge overall value. When differences in single outcomes can be used to compare the cost-effectiveness of multiple competing treatments to a single standard of care, we suggest using a cost consequences approach to determine which treatment is more cost-effective. But when this is not possible, because of its wider acceptance and more grounded benchmarks, we recommend the use of a cost per QALY analysis tied to WHO thresholds over an efficiency frontier methodology. This approach offers the widest generalizability of application across different kinds of treatments and clinical conditions, and is more likely to align with evidence on cost-effectiveness generated by manufacturers, particularly those who work with international evidence review agencies and payers.

We recognize, however, that despite the factors in favor of using incremental cost-effectiveness as the primary, or even the sole, metric for rating value, there are many reasons it is unlikely to serve that function well in the US. First, there is no national infrastructure for performing the analyses necessary to calculate cost per QALYs on all new treatments being considered for coverage.

Further delineation of methodological and political disadvantages of using cost/QALY in the US

The importance of affordability
As central as cost-effectiveness can be as a method for rating value, it does not capture directly a primary payer concern regarding the budget impact of potential new costs associated with treatments. Indeed, for US payers, budget impact, often called “affordability,” has been far more central to judgments regarding overall value than has considerations of cost-effectiveness. A new treatment that offers net comparative benefits at a reasonable incremental cost may meet cost-effectiveness thresholds and be considered “high value” by patients, clinicians, and manufacturers, but if that treatment
will be used by a large number of patients the added cumulative costs may adversely affect the ability of payers to manage overall health care budgets. Substantial new costs raise the risk of either displacing care of greater value within the health care system, or contributing to an unsustainable rise in health insurance premiums. Ultimately, therefore, no treatment, even if it is judged effective and meets cost-effectiveness benchmarks, will be considered to be “high value” by a US payer if it adds substantially to overall health care costs.

The difficulties of justifying a threshold for incremental cost-effectiveness are well established, but it is even more challenging to designate a reliable, generalizable measure by which to determine whether any given budget impact for a payer would represent high value or low value. The draft value framework proposes two potential standards: 1) a qualitative judgment of whether the added costs would cause “more valuable” care to be displaced due to rationing; and 2) a quantitative analysis of whether the added costs would contribute to a rise in overall health care spending by more than the expected annual rise in GDP.

These two potential “standards” need to be discussed further with the policy workgroup before confirming their utility.

As with cost-effectiveness, it would be equally incorrect to presume that affordability should serve as the only measure that US payers should use to assess the value of a treatment. Cost-effectiveness captures an important facet of value, as does affordability; neither can serve alone. And both also need to be viewed within the context of the evidence on comparative clinical effectiveness and all the contextual considerations mentioned above.

There are different options for integrating all these elements as part of value assessment. Consideration should be given to the idea of using formal “multi-criteria decision analysis” (MCDA) as an approach to integrate the four [or five] elements of value assessment into a structured decision-making process. MCDA assigns quantitative weights to whatever criteria are considered relevant for decision-making, and usually sums up these weights in the assessment of a treatment to determine its overall value based on a cumulative value score. MCDA offers advantages in transparency but presents its own set of problems regarding how to weight and combine different criteria. Ultimately, this value framework adopts the intent of MCDA and the conceptual approach of integrating multiple criteria in a value assessment, but it does not use quantitative weights to do so.

We conclude that formal MCDA, with quantitative weights assigned to different criteria, has greater disadvantages than advantages for US payers. Instead, we recommend that payers use an explicit value table as part of a step-wise approach to integrating all the elements to be considered in a value assessment. The value table is
designed to highlight how certain combinations of findings on comparative clinical effectiveness and other elements can lead to an overall rating of “high” or “low value.” The value table seeks to capture all the key elements necessary to a valid process of value assessment, and to provide a model for an approach that will enhance the rigor, reliability, and legitimacy of these determinations with all stakeholders. Starting with a categorical rating for the comparative clinical effectiveness of the treatment, the value table has corresponding criteria for cost-effectiveness and affordability that must both be met in order to achieve an overall value assessment rating of “high value.” The relative role of contextual considerations is indicated as well, but we do not recommend trying to quantify contextual considerations as a means to weighting either the comparative clinical effectiveness rating and/or element within the cost-effectiveness analysis.
Figure XX. Mock value table demonstrating use of *hypothetical* criteria in sequence to assign overall value ratings to treatments.

<table>
<thead>
<tr>
<th>Comparative Clinical Effectiveness</th>
<th>Additional Benefits</th>
<th>Cost-effectiveness</th>
<th>Affordability</th>
<th>Contextual Considerations regarding the illness and therapy</th>
<th>Value Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superior: A</td>
<td>Less relevant</td>
<td>Cost/QALY: more effective and less expensive; or Cost/key outcome &lt; comparator</td>
<td>Cost-saving or budget neutral</td>
<td>Least relevant</td>
<td>Very High: a+</td>
</tr>
<tr>
<td>Superior: A</td>
<td>Less relevant</td>
<td>Cost/QALY &lt; 50-100K or Cost/key outcome ≤ comparator</td>
<td>Added costs not enough to displace more valuable care (qualitative) and/or raise premiums more than growth in GDP (quantitative)</td>
<td>Least relevant</td>
<td>High: a</td>
</tr>
<tr>
<td>Incremental or Better: B, B+</td>
<td>More relevant</td>
<td>Cost/QALY &lt; 50K or more effective and less expensive; or Cost/key outcome ≤ comparator</td>
<td>Cost-saving, budget neutral, or added costs not enough to displace more valuable care (qualitative) and/or raise premiums more than growth in GDP (quantitative)</td>
<td>Less relevant</td>
<td>High: a</td>
</tr>
<tr>
<td>Comparable or Better: C, C+</td>
<td>Very relevant</td>
<td>Cost/QALY: cost-saving (dominates) or budget neutral if significant additional benefits</td>
<td>Cost-saving or budget neutral if significant additional benefits</td>
<td>More relevant</td>
<td>High: a</td>
</tr>
</tbody>
</table>
Figure XX. Mock value table demonstrating use of \textit{hypothetical} criteria in sequence to assign overall value ratings to treatments.

<table>
<thead>
<tr>
<th>Comparative Clinical Effectiveness</th>
<th>Additional Benefits</th>
<th>Cost-effectiveness</th>
<th>Affordability</th>
<th>Contextual Considerations regarding the illness and therapy</th>
<th>Value Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Superior: A</td>
<td>Minimal/none</td>
<td>Cost/QALY &lt; 300K</td>
<td>Added costs likely to displace more valuable care (qualitative) and/or raise premiums more than growth in GDP (quantitative)</td>
<td>More relevant</td>
<td>Low: c</td>
</tr>
<tr>
<td>Incremental or Better: B, B+</td>
<td>Minimal/none</td>
<td>Cost/QALY &lt; 150K</td>
<td>Added costs likely to displace more valuable care (qualitative) and/or raise premiums more than growth in GDP (quantitative)</td>
<td>Less relevant</td>
<td>Low: c</td>
</tr>
<tr>
<td>Comparable: C, C+</td>
<td>Minimal/none</td>
<td>Cost/QALY not relevant if clinically comparable</td>
<td>Any additional cost creates low value; if C+ then additional clinical evidence needed to justify any cost increase</td>
<td>More relevant</td>
<td>Low: c</td>
</tr>
<tr>
<td>Promising but Inconclusive: PI</td>
<td>Minimal/none</td>
<td>Cost/QALY &gt; 150K</td>
<td>Added costs likely to displace more valuable care (qualitative) and/or raise premiums more than growth in GDP (quantitative)</td>
<td>More relevant</td>
<td>Low: c</td>
</tr>
</tbody>
</table>