2020 Value Assessment Framework

Proposed Changes

August 21, 2019

Contents

Executive Summary

This paper describes proposed updates to the ICER value assessment framework, including refinements of its conceptual structure and modifications to the specific methods used to gather, assess, and appraise evidence of different types. These proposals build on several years of experience with the current framework, which applied to reviews launched in July 2017 and later, and 55 letters from 97 organizations and individuals that were submitted as part of a national call for open input that ran from May 2 to June 10, 2019. These comments can be found at http://icer-review.org/material/2020-value-assessment-framework-open-input-comments/.

In this executive summary, we describe proposed changes to the current value assessment framework; the full text contains additional discussion of the rationale behind the proposed changes. We also address several key elements of the framework for which we are not proposing any change in order to explain our reasoning for continuing with current methods despite suggestions for change contained in public comments. Otherwise, elements of the framework that will remain unchanged are generally not discussed in detail, and full descriptions can be found in the 2017-2019 Value Assessment Framework and its adaptations posted to ICER's website (https://icer-review.org/methodology/icers-methods/icer-value-assessment-framework/). Other supporting documents (i.e., ICER's methods for health technology assessment and economic evaluation reference case) can be found at https://icer-review.org/methodology/icers-methods/.

The proposals in this document will be subject to a public comment period from August 16 through October 18, 2019. ICER hopes to receive further comments on these proposed changes. After reviewing all public comments, ICER will reflect further and make any final changes before releasing its Final 2020 Value Assessment Framework on December 18, 2019. This document will present a comprehensive description of all elements of the value assessment framework, and will be released with several companion updated documents, including special methods adaptations for treatments of ultra-rare diseases, the ICER Evidence Rating Matrix, the ICER reference case for economic evaluations, and ICER guides to patient and manufacturer engagement. An additional document detailing methods adaptations for assessments of single or short-term transformative therapies is currently undergoing public consultation through September 6, following which ICER will release a final version on or before November 15.

Comparative Clinical Effectiveness

Sources of Evidence

1. ICER reaffirms use of existing real-world evidence. ICER reaffirms its ongoing commitment to seek and use existing RWE in its reviews. RWE may help complement other types of evidence in assessments of comparative clinical effectiveness, in contributing to assessment of the potential

other benefits of interventions, and in providing useful information to inform the assumptions of economic models. As with all evidence, ICER will assess the internal and external validity of RWE as part of a larger judgment of whether and how that evidence should be incorporated in an assessment. As part of this broad commitment, ICER will continue to formally request that stakeholders who are engaging on a review project submit relevant RWE for consideration in the evidence review.

2. ICER will seek opportunities to generate new RWE for incorporation in reviews. ICER will explore collaborative relationships with organizations that may serve as sources of real-world data in order to generate RWE during reviews that can complement published data sources.

Evidence Rating Matrix: Addition of a New Summary Rating

1. ICER will change its EBM Matrix Evidence Rating categories. ICER will introduce a new rating of C++ and modify the definition of the C+ rating. A rating of C+ will now signify that, versus the comparator, the evidence provides moderate certainty of a comparable or small (but not substantial) net health benefit, with high certainty of at least a comparable net health benefit. The rating C++ will signify that, versus the comparator, the evidence provides moderate certainty of a comparable, small, or substantial net health benefit with high certainty of at least a comparable net health benefit.

2. ICER will revise previous ratings to match new Evidence Rating categories. In order create greater consistency between previous ICER reports and those that will adopt the new definitions of C+ and C++ going forward, we will retrospectively revise all relevant Evidence Ratings in ICER reports from 2017-2019. These revisions will reflect the evidence available at the time of the report, and not rely on subsequent information.

Cross-Reference with German Evidence Ratings

1. ICER will provide complementary evidence ratings using the German categories of "added benefit." Along with its own evidence ratings, ICER will seek to translate its judgment of the evidence into the rating system for added clinical benefit used in Germany to summarize drug assessments and guide pricing considerations.

Long-Term Cost Effectiveness

Measures of Health Gain

1. Quality-Adjusted Life Years (QALY) Analyses. No changes – see full-text discussion section.

2. Equal Value of Life Years Gained (evLYG) Analyses. No changes – see full-text discussion section.

Quantifying Additional Dimensions of Value

1. No changes are proposed through which additional dimensions of value would receive a quantified weighting in the ICER base-case cost-effectiveness model. Within assessments of "single or short-term transformative therapies" we are proposing that additional dimensions of value be included as new categories of "other potential benefits or disadvantages" for appraisal committee voting. However, we are not proposing that these dimensions be quantified separately and used to weight the results of cost-effectiveness analyses. These proposals and their rationale are described in two documents available here: methods proposal, technical brief.

Cost-Effectiveness Threshold Ranges

1. In all reports, ICER will provide a set of results using standardized cost-effectiveness thresholds from $50,000-$200,000 per QALY and per evLYG. ICER will provide cost-per-QALY results at $50,000, $100,000, $150,000 and $200,000 per QALY and per evLYG for all assessments, including those for treatments of ultra-rare disorders.

2. ICER will continue to use the range of $100,000-$150,000 per QALY and per evLYG in presenting value-based price benchmarks. ICER will continue to use the threshold range from $100,000-$150,000 per QALY as the standard for its value-based price benchmarks for all assessments. Value-based price benchmarks using $100,000-$150,000 per evLYG will also be provided.

Base-Case Perspective in Economic Models

1. Base-Case Perspective. No changes proposed – see full-text discussion section.

Discounting

1. Discounting. No changes proposed – see full-text discussion section.

Alternative Economic Modeling Assumptions

1. ICER will add a "Controversies and Uncertainties" section to the cost-effectiveness section of its reports in order to broaden discussion of alternative model structures and assumptions suggested by manufacturers or other stakeholders.

Other Changes

1. ICER will exclude unrelated costs in some cost-effectiveness analyses.

2. When relevant, ICER will seek information from manufacturers and payers with which to model as a scenario analysis a limited number of outcome-based payment arrangements for the intervention under review.

3. Sources of Evidence. No changes proposed – see discussion section.

4. Caregiver Utilities and Costs. No changes proposed – see discussion section.

5. Dynamic Pricing. No changes proposed – see discussion section.

6. Subgroup Analyses. No changes proposed – see discussion section.

7. Public Payer Perspective Incorporating Behavioral Health Outcomes. No changes proposed – see discussion section.

8. Reference Case. ICER's Reference Case will be revised to reflect any of the proposed revisions that are adopted.

Potential Other Benefits and Contextual Considerations

As displayed in Table ES1 below, ICER proposes the following changes:

1. ICER will change the wording of all questions related to potential other benefits and contextual considerations to improve clarity and consistency of interpretation.

2. ICER will add a first question related to whether appraisal committee members believe that uncertainty and/or model assumptions lead the base-case cost-effectiveness results to be overly optimistic or pessimistic.

3. ICER will add several new potential other benefits of a new intervention compared to the selected comparator:
    a. For interventions that may offer special advantages by virtue of presenting an option to patients with a notably different balance or timing of risks and benefits versus other treatments.

b. For interventions that have a delivery mechanism or complexity of regimen that may improve or decrease real-world adherence relative to comparator treatments

4. ICER will provide empirical results for absolute QALY shortfall and proportional QALY shortfall to support deliberation and voting on a single question on relative "health loss" as a contextual consideration. This question will take the place of two separate questions on severity of illness and lifetime burden of illness.

5. ICER will add one new potential disadvantage related to treatments that, if not entirely curative, could reduce or even preclude the potential effectiveness of future treatments.

6. ICER will change the voting structure for all questions from a yes/no format to a Likert scale from 1-3. The intent of the new voting structure is to enhance the application of these considerations by decision-makers within a cost-effectiveness range suggested by the base-case economic model.

Table ES1. Proposed List of Voting Questions on Potential Other Benefits/Disadvantages and Contextual Considerations

1	Intermediate (2)	3
Uncertainty or overly favorable model assumptions creates significant risk that base-case cost-effectiveness estimates are too optimistic		Uncertainty or overly unfavorable model assumptions creates significant risk that base-case cost-effectiveness estimates are too pessimistic
Very similar mechanism of action to that of other active treatments		New mechanism of action compared to that of other active treatments
Delivery mechanism or relative complexity of regimen likely to lead to much lower real-world adherence relative to the comparator		Delivery mechanism or relative complexity of regimen likely to result in much higher real-world adherence relative to the comparator
The intervention offers no special advantages to patients by virtue of presenting an option to patients with a notably different balance or timing of risks and benefits		The intervention offers special advantages to patients by virtue of presenting an option to patients with a notably different balance or timing of risks and benefits
Small health loss without this treatment as measured by proportional and/or absolute QALY shortfalls		Substantial health loss without this treatment as measured by proportional and/or absolute QALY shortfalls
Will not significantly reduce caregiver or broader family burden versus the comparator		Will significantly reduce caregiver or broader family burden versus the comparator
Will not have a significant impact on improving return to work and/or overall productivity versus the comparator		Will have a significant impact on improving return to work and/or overall productivity versus the comparator
Other		Other

Potential Budget Impact Analysis

1. ICER will extend the time period over which we average the annual number of drugs approved by the FDA from two years to five years.

   ICER recalculates the potential budget impact threshold each calendar year, using the most recent inputs available. In the recalculation of ICER's potential budget impact threshold for calendar year 2019, we have now extended the time period over which we average the annual number of drugs approved by the FDA from two to five years, to reduce fluctuations in the threshold due to this variable. See Table ES2 for the updated calculations used to derive the threshold for 2019.

Table ES2. Potential Budget Impact Threshold Calculations

Item	Parameter	Estimate	Source
1	Growth in US GDP, 2019 (est.) +1%	3.5%	World Bank, 2019
2	Total personal medical health care spending, 2018	$2.95 trillion	CMS National Health Expenditure, 2019
3	Contribution of drug spending to total health care spending (%) (Row 4 ÷ Row 2)	16.9%	Calculation
4	Contribution of drug spending to total health care spending, 2018	$498.6 billion	CMS National Health Expenditures, 2019 Altarum Institute, 2018
5	Annual threshold for net health care cost growth for ALL drugs (Row 1 x Row 4)	$17.4 billion	Calculation
6	Average annual number of new molecular entity approvals, 2014-2018	42.6	FDA, 2019
7	Annual threshold for average cost growth per individual new molecular entity (Row 5 ÷ Row 6)	$409.6 million	Calculation
8	Annual threshold for estimated potential budget impact for each individual new molecular entity (Doubling of Row 7)	$819 million	Calculation

2. <u>ICER will add the following new language to our economic reference case providing greater detail regarding our methods of potential budget impact analysis:</u>

"ICER uses the cost-effectiveness model in an economic evaluation to estimate the potential total budgetary impact of new treatments in the US, assuming different prices, including the treatment's list and net prices, and the three threshold prices to achieve cost effectiveness at $50,000, $100,000, and $150,000 per QALY. Results from the cost-effectiveness model are used to provide undiscounted net costs (including intervention/comparator costs, other health care costs, and total costs) broken out by year for years one through five, for use in the potential budget impact analyses. Potential budget impact is defined as the total differential cost of using each new therapy rather than relevant existing therapy for the treated population, calculated as differential health care costs (including drug costs) minus any offsets in these costs from averted health care events.

Potential budget impact analyses are based on net cost per patient and estimates of the proportion of the US population eligible for treatment with the new intervention. ICER uses

epidemiologic and other data to estimate the size of the potential candidate population for each new treatment. We then assume that an equal proportion of patients (20%) would be treated with the new treatment each year over five years, arriving at a cumulative 100% uptake at five years.

The goal of ICER's potential budget impact analysis is to estimate the net cost per patient treated with new interventions so that decision-makers can use their own assumptions about uptake and pricing to determine their own estimate of potential budget impact. We also seek to produce calculations that will help policy makers identify situations in which the potential uptake of a new treatment, at various pricing levels, might exceed a budget impact threshold that signifies that the budget impact in the near term (over 5 years) would contribute to overall health care cost growth at a higher rate than growth in the national economy (plus 1%).

To accomplish these goals, ICER's potential budget impact analyses must evaluate whether a new drug would be likely to take market share from one or more drugs. ICER will continue to use clinical expert opinion regarding the treatments likely to be displaced by use of a new treatment within the eligible population. ICER will then follow one of the procedures listed below, dependent on whether existing treatments are being displaced. These are explicitly NOT meant to represent our assumptions of the budget impact of new interventions that are most likely in the real world. Our methods are intended to provide the calculations that can underpin a graphic figure that allows decision-makers and policy makers to make their own assumptions.

- No existing active treatment: If the intervention is for a condition which has no existing active treatment in the market (other than best supportive care), we will calculate potential budget impact for 100% of the eligible population at the end of five years (20% marginal new uptake per year).
- Existing treatments launched within prior 2 years: If the intervention is for a condition with existing active treatment(s), one or more of which was launched within the last two years, equal proportions of the eligible population will be split among the intervention and the recently launched treatment(s), with 100% displacement of relevant treatments launched more than two years ago.
- Existing treatments all on market >2 years: If the intervention is for a condition with existing active treatment(s) all launched more than two years ago, we will calculate potential budget impact for 100% of the eligible population at the end of five years, with displacement of existing treatments.
- Multiple existing treatments: When there are multiple existing treatments on the market, clinical expert opinion will be used to estimate the percentage of patients converted from each existing treatment to the new treatment.

- Untreated patients: For all cases, we will include the untreated portion of the eligible population, as long as they are considered eligible for the new treatment.

3. ICER will present a cumulative per-patient potential budget impact. ICER will now present a cumulative per-patient potential budget impact for each year over the five-year time horizon, with results being presented graphically for each intervention assessed, and numerical data presented in tabular format in an appendix of the report. This graph will replace the prior tables that reported five-year annualized potential budget impact per patient.

Report Development and Public Meetings

Report Development

1. ICER will extend the timeline for large class reviews by nine weeks.

2. ICER will implement a formal process through which to reassess whether new evidence has emerged that should be included in an update to the report one year after the release of a Final Evidence Report.

3. ICER will make the following changes to public comment periods:
    a. Extend the draft report public comment period for class reviews by one week as part of the aforementioned timeline extension.
    b. Extend the word limit for written summaries of oral public comments included in the final report from 250 to 750 words.

4. ICER will create a new "Patient Perspectives" chapter for its reports that will describe the input we have received from patients, families, and patient organizations, as well as relevant sources of patient-generated evidence. We will also summarize relevant sources of patient-generated evidence that have been shared by patients and identified through our research process.

5. Methods Transparency. No changes – see discussion section.

6. Policy Guidance for Stakeholders. No changes – see discussion section.

Public Meetings

1. Council Membership. No changes – see discussion section.

2. ICER will post annual COI disclosure statements to its website for each voting council.

3. ICER will adopt a code of conduct for public meetings.

Stakeholder Engagement

1. ICER will update the following patient engagement materials and approaches:
    a. Revise patient engagement materials to include examples of how patient input informed reviews.
    b. Revise the language of its patient input survey to include PICOTS language
    c. Continue to include suggestions that were adopted in the "Stakeholder Input" section of scoping documents, and will expand the section to include discussion of suggestions that were not adopted.

2. Economic Model Transparency. No changes – see discussion section.

3. ICER will formalize the practice of debriefing with patient groups after a review has concluded.

4. ICER will produce a series of lay-friendly seminars that will provide background on evidence-based medicine and its application to health technology assessment.

1. Introduction

This paper describes proposed updates to the ICER value assessment framework, including refinements of its conceptual structure and modifications to the specific methods used to gather, assess, and appraise evidence of different types. These proposals build on several years of experience with the current framework, which applied to reviews launched in July 2017 and later, and 55 letters from 97 organizations and individuals that were submitted as part of a national call for open input that ran from May 2 to June 10, 2019. These comments can be found at http://icer-review.org/material/2020-value-assessment-framework-open-input-comments/.

In the sections that follow, we describe proposed changes to the current value assessment framework. We also address several key elements of the framework for which we are not proposing any change in order to explain our reasoning for continuing with current methods despite suggestions for change contained in public comments. Otherwise, elements of the framework that will remain unchanged are generally not discussed in detail, and full descriptions can be found in the 2017-2019 Value Assessment Framework and its adaptations posted to ICER's website (https://icer-review.org/methodology/icers-methods/icer-value-assessment-framework/). Other supporting documents (i.e., ICER's methods for health technology assessment and economic evaluation reference case) can be found at https://icer-review.org/methodology/icers-methods/.

The proposals in this document will be subject to a public comment period from August 16 through October 18, 2019. ICER hopes to receive further comments on these proposed changes. After reviewing all public comments, ICER will reflect further and make any final changes before releasing its Final 2020 Value Assessment Framework on December 18, 2019. This document will present a comprehensive description of all elements of the value assessment framework, and will be released with several companion updated documents, including special methods adaptations for treatments of ultra-rare diseases, the ICER Evidence Rating Matrix, the ICER reference case for economic evaluations, and ICER guides to patient and manufacturer engagement. An additional document detailing methods adaptations for assessments of single or short-term transformative therapies is currently undergoing public consultation through September 6, following which ICER will release a final version on or before November 15.

1.1. Overarching Purpose and Principles of the ICER Value Assessment Framework

For more than 10 years ICER has been active in developing methods for evidence assessment. Evidence assessment, however, is only one component of ICER's broader effort to provide mechanisms through which all stakeholders and the general public can engage in discussions on how best to use evidence as the foundation for a more effective and sustainable health care

system. A formal effort was undertaken between 2014-2015 to gain input through a multi-stakeholder advisory group on ways to define with greater detail the conceptual and methodological underpinnings of ICER reports – a "value assessment framework." Ultimately, the purpose of the value assessment framework is to form the backbone of rigorous, transparent evidence reports that, within a broader mechanism of stakeholder and public engagement, will help the United States evolve toward a health care system that provides fair pricing, fair access, and a sustainable platform for future innovation.

In this effort ICER is guided by several key underlying principles. One is that we act with respect for all, in concordance with a presumption of good will on the part of all participants and stakeholders in the health care system. ICER does not intend to target any particular interest group or organization. There are many areas in which the US health system fails to serve patients well, in which access to care is suboptimal, waste and inefficiency pose major problems, and costs to patients and the health system fail to align with added value. ICER believes that only through collaborative efforts, built upon a foundation of civil discourse and honest consideration of evidence on effectiveness and value, can lasting progress be made on behalf of patients today and those of the future.

The ethical vision inherent in ICER's work recognizes that many choices that are made in health care – choices in clinical care, insurance coverage, pricing, payment, and allocation of resources within health systems – must address the basic reality that societal resources for health care are not unlimited, and that there will always be trade-offs and dilemmas over how to organize and pay for the services provided within a health system. Too often, these decisions are made without rigorous evidence and with little transparency. Too often, there is little chance for reflection or public engagement in managing the tensions that can arise between innovation, access, and costs. ICER's value assessment framework seeks to place scientific methods of evidence analysis at the heart of a clearer and more transparent process. The value framework reflects our strong underlying belief that rigorous thinking about evidence can prevent the kind of waste that strains our ability to provide patient-centered care. The framework also is intended to support discussions about the best way to align prices for health services with their true added value for patients. While considering value and linking it to pricing and insurance coverage cannot solve every dilemma, nor satisfy every need, ICER believes it offers the best hope of avoiding rationing of care by the ability of patients to pay for care, and that it can promote a more dynamic, innovative health care system that will make the best use of available resources in caring for all patients.

1.2. The Population Perspective and Intended Uses of the ICER Value Framework

The ICER value framework describes the conceptual framework and set of associated methods that guide the development of ICER evidence reports. ICER reports are intended to support deliberation

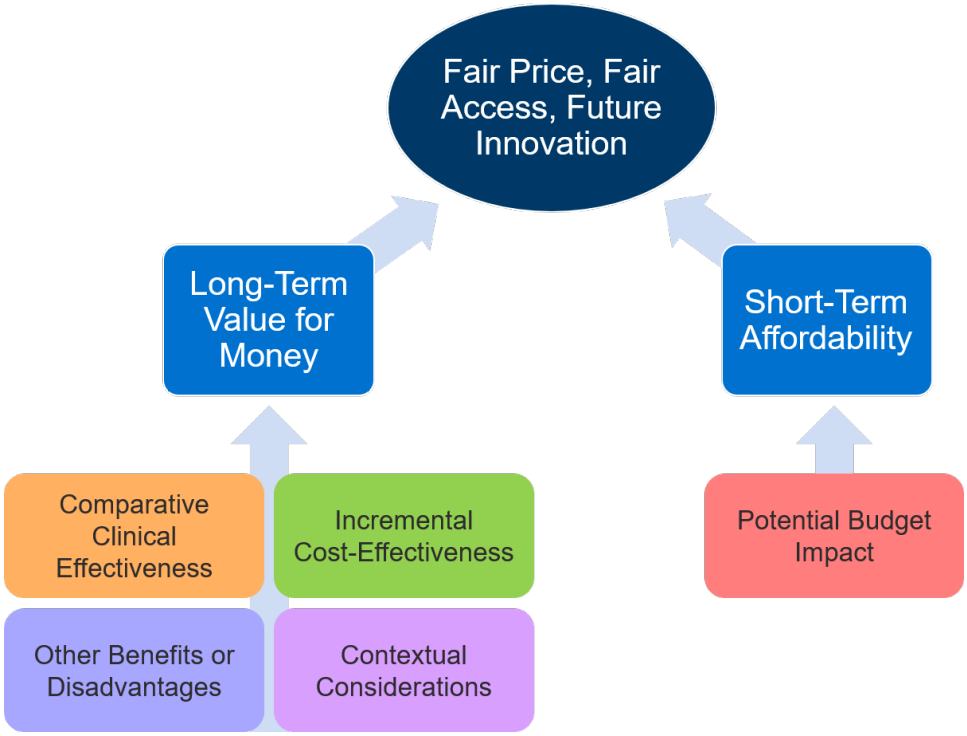
on medical policies related to health services (e.g., tests or treatments) and delivery system interventions (e.g., preventive programs, changes to the organization of medical personnel). To inform these kinds of medical policies the ICER value framework takes a "population" level perspective as opposed to trying to serve as a shared decision-making tool to be used by individual patients and their clinicians. Taking a population perspective implies that the ICER value framework seeks to analyze evidence in a way that supports population-level decisions and policies, such as broad guidelines on appropriate care, pricing, insurance coverage determinations, and payment mechanisms. A value framework intended to support decisions about the care of individual patients requires a structure that invites weighting of benefits, harms, and costs from the individual patient's perspective. There is an important need for better evidence-based shared decision-making tools for individual patients and clinicians, but this is not the primary intended purpose of the ICER value framework or of ICER reports.

Even with its population-level focus, however, the ICER value framework seeks to encompass and reflect the experiences and values of patients. Representing the diversity of patient outcomes and values in a population-level framework is difficult because there will always be an inherent tension between average findings in clinical studies and the uniqueness of every patient. There will also always be diversity in the way that patients view the balance of risks and benefits of different treatment options. The ICER value framework does not solve these tensions, but neither does it obscure them. Population-level decisions and policies have always been made by life science companies, insurers, and clinical organizations looking at evidence in the same general way. One important goal of the ICER value framework is to provide an evidence report that does a better job of analyzing the strengths and limitations of the available evidence, including what is or is not known about the variation in response to different treatments among patients with different personal and clinical characteristics. The ICER value framework also creates an explicit place and role for consideration of elements of value that are important to individual patients but that fall outside traditional clinical measures.

2. Conceptual Structure

Below, we present the updated conceptual structure for ICER's value framework. The only change that has occurred since 2017 is new language describing that the goal we believe should be sought by all stakeholders is "fair price, fair access, and future innovation." We believe this describes more clearly the ultimate aim for decision-making that we intend our value assessment framework to support. Otherwise, there are no proposed changes to the general conceptual structure of the value assessment framework. A detailed description of this conceptual structure may be found on pages 5-9 of the 2017-2019 framework, available at http://icer-review.org/wp-content/uploads/2017/06/ICER-value-assessment-framework-Updated-050818.pdf.

Figure 2.1. Updated Conceptual Structure of the ICER Value Assessment Framework



3. Comparative Clinical Effectiveness

3.1 Sources of Evidence

Proposed Changes

1. ICER reaffirms use of existing real-world evidence. ICER reaffirms its ongoing commitment to seek and use existing RWE in its reviews. RWE may help complement other types of evidence in assessments of comparative clinical effectiveness, in contributing to assessment of the potential other benefits of interventions, and in providing useful information to inform the assumptions of economic models. As with all evidence, ICER will assess the internal and external validity of RWE as part of a larger judgment of whether and how that evidence should be incorporated in an assessment. As part of this broad commitment, ICER will continue to formally request that stakeholders who are engaging on a review project submit relevant RWE for consideration in the evidence review.

2. ICER will seek opportunities to generate new RWE for incorporation in reviews. ICER will explore collaborative relationships with organizations that may serve as sources of real-world data in order to generate RWE during reviews that can complement published data sources.

Discussion

ICER has consistently sought to incorporate analysis of RWE into our reports whenever it can provide additional perspective on comparative clinical effectiveness or cost-effectiveness. In addition to searching for published RWE and seeking RWE in the grey literature, on several occasions we have collaborated with patient and other stakeholder organizations to obtain new patient and caregiver survey information when it was not available in the medical literature. Findings from this work have been included in our Evidence Reports and helped inform discussions during our Public Advisory meetings and Council member votes.

RWE often has greater vulnerability to known and unknown biases that create limitations in our ability to rely on it when making judgments about relative effectiveness of different care options. Nonetheless, we understand that randomized controlled clinical trials have their own limitations and are often inadequate to address all questions relevant to assessments of comparative clinical effectiveness. RWE can be particularly helpful under certain circumstances such as when long-term safety of a treatment or durability of a medication's effect is unclear. We have also emphasized how RWE can be helpful in supporting consideration of a treatment's "potential other benefits" that lie outside traditional clinical trials. Patient-reported outcome studies and studies that capture broader patient and family effects of treatment are especially desired as they can provide evidence usually not included in clinical trials.

Aside from peer-reviewed and published real-world evidence, there are numerous sources of real-world data that could prove informative in an assessment, including data from anonymized electronic medical records, insurance claims, and patient and caregiver surveys and questionnaires. Our role has not included an emphasis on using these data sources to perform *de novo* studies. However, looking forward, we are announcing with this framework update a commitment to explore forming collaborations with organizations to leverage these kinds of data for new analyses. Such analyses would need to address key gaps in the evidence base and be feasible within the timelines of an ICER review. Any *de novo* analyses would also need to be transparent to all stakeholders so that all participants can engage in deliberation on their validity and relevance.

3.2 Evidence Rating Matrix: Addition of a New Summary Rating

Proposed Changes

1. ICER will change its EBM Matrix Evidence Rating categories. ICER will introduce a new rating of C++ and modify the definition of the C+ rating. A rating of C+ will now signify that, versus the comparator, the evidence provides moderate certainty of a comparable or small (but not substantial) net health benefit, with high certainty of at least a comparable net health benefit. The rating C++ will signify that, versus the comparator, the evidence provides moderate certainty of a comparable, small, or substantial net health benefit with high certainty of at least a comparable net health benefit.

2. ICER will revise previous ratings to match new Evidence Rating categories. In order create greater consistency between previous ICER reports and those that will adopt the new definitions of C+ and C++ going forward, we will retrospectively revise all relevant Evidence Ratings in ICER reports from 2017-2019. These revisions will reflect the evidence available at the time of the report, and not rely on subsequent information.

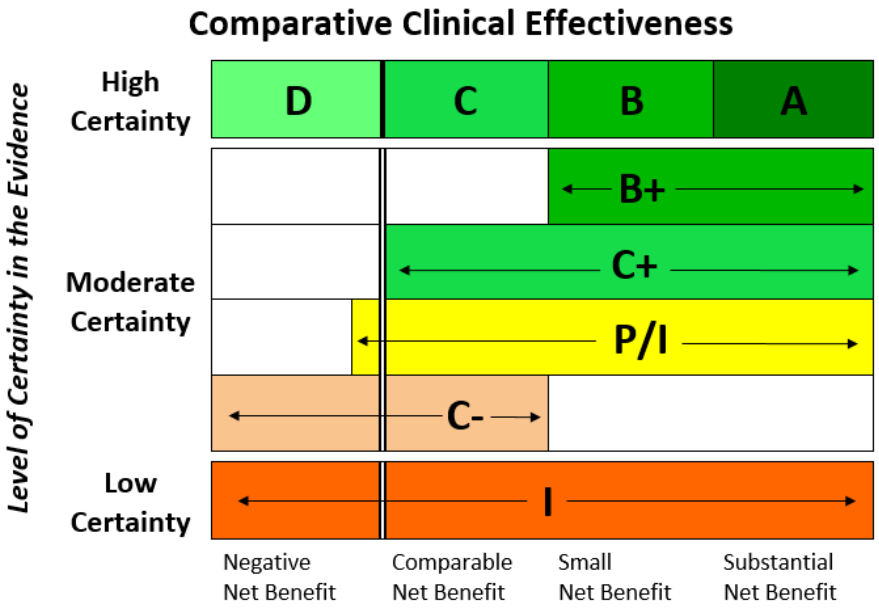
Discussion

1. Evidence Rating Categories. The current ICER Evidence Rating Matrix includes four evidence grades (B+, C+, C-, and P/I) in the moderate certainty domain (Figure 1a). These ratings are assigned when the conceptual confidence interval surrounding a point estimate extends across two or three categories of comparative net health benefit. The precision of a judgement of comparative net health benefit may vary for different evidence ratings that fall within the moderate certainty domain. For example, a new drug ("Drug X") may offer a distinct advantage over existing treatments, but the true level of incremental benefit (i.e., small vs. substantial) is not yet known. In this situation, the conceptual confidence interval would extend across two categories of benefit (small and substantial), and Drug X would receive a B+ rating. The evidence for another drug ("Drug Y") may provide high certainty that "Drug Y" is not inferior to its comparator, but there may be

insufficient evidence to determine whether the net health benefit is comparable, small, or substantial. The conceptual confidence interval surrounding the point estimate for Drug Y would therefore extend across three categories of benefit (comparable, small, substantial), and Drug Y would receive a C+ rating.

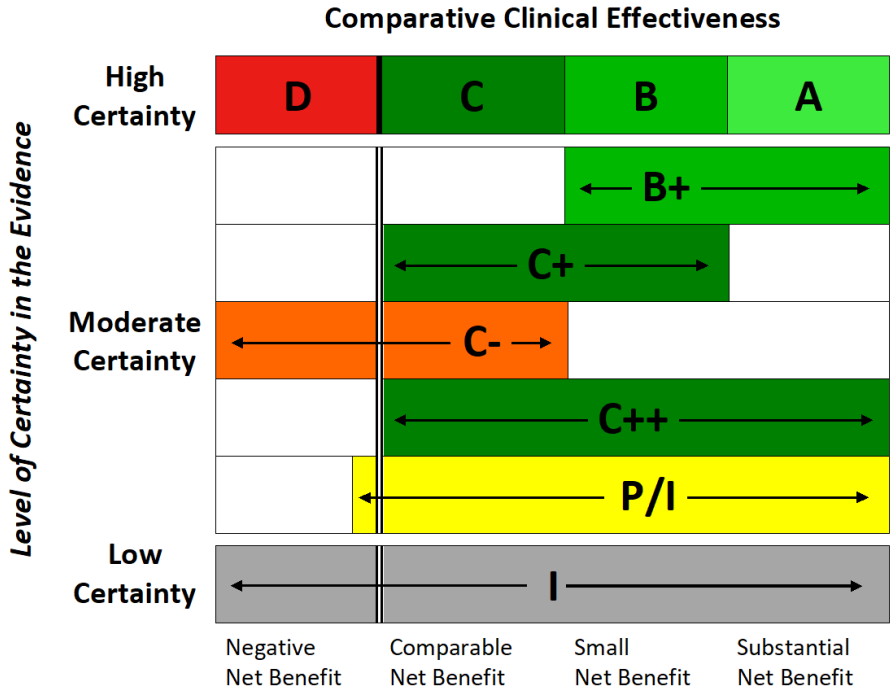
When the evidence supports greater precision, we think it is important to specify where the upper and lower limits of our conceptual confidence interval fall. Accordingly, we will introduce an additional rating of C++ and modify the definition of the C+ rating. Under the new terminology, C++ will signify that, versus the comparator, the evidence provides moderate certainty of a comparable, small, or substantial net health benefit, with high certainty of at least a comparable net health benefit. A rating of C+ will now signify that, versus the comparator, the evidence provides moderate certainty of a comparable or small (but not substantial) net health benefit, with high certainty of at least a comparable net health benefit. The updated matrix (Figure 1b) is intended to provide greater specificity, when the evidence supports such precision. We believe this will assist decision-makers in applying the ICER Evidence Rating Matrix in a more transparent, reliable, and consistent fashion.

Figure 1a. Current ICER Evidence Rating Matrix



A = "Superior" *- High certainty of a substantial (moderate-large) net health benefit*
B = "Incremental" *- High certainty of a small net health benefit*
C = "Comparable" *- High certainty of a comparable net health benefit*
D = "Negative" *- High certainty of an inferior net health benefit*
B+ = "Incremental or Better" *- Moderate certainty of a small or substantial net health benefit, with high certainty of at least a small net health benefit*
C+ = "Comparable or Better" *- Moderate certainty of a comparable, small, or substantial net health benefit, with high certainty of at least a comparable net health benefit*
P/I = "Promising but Inconclusive" *- Moderate certainty of a comparable, small, or substantial net health benefit, and a small (but nonzero) likelihood of a negative net health benefit*
C- = "Comparable or Inferior" *- Moderate certainty that the point estimate for comparative net health benefit is either comparable or inferior*
I = "Insufficient" *- Any situation in which the level of certainty in the evidence is low*

Figure 1b. Proposed ICER Evidence Rating Matrix



A = "Superior" *- High certainty of a substantial (moderate-large) net health benefit*
B = "Incremental" *- High certainty of a small net health benefit*
C = "Comparable" *- High certainty of a comparable net health benefit*
D= "Negative" *- High certainty of an inferior net health benefit*
B+= "Incremental or Better" *– Moderate certainty of a small or substantial net health benefit, with high certainty of at least a small net health benefit*
C+ = "Comparable or Incremental" *- Moderate certainty of a comparable or small net health benefit, with high certainty of at least a comparable net health benefit*
C- = "Comparable or Inferior" *– Moderate certainty that the point estimate for comparative net health benefit is either comparable or inferior*
C++ = "Comparable or Better" *- Moderate certainty of a comparable, small, or substantial net health benefit, with high certainty of at least a comparable net health benefit*
P/I = "Promising but Inconclusive" *- Moderate certainty of a small or substantial net health benefit, small (but nonzero) likelihood of a negative net health benefit*
I = "Insufficient" *– Any situation in which the level of certainty in the evidence is low*

3.3 Cross-Reference with German Evidence Ratings

Proposed Changes

1. ICER will provide complementary evidence ratings using the German categories of "added benefit." Along with its own evidence ratings, ICER will seek to translate its judgment of the evidence into the rating system for added clinical benefit used in Germany to summarize drug assessments and guide pricing considerations.

Discussion

As ICER's work has gained use internationally, interest has been expressed in comparing ICER evidence ratings to those from health technology assessment groups that provide similar reviews for policy making purposes in other countries. Germany is the largest pharmaceutical market in Europe and has a sophisticated evidence review system grounded in assessments by the Institute for Quality and Efficiency in Health Care (IQWiG) that are then deliberated upon by the nation's Federal Joint Committee (G-BA).[1,2] The evidence rating system used by these German organizations results in assignment of a rating for "added benefit" that is separated into six categories: 1) major added benefit, 2) considerable added benefit, 3) minor added benefit, 4) non-quantifiable added benefit, 5) no added benefit proven, 6) less benefit.[3]

We propose to provide our own judgment of "added benefit" within the German categories to complement ICER's own methods. We propose to translate the ICER assessment into the German categories, rather than rate the evidence in the same manner as would be done in Germany as there are important differences in the two methods that must be acknowledged. First, the German categories do not have an explicit axis related to "level of certainty" that modulates the evidence rating; instead, uncertainty is factored into whether there is adequate evidence to demonstrate any added benefit or not, and whether that benefit can be quantified at all, or not. Second, the German methods stipulate specific patient outcomes, such as mortality, serious symptoms, health-related quality of life, and non-serious symptoms, that are the sole focus for judgments of added benefit. Notably, orphan drugs, by their very designation, are automatically deemed to have some added benefit, although the manufacturer is still required to demonstrate how much. ICER's conceptualization of "net health benefit" may in some cases be broader than the specific outcomes viewed as relevant by the German system, and our rating blends consideration of harms and benefits more explicitly than the German rating system.

A third distinction is that the German methodology has suggested specific quantitative thresholds for improvements in the specified patient outcomes to merit placement in a particular category of added benefit. ICER has chosen not to seek a quantitative threshold for its judgments between comparable, incremental, and substantial net health benefit.

A final difference in the two rating systems is linked to judgments regarding the role of indirect assessments in judgments of comparative clinical effectiveness. The German system tends not to admit indirect assessments (e.g. network meta-analyses) as adequate for demonstrating added benefit, whereas ICER has favored the inclusion of indirect assessments in its reports, particularly when there are no head-to-head trials of active comparator agents. This difference in opinion on the relative validity and utility of indirect assessments creates the likelihood that the German system will rate a body of evidence differently from ICER, even if both organizations are using the same evidence rating scheme.

Despite these important differences, we feel providing our judgment of the evidence within a secondary rating system may help decision-makers consider different ways to consider the strength of evidence behind new interventions, and it may spur further dialogue and calibration of evidence assessments across important pharmaceutical markets.

ICER will seek to be fully transparent in describing our rationale for assigning both our own evidence rating and that within the German categorical system of added benefit. As a rough algorithm for the crosswalk between the two rating systems, we envision the following. We will note for orphan drugs that the German system would, at minimum, rate them as "non-quantifiable added benefit" but we will also give our judgment of an added benefit rating without this consideration.

Table 3.1. Crosswalk Between German and ICER Evidence Rating Categories

German Rating of "Added Benefit"	ICER EBM Matrix Rating of "Comparative Clinical Effectiveness"
Major Added Benefit	A
Considerable Added Benefit	A
Minor Added Benefit	B
Non-quantifiable Added Benefit	B+
No Added Benefit Proven	C+, C++, Promising but Inconclusive (P/I), C, I
Less than Comparator	D

3. Long-Term Cost Effectiveness

3.1 Measures of Health Gain

Proposed Changes

1. Quality-Adjusted Life Years (QALY) Analyses. No changes – see discussion section.

2. Equal Value of Life Years Gained (evLYG) Analyses. No changes – see discussion section.

Discussion

1. QALY Analyses. ICER does not propose any changes to our use of the QALY as part of assessments that compare therapies on their ability to improve quality of life and lengthen life. The QALY is the gold standard for measuring how well a medical treatment improves and lengthens patients' lives, and therefore has served as a fundamental component of cost-effectiveness analyses in the US and around the world for more than 30 years. Cost-effectiveness analysis examines evidence for entire patient populations, comparing the health benefits and economic costs of different treatment options. A common measure of improved outcomes for patients is needed for these analyses to support broader efforts by governments, private insurers, and drug manufacturers to make more transparent, evidence-based coverage policies and pricing decisions.

Economic analyses using the QALY make treatments that alleviate serious illness look especially valuable. Because the QALY records the degree to which a treatment improves patients' lives, treatments for people with serious disability or illness have the greatest opportunity to demonstrate more QALYs gained and justify a higher price. For this reason, ICER has found that many innovative and expensive new treatments are highly cost-effective, including CAR-T for childhood leukemia at $475,000/treatment, emicizumab for hemophilia at $450,000/year, personalized lung cancer drugs at $90,000/year, and Zolgensma gene therapy for spinal muscular atrophy at $2.1 million for a single treatment.[4-7]

2. evLYG Analyses. We received several comments on the use and inclusion of the evLYG analysis in our reports. Most comments received on this topic recommended against its use, citing its inability to accurately value treatments or capture patient quality of life. However, we also received some comments encouraging its inclusion in our economic evaluations as a complement to the QALY that provides policymakers with additional information to support the development of evidence-based policies, especially for rare diseases.

Concerns have been raised that the QALY potentially undervalues treatments that improve survival in conditions associated with disability or serious illness. In most cases, the QALY would capture the

benefits of improved survival, as our assessments examine the incremental changes in quality of life from treatment, regardless of the baseline level of quality of life. However, in cases where life is prolonged without substantial improvements in quality of life, there may be a perception among some that the QALY could discriminate against treatments for certain patient groups. To help place treatment outcomes in a broader context, ICER will continue to highlight an element in our reports that provides policymakers with information that weighs extension of life expectancy equally across all conditions.

The evLYG analysis counts any gains in length of life equally, regardless of the treatment's ability to improve patients' quality of life. For all additional years of life gained, this analysis will award full health (i.e., the quality of life of the general population), irrespective of the health state patients are in during these additional years of life gained. In other words, if a treatment adds a year of life to a vulnerable patient population – whether treating individuals with cancer, multiple sclerosis, diabetes, epilepsy, or a severe lifelong disability – that treatment will receive the same evLYG as a different treatment that adds a year of life for healthier members of the community.

ICER reaffirms the continued use of the evLYG in its economic analyses, whenever relevant and feasible given model structure, as a supplement and not a replacement to the cost per QALY analysis, which reflects the true benefits a treatment may have on the quality of life on the population in which it is used. In certain situations, model structure may make the calculation of evLYG intractable, in which case we will report life years gained rather than evLYG.

Using both the cost per QALY and the cost per evLYG results will enable policy makers to gain a broad overview of the cost-effectiveness of treatments while ensuring that results will be available to demonstrate whether there is any impact of extended life at a low quality of life. By understanding a treatment's cost per evLYG, as well as its traditional cost per QALY, we believe policymakers can be reassured that they are considering information that poses no risk of discrimination against any patient group. If ICER's analysis finds a major difference in these two measures, we will include specific language in our report describing the underlying characteristics of the treatment and the condition that lead to the difference. More information on the evLYG analysis is available here.

3.2 Quantifying Additional Dimensions of Value

Proposed Change

1. No changes are proposed through which additional dimensions of value would receive a quantified weighting in the ICER base-case cost-effectiveness model. Within assessments of "single or short-term transformative therapies" we are proposing that additional dimensions of value be included as new categories of "other potential benefits or disadvantages" for appraisal committee voting. However, we are not proposing that these dimensions be quantified

separately and used to weight the results of cost-effectiveness analyses. These proposals and their rationale are described in two documents available here: methods proposal, technical brief.

Discussion

1. Additional Dimensions of Value. We received public comments urging ICER to include additional elements of value in our analysis quantitatively, rather than approaching them qualitatively in our reviews and having them voted upon as part of the appraisal committee meeting. A recent report from the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Special Task Force on Value Frameworks by Lakdawalla et al. highlighted eight elements of value that have been proposed by some academics and policymakers as important to decision-making but which may not be adequately captured by the standard QALY.[8] These suggested elements of value include reduction in uncertainty, fear of contagion, insurance value, severity of disease, value of hope, real option value, equity, and scientific spillovers. Although recommending that consideration be given to incorporating these additional dimensions of value whenever relevant, Lakdawalla et al. and the Second Panel on Cost-Effectiveness both acknowledge that these additional elements of value remain controversial and that methods for empirically integrating them into a value-based price are not well established.[8-10] As a result, recommendations focus on the need for further research into methods for quantitative and/or qualitative incorporation into technology assessments.

In evaluating the potential for alternative assessment methods that would integrate these additional dimensions of value, there are several key challenges. First, as noted above, methods for the quantification of these value dimensions are viewed by many health economists as too exploratory for routine incorporation into assessments. For example, the value of hope may be tied empirically to the risk attitudes of patient groups that vary widely depending on the severity of the condition and the prospects for future treatments to be effective. While scientific spillover effects can be demonstrated, it remains unclear how to identify which new treatment approaches are more or less likely to lead to future positive spillover effects, and to estimate in any way how much weight to lend to this forecast. Similar difficulties confront efforts to quantify real option value, whereas insurance value overlaps significantly with considerations around severity or burden of illness.

All of these potential additional elements of value raise questions of whether there needs to be some form of "negative" scoring on these dimensions to balance the positive added value for some interventions within an overall understanding of opportunity costs within the health system. Thus, it is unclear how the inclusion of these additional elements should change the cost-effectiveness threshold used as a general guide to decision-making in order to accommodate an increased valuation for some interventions. ICER therefore believes that there are strong conceptual and practical reasons not to add quantified additional dimensions of value into our cost-effectiveness analyses at this time.

3.3 Cost-Effectiveness Threshold Ranges

Proposed Change

1. In all reports, ICER will provide a set of results using standardized cost-effectiveness thresholds from $50,000-$200,000 per QALY and per evLYG. ICER will provide cost-per-QALY results at $50,000, $100,000, $150,000 and $200,000 per QALY and per evLYG for all assessments, including those for treatments of ultra-rare disorders.

2. ICER will continue to use the range of $100,000-$150,000 per QALY and per evLYG in presenting value-based price benchmarks. ICER will continue to use the threshold range from $100,000-$150,000 per QALY as the standard for its value-based price benchmarks for all assessments. Value-based price benchmarks using $100,000-$150,000 per evLYG will also be provided.

Discussion

1. Standardized Cost-Effectiveness Thresholds for all Assessments. We received comments arguing for the use of different cost-effectiveness threshold values, either in general or in specific cases, such as for end-of-life treatments. We also received requests that we not use thresholds at all, or that we adopt thresholds that vary depending on patient or disease characteristics.

ICER's first draft proposals for the 2017-2019 methods update included a proposal to create a stepwise set of cost-effectiveness thresholds related to different levels of severity of illness and/or lifetime burden of illness.[11] Public comment from patient groups and manufacturers was nearly uniformly negative to this proposal and it was dropped in favor of retaining a single cost-effectiveness threshold range for all assessments. Current public comment has again included some recommendations to adopt differential cost-effectiveness thresholds for different types of treatments and/or different types of conditions. In part, the challenge in this area is that many people accept a broad ethical value to prioritize treatments for the worst off, but arriving at a single quantifiable measure for this concept is difficult and raises thorny questions about whether the goal should be to prioritize the absolute loss of health ("absolute QALY shortfall") or the loss of health in relation to the amount of time patients have left to live ("proportional QALY shortfall"). Either approach creates "winners and losers" among treatments that often causes equity concerns and other concerns about unintended consequences.

Given that there continues to be no strong consensus among academic health economists or ethicists on whether or how to quantify and integrate these values into cost-effectiveness analyses, we have judged that it remains premature to seek to create a separate series of cost-effectiveness thresholds related to severity, burden of illness, or "need." As discussed later in this set of proposed changes, we will propose to bring greater clarity and empiric results to these issues as

part of the deliberation and voting on "contextual considerations" performed as part of every public meeting of our independent appraisal committees.

As a consequence, ICER proposes to use a common set of cost-effectiveness thresholds for all assessments. Moreover, we propose to extend that common set of provided thresholds to treatments of ultra-rare disorders, where previously we have provided a broader range of results, from $50,000 per QALY/evLYG up to $500,000 per QALY/evLYG. Instead, we propose to provide a uniform range of results from $50,000 to $200,000 per QALY/evLYG for all assessments.

We are making this proposal for several reasons. First, there remain important equity concerns related to extending the threshold range higher for treatments just because they treat a small population.[12] In addition, the economic landscape for treatments of rare and ultra-rare conditions has shifted. Years ago, when drug prices were far lower on average, it could be reasonably argued that the profit required to sustain innovation in rare disease treatments required pricing that far exceeded standard cost-effectiveness thresholds. But in today's market environment, it only takes $100,000 per treatment course, multiplied by a mere 10,000 patients, to provide $1 billion per year in revenue. We therefore judge that today it no longer seems necessary to make important exceptions to applying standard cost-effectiveness thresholds to analyzing the value of treatments of rare or ultra-rare conditions.

A final reason for shifting to presenting results for all assessments from $50,000-$200,000 per QALY and evLYG comes from mischaracterization of our current methods for treatments of ultra-rare disorders, in which we present results extending up to $500,000 per QALY/evLYG. Some manufacturers have messaged publicly that this implies that ICER has formalized $500,000 per QALY as the acceptable cost-effectiveness ceiling for these treatments. We have not. As we state in our current methods, our view of treatments for ultra-rare conditions includes the historical perspective that decision-makers have often accepted prices beyond standard cost-effectiveness ranges, particularly for treatments of very small ultra-rare populations. We will continue to include standard language to this effect when presenting value-based price benchmarks for these treatments. But we feel that the unintended consequence of presenting results up to $500,000 per QALY is serious enough that we should no longer provide results within this much broader spectrum. Since our range for value-based price benchmarks remains $100,000-$150,000 per QALY and evLYG, we will provide a broader range of results symmetrically around this range, from $50,000-$200,000 per QALY/evLYG. We believe this is a broad enough range to accommodate the needs of decision-makers in the US to think about their own desired interpretation of cost-effectiveness thresholds.

Although ICER proposes to use a standardized threshold range across all assessments, our reports will continue to include discussion of contextual factors and other important considerations for all therapies, including those for ultra-rare disease or short-term transformative treatments. We also acknowledge that, no matter the threshold or range selected, ICER and the broader HTA community

have a responsibility to educate potential users of our work about the need to embed CEA in a broader decision-making structure that is sensitive to the benefits and disadvantages of treatments that do not feature in the outcomes of clinical trials, as well as the ethical dimensions that are always inherent in any priority-setting process.

2. Cost-effectiveness threshold range for value-based price benchmark recommendations. ICER recognizes the variety of academic and conceptual work over the years that has explored methods for establishing cost-effectiveness thresholds.[13] There are two basic theoretical approaches to determining cost-effectiveness thresholds: 1) demand-side, or willingness to pay (WTP), and 2) supply-side, or opportunity cost.

Ryen and Svensson reviewed the literature on WTP for a QALY and found that results from studies based in the US differed by orders of magnitude, with the most recent (2010) estimate at approximately $60,000 per QALY.[14] Demand-side approaches have often focused on measures of per capita GDP, surveys of individual WTP, or revealed choices (e.g., estimates from job choices). Benchmarks for cost-effectiveness thresholds have been frequently justified by estimates of societal WTP, which, based on earlier consensus efforts at the World Health Organization (WHO), have commonly been cited as approximately 1-3 times the per capita GDP of the country per additional QALY.[15,16] However, Marseille et al. point out that thresholds based on per capita GDP have little theoretical basis, are too high to distinguish among most interventions, and are not likely to reflect affordability in many settings.[17] WHO itself has recently commented on the "misuse" of its earlier recommendations, and has argued that thresholds in this range are likely to prove unaffordable over the long-term.[18]

Attempts have also been made to use the value of statistical life (VSL) as a measure of societal WTP, especially in transportation and environmental assessments.[19] VSL estimates are based on evidence from market decisions such as wages for jobs with different risks of death, or on surveys that ask about similar risk-money tradeoffs. However, there are several important limitations of this approach.[19] Using VSL estimates in this way conflates WTP to avoid risk and willingness to accept risk, which may be quite different. In addition, using VSL as an estimate of WTP requires the assumption that VSL can be converted to calculate the value of a life year, but how to "spread" the VSL over life years remains unresolved. Using data on job choice to determine WTP also requires several strong assumptions about the fairness and rationality of the labor market, such as that workers have free choice of employment across jobs with different levels of risk. Lastly, the literature finds a wide range of estimates for VSL across different studies, with Hirth et al. reporting upper-bound estimates that were greater than 20 times the lowest estimate.[20]

Another suggestion as a basis for setting cost-effectiveness thresholds in the US has been to use prior funding decisions to benchmark WTP for future interventions. However, there is no certainty that previous funding choices were made with cost-effectiveness in mind. In addition, estimates of demand based on current funding may be distorted because health insurance is a tax-credited

employment benefit, meaning that health insurance coverage decisions do not necessarily match population preferences.

In an important recent conceptual contribution, Phelps[21] built on earlier work he had done with Garber[22] to look at how the optimal (i.e., utility-maximizing) threshold would vary with income and relative risk aversion. In this recent work, Phelps estimated optimal WTP by specifying utility as a function of income and using estimates of relative risk aversion – a measure of the rate at which marginal utility changes as income changes – to calibrate the function. This analysis assumed a Weibull utility function, which was parameterized to have declining absolute risk aversion (DARA) because the quantity of risky assets rises with wealth, and increasing relative risk aversion (IRRA) because the share of risky assets declines with wealth, as observed by Arrow.[23]

Results from these analyses confirm previous work suggesting that the optimal WTP threshold rises with income, as does the ratio of the threshold to income. That is, as income rises, trading off other goods and services for health care becomes less painful in terms of loss of utility and spending on health care should increase. Assuming an income of $50,000 and plausible values for other parameters, Phelps found that the optimal threshold was approximately two times income, or approximately $100,000-$110,000 per QALY if using the mean personal income in the US (approximately $54,000 in 2018).[24] Phelps notes that this work focuses on a representative, utility-maximizing individual, and expansion from this to decisions at the societal level may not be straightforward.

Phelps' approach to estimating WTP represents an important contribution, but WTP may be considered a more relevant approach to thinking about thresholds in a consumer sovereignty-based (i.e., welfarist) system. Value may vary by individual income and over time, and it is not clear whether WTP should be measured at the individual or household level. In addition, all WTP methods need to account for the mix of those who can afford to pay something and those who cannot, as a "median voter rule" for this mixed population would give a different answer than among those who can afford to pay some amount. Phelps has pointed out that a skewed income distribution means that the median voter model would almost always lead to lower thresholds than would be utility maximizing.[21] A central question in considering health economics is who captures the "value" of an intervention. Using a central measure of WTP, such as the median WTP, could lead to reduced access for those who have lower ability to pay. If an "average" WTP is selected, people with lower incomes may be forced to pay too much for health care to satisfy the WTP of the rich. Societal resources may be drawn into health spending from other domains of social spending that are much more important to people with lower incomes (such as public education). Some people with lower incomes are likely to be forced out of insurance markets all together.

In the US market-based system with multiple payers, there is a case for multiple thresholds based on WTP which may differ by payer type (e.g., government vs. commercial insurance).[25] However, there are broad requirements across the US health care system to fund all "medically necessary"

care. We also believe that there exists a widely accepted ethical goal in the US to have a common standard of care available for all patients, albeit with acknowledged differences in access due to network constraints, out-of-pocket payment, and other benefit design features. That the US does not yet achieve the goal of a common standard of care available for all patients does not imply, in our view, that ICER should abstain from framing a range of cost-effectiveness that should apply broadly across many, if not all, health insurance systems in the US.

Turning from the WTP approach, the other major paradigm for determining cost-effectiveness thresholds is a supply-side approach based on the idea that thresholds should reflect the opportunity cost of additional health care spending. Opportunity cost approaches based on health system outcomes and costs look at the trade-off between spending on a new intervention when that spending must come from curtailing current spending elsewhere in the health system on existing interventions, or from reducing spending on other social goods outside the health system, such as education or public safety. This approach has its strongest theoretical foundation in situations where the health system budget can be considered fixed. In such cases, the threshold can be considered as reflecting the point at which a higher price for a new intervention will lead to more health being lost within the health care system than will be gained by the patients who will benefit from the new treatment.

The best recent evidence on opportunity cost suggests that the previous WHO-recommended ranges for cost-effectiveness of one to three times per capita GDP are too high.[26] Claxton has argued for a lower cost-effectiveness threshold in the UK, US, and other countries, given the marginal productivity of the respective health care systems.[26-28] For the US, Claxton estimates an opportunity cost threshold of approximately $30,000-$40,000 per QALY.[28]

More recently, there has been a seminal attempt to ground an opportunity cost analysis directly from US data. In this work, Vanness has estimated health opportunity costs for private plans in the US.[29] Taking account of the effect of premium changes on coverage and the morbidity and mortality effects of loss of coverage, Vanness estimated the negative QALY impacts that result in the US health care system with rising costs and premiums. His research produces an estimate of $84,000 per QALY as the threshold. Working within this paradigm, this means that any new intervention introduced at a price that leads to an incremental cost-effectiveness ratio greater than $84,000 per QALY produces a net loss of health due to its impact on premium increases and thereby loss of insurance, especially among poorer members of the insurance pool. Vanness's work does not capture the potential impact of rising premiums on increasing deductibles and other out-of-pocket requirements that can lead to delayed or foregone care, nor does it capture the impact that rising premiums have on suppressing spending on other workplace benefits and wages. In some ways, therefore, it could be considered an upper-bound estimate of a threshold at which greater net losses occur despite the introduction of a treatment that will benefit those patients who can obtain it.

Which approach – WTP or opportunity cost -- should ICER take in its determination of the cost-effectiveness thresholds we use when presenting value-based price benchmarks to inform decision-making? For several reasons, we believe the opportunity cost is the strongest theoretical foundation. Despite the lack of an explicit overall budget for health care in the US, we believe the current environment of the US health care system indicates that we have reached a point where policymakers are no longer willing to accept cost increases in the US health care system that outpace growth in the overall economy. We hear this continuously from employers and many unions and other plan sponsors who are trying to maintain health benefits for their members. We hear this in broader concerns from consumer groups such as FamiliesUSA and AARP, who are aware of the opportunity costs faced by the public due to increasing health care costs. We hear it repeatedly from representatives of state government and state Medicaid programs, where rising health care costs have stripped out state spending on other needs such as education, police, and public infrastructure. And we also view the goals of several state laws as indicative. Maryland has a long-standing arrangement that limits hospital cost growth to the growth rate estimated for the state's overall economy.[30] Massachusetts already links policy actions to growth in health care costs that outstrip growth in the state per capita GDP; and recent initiatives may extend state oversight to prescription drugs as well.[31]

Overall, therefore, we believe that ICER functions in a system where health expenditure may continue to grow, but that it has reached the point at which policymakers sense that the opportunity cost for current spending is already substantial. This implies that an opportunity cost paradigm is justifiable as the predominant theoretical foundation for our cost-effectiveness thresholds. We believe that the opportunity costs are real, both within the health system and beyond, and that our goal should be to recommend value-based prices that will ensure that new interventions are adopted at a price that leads to a net increase in health over the entire population. It is not a matter of saving money; it is a commitment to improving health.

Following this line of reasoning, and reflecting on the most recent conceptual and empirical research, we have contemplated reducing our value-based price benchmark range to $50,000-$100,000 per QALY. We note, however, that the top end of our price benchmark range is usually interpreted as a "ceiling" price beyond which a treatment will be viewed as not cost-effective. We are aware that the opportunity cost empirical data for the US need formal peer review and further delineation. It is reassuring that the most recent highly respected work using the WTP paradigm for determining thresholds arrived at a very similar approximate result: $100,000 per QALY. And we believe there is some value in ICER retaining a consistent threshold range as a level playing field for all stakeholders. Therefore, for all the above reasons we are proposing to retain our current cost-effectiveness range to support our value-based price benchmark recommendations. We recognize that single cost-effectiveness thresholds should not be used as a blunt decision rule, and that decision-makers may want to consider different thresholds given their own view of their

opportunity costs and their interpretation of a treatment's potential other benefits and contextual considerations.

3.4 Base-Case Perspective in Economic Models

Proposed Change

1. Base-Case Perspective. No changes proposed – see discussion section.

Discussion

1. Base-Case Perspective. We received several comments urging ICER to use the societal perspective in the base-case analysis instead of the health care sector perspective.

The Second Panel on Cost-Effectiveness in Health and Medicine recommends reporting results from both the health system perspective and the societal perspective, with an "impact inventory" used to make transparent which elements of a full societal perspective are included.[9] ICER does provide results from both perspectives but chooses to use the health system perspective as the basis for its primary base-case results. The reasons for this are both conceptual and practical. Most importantly, we believe that our reports are primarily intended to inform population-based medical policy and pricing decisions within the US health care system. Decision-makers in the US health care system are not responsible for making trade-off decisions that involve broader societal resources. Of course, decision-makers may wish to consider the influence of health care on societal factors such as worker productivity, educational outcomes and spending, correctional system spending, tax revenues and payouts from Social Security. Our modified societal perspective tends to be able to model productivity effects but occasionally can include other factors when there are data or sources for reasonable assumptions. But the primary frame of reference for those entities involved in coverage and pricing policy is the health system. This is a feature of health technology assessment at the national level across most developed nations and is one reason that ICER and nearly all international HTA agencies use the health system perspective as that taken for the reference case for cost-effectiveness modeling.[32]

A second important reason that ICER prefers the health system perspective for its base case is the risk for discrimination against the disabled and elderly when a true societal perspective is taken in economic modeling. Giving "extra credit" to treatments of younger, working-age adults over patients who may never work again does not reflect the ethical principles that guide ICER's work. We understand that for some health care interventions there may be important value in the broader effects of treatment on productivity, both for patients and their families. But to hard-wire this consideration into lower price recommendations for treatments of elderly or disabled patients seems unreasonable in our view. As per our methods adaptations for treatments of ultra-rare diseases, however, when the societal costs of care for any disease are large relative to the direct

health care costs, the societal perspective will be included as a co-base case, presented directly alongside the health care sector perspective analysis.

3.5 Discounting

Proposed Change

1. Discounting. No changes proposed – see discussion section.

Discussion

1. Discounting. We received public comments suggesting lowering the discount rates or even removing the use of discount rates from our analyses entirely. Commenters were concerned that the time divergence between short-term costs and long-term health benefits could result in an unfair judgment in certain cases, such as in the evaluation of curative therapies. There was also concern that discounting of benefits prioritizes the needs and health of current generations over those in the future.

Discounting is a standard method in economic modeling, although the choice of the discounting rate and whether costs and benefits should be discounted uniformly or in some differential way are matters of debate.[33,34] In the US, the standard approach has been recently confirmed by the Second Panel on Cost-Effectiveness in Health and Medicine as a uniform discount rate of 3% applied to both costs and benefits.[35,36] Other countries may use a different discount rate, ranging somewhere between 1.5% and 5%, but most, including the UK and Canada, also use a single discount rate for both costs and effects.[33]

The use of a 3% discount rate in the US as standard for both costs and outcomes is based on estimates of the real consumption rate of interest and data on real economic growth, which are thought to reflect the social rate of time preference. While some have criticized the use of the 3% discount rate or discounting itself, we have made the judgment that there is no persuasive evidence for the use of another rate or scheme at this time. The use of a single, uniform discount rate for all assessments will allow for consistent comparisons across different or prior evaluations. We also do not propose presenting sensitivity analyses that vary the discount rate, as we do not believe this would provide additional information that is useful to decision-makers in this context. ICER encourages continued research into the appropriate discount rate to use for health economic evaluations, as well as periodic updates of the appropriate discount rate, as necessary.

3.6 Alternative Economic Model Assumptions

Proposed Change

1. ICER will add a "Controversies and Uncertainties" section to the cost-effectiveness section of its reports in order to broaden discussion of alternative model structures and assumptions suggested by manufacturers or other stakeholders.

Discussion

1. "Controversies and Uncertainties" Sub-Section. We received comments urging greater model transparency through public release of fully executable models and by providing additional details about the rationale behind the judgments that underpin the base case. Other comments recommended that ICER consider the acceptance of manufacturer-developed models and additional opportunities for input into model development for interested stakeholders.

The new proposed sub-section on "Controversies and Uncertainties" will allow exploration of different model variations that could be viewed as more conservative or optimistic. In particular, this sub-section will expand discussion of any alternative model structures or inputs suggested by manufacturers or other stakeholders that differ importantly from the base case. Although the current layout of ICER reports includes information on these issues, we feel it will be helpful to consolidate and expand discussion of factors related to uncertainty, including lack of information on natural history, limitations of the data on patient outcomes, difficulties translating existing data into measures of quality of life, and disagreements over the plausibility of certain inputs or assumptions.

Summaries of relevant published cost-effectiveness analyses will also be moved to this sub-section, pointing out differences in model structure, inputs and assumptions, and the impact of these differences on model results. This sub-section will allow for the acknowledgment of uncertainties and controversies raised by various stakeholders, while lending greater transparency to the rationale behind methodological decisions that underpin the base case. This new section will serve as an avenue to discuss how different assumptions or scenarios might affect model results and as a useful tool for decision-makers to understand the issues and uncertainties that may remain controversial.

3.7 Other Changes

Proposed Changes

1. ICER will exclude unrelated costs in some cost-effectiveness analyses.

2. When relevant, ICER will seek information from manufacturers and payers with which to model as a scenario analysis a limited number of outcome-based payment arrangements for the intervention under review.

3. Sources of Evidence. No changes proposed – see discussion section.

4. Caregiver Utilities and Costs. No changes proposed – see discussion section.

5. Dynamic Pricing. No changes proposed – see discussion section.

6. Subgroup Analyses. No changes proposed – see discussion section.

7. Public Payer Perspective Incorporating Behavioral Health Outcomes. No changes proposed – see discussion section.

8. Reference Case. ICER's Reference Case will be revised to reflect any of the proposed revisions that are adopted.

Discussion

1. Excluding unrelated costs. In cases where an intervention that increases QALYs would not be found to be cost effective, even with a zero-dollar price, we will exclude unrelated (non-drug) health care costs as a separate scenario analysis. Even though it may be controversial to treat such costs as unrelated, we believe it is still important to explore the effect of excluding these costs from the analysis especially when the disease already has very high health care costs.

We have encountered specific situations in assessments where the cost-effectiveness analysis is not able to produce a non-negative threshold price that would make a given treatment cost-effective. In addition, we have received comments during specific assessments that have suggested excluding unrelated costs in scenario analysis.

In some cases, there are no positive prices for an intervention that will reach specific cost-effectiveness thresholds. This may occur in situations where a new treatment is added on to existing treatment that is already near or beyond the cost-effectiveness threshold. One option in such cases would be to re-price the entire regimen, including the older, existing treatments, rather than just the new intervention, but this would not generally be a real-world option for regimens with multiple manufacturers. Another example where this may occur is when a new treatment results in more time spent in health states that have very high costs and/or a low utility value, making it impossible for the incremental cost effectiveness ratio to reach specific thresholds even at zero price.[37] In such cases a scenario analysis excluding health state costs that are not related to the intervention *per se,* may be informative.

2. Payment Models. We received comments asking that our analyses include the potential impact of outcome-based contracts on the cost-effectiveness of treatments. We agree that these contracts can be a useful tool in managing uncertainty and increasing the ultimate cost-effectiveness of treatment. We will actively seek information from manufacturers and payers about the potential outline of outcomes-based contracts for scenario analyses in our reports. It will only be helpful to run these kinds of scenario analyses if the list price of the treatment is known. If we do know the list price but do not receive any guidance from stakeholders, we may do an exploratory scenario analysis using outcomes and levels of financial risk-sharing that could meet our cost-effectiveness range.

3. Sources of Evidence. ICER received comments from multiple stakeholders recommending that we develop and utilize standard methods for incorporating RWE into our analyses, such as claims databases, electronic records, and registry data.

ICER has used and commits to continue using RWE provided the data are considered to be fit for purpose and of high quality, as judged by ICER's evidence review team. For example, ICER assessments have used analyses of commercial payer and Medicaid claims data to estimate costs for stem cell transplantation in an analysis of CAR-T treatments,[38] and to provide more current estimates of best supportive care costs for cystic fibrosis patients[39] than could be found in the literature. In the absence of high-quality randomized controlled trial data, ICER will rely on the highest quality RWE to provide critical inputs into our economic evaluations and context for the interpretation of both clinical effectiveness and cost effectiveness. However, in the absence of high quality RCT or RWE data, ICER will continue to report on the need for this data.

While RWE can reflect treatment effectiveness, adherence, and practice patterns seen outside a controlled trial setting, this type of evidence can also be fraught with confounding and bias and is highly dependent on study methodology. In addition, as we are often evaluating new health technologies that have not yet been launched in the market, high quality RWE may not exist. While some stakeholders have urged us to delay value assessments until after RWE has been generated, we strongly believe that such value assessments need to be conducted around the time of launch, to allow policymakers to make coverage and treatment decisions based on the best information available at the time.

4. Caregiver Utilities and Costs. We received several comments suggesting the inclusion of economic and utility impact on family members (caregiver spillover effects) in our economic evaluations. These caregiver effects include caregiver and/or family productivity loss, as well as quality of life impacts as a result of caregiving for patients.

The Second Panel on Cost-Effectiveness in Health and Medicine recommends including family and caregiver impacts in specific therapeutic areas where the introduction of a health technology is shown to alleviate such family/caregiver burden, leading to better overall health and economic

outcomes.[9] ICER has included caregiver and/or family economic burden when relevant and when appropriate data were available. However, we have only rarely included utility-specific caregiver/family effects, for several reasons.

While it may appear logical to include effects on caregiver utility, there continue to be many unresolved questions about whether and how to incorporate caregiver utilities. Key areas of uncertainty include the number of family members to include and how to account for changes in caregivers and their health-related quality of life over time.[40] Information is also needed on the stabilization or decrease of caregiving burden over time as caregivers become accustomed,[41] as well as on the magnitude and duration of change in caregiver utility following changes in health status such as the cure or death of a patient. We encourage future research on caregiver effects to address these areas of uncertainty. As research continues, we will consider scenario analyses that include the utility impact to patients' families and/or caregivers when compelling data exist. In analyses using a modified societal perspective, ICER will continue to include economic impacts on caregivers and family when published or grey literature data on productivity and other indirect costs are available.

5. Dynamic Pricing. We received public comments recommending the adoption of dynamic pricing for drugs and other health care costs in our economic evaluations to account for relative changes in the cost of providing health care over time, such as a decrease in the price of a drug following loss of exclusivity.

Questions have been raised as to whether ICER's cost-effectiveness analyses should account for changes in pricing over time. The topic of drug price changes is often raised in the context of anticipated loss of exclusivity for one or more drugs, or the anticipated introduction of biosimilars. Standard practice in cost-effectiveness analysis is to use current prices throughout an analysis, and there is at present no well-developed methodology for computing cost-effectiveness measures of health care interventions throughout their life cycle. Limited work has been done in this area,[42,43] but the results may not be generalizable to other therapy areas, health care settings, or geographies. In addition, analyses using a health care sector perspective and static pricing are more consistent with an opportunity cost paradigm as the foundation for cost-effectiveness analysis and decision-making. For health care decision-makers considering cost-effectiveness at the margin, decisions should theoretically be driven by the opportunity cost of existing services, making price changes in the future less relevant.

Attempts to model price changes over time would add an additional layer of uncertainty and speculation to cost-effectiveness analyses. In the US market, where drug prices are mostly unregulated, changes in prices occur relatively frequently and are difficult to predict. Prices for specific branded drugs may decrease over time, especially as competing drugs come to market, but also often increase over time, sometimes repeatedly. The entry of other branded competitors in the future can be difficult to predict, as FDA approval of anticipated new drugs may be delayed or

denied. Price increases may also occur in anticipation of loss of exclusivity. Generic drugs and biosimilars are expected to have discounted pricing relative to branded or bio-original competitors, but the size of that discount may be difficult to estimate, especially if it occurs years in the future. It also may be difficult to predict the timing of market entry for generic drugs or biosimilars, due to the possibility of patent litigation or other barriers to entry. Finally, even products with historically stable pricing may be sold to or acquired by another manufacturer, who could decide to change pricing in dramatic and unpredictable fashion.

6. Subgroup Analyses. We received several comments recommending the inclusion of different patient subgroups seen in the real-world when analyzing cost-effectiveness of different health technologies.

ICER clarifies that we have and will continue to include analysis of patient subgroups when robust data and relevant inputs from clinical trials and/or real-world evidence are available to do so. While cohort models tend to reflect homogeneity in patient populations for whom health technologies are assessed, we include scenarios with different patient subgroups to account for the heterogeneity within patient groups within a specific disease area.

As an example, in the 2017 ICER review of targeted immunomodulators for the treatment of moderate-to-severely active rheumatoid arthritis, we included not only those patients in whom conventional disease modifying agents failed, but also those in whom such conventional therapies were not well-tolerated, as well as those who were naïve to such therapies, to align with treatment practice patterns in the real world. Such subgroup analyses have been and will continue to be undertaken when ICER believes that health technologies are likely to be approved or have been used extensively within these subgroups of interest, and as mentioned earlier, pending data availability.

7. Public Payer Perspective Incorporating Behavioral Health Outcomes. We received comments urging ICER to consider the potential impact of changes in behavioral health outcomes on income levels and eligibility for means-tested public programs. Public payers may have a very different perspective on the cost-effectiveness of treatments that alleviate poverty or disability, thereby allowing patients to move from public programs to commercial insurance.

While ICER acknowledges that different payer types may have different perspectives, we believe it is important for policymakers not to view health investment as less worthwhile if the return on investment is realized by a different (type of) payer. ICER's economic evaluations will therefore continue to be conducted using a broad health care sector perspective, with a societal perspective as a scenario analysis. ICER may consider payer-specific analyses where considered particularly relevant and when data are available.

8. Reference Case. To ensure the comparability and transparency of ICER's economic analyses, Reference Case specifications will be updated to reflect the most currently recommended methods.

4. Potential Other Benefits or Disadvantages and Contextual Considerations

4.1 List of Voting Questions and Voting Format

Proposed Changes

As displayed in Table 4.1 below, ICER proposes the following changes:

1. ICER will change the wording of all questions related to potential other benefits and contextual considerations to improve clarity and consistency of interpretation.

2. ICER will add a first question related to whether appraisal committee members believe that uncertainty and/or model assumptions lead the base-case cost-effectiveness results to be overly optimistic or pessimistic.

3. ICER will add several new potential other benefits of a new intervention compared to the selected comparator:
    a. For interventions that may offer special advantages by virtue of presenting an option to patients with a notably different balance or timing of risks and benefits versus other treatments.
    b. For interventions that have a delivery mechanism or complexity of regimen that may improve or decrease real-world adherence relative to comparator treatments

4. ICER will provide empirical results for absolute QALY shortfall and proportional QALY shortfall to support deliberation and voting on a single question on relative "health loss" as a contextual consideration. This question will take the place of two separate questions on severity of illness and lifetime burden of illness.

5. ICER will add one new potential disadvantage related to treatments that, if not entirely curative, could reduce or even preclude the potential effectiveness of future treatments.

6. ICER will change the voting structure for all questions from a yes/no format to a Likert scale from 1-3. The intent of the new voting structure is to enhance the application of these considerations by decision-makers within a cost-effectiveness range suggested by the base-case economic model.

Table 4.1 Proposed List of Voting Questions on Potential Other Benefits/Disadvantages and Contextual Considerations.

1	Intermediate (2)	3
Uncertainty or overly favorable model assumptions creates significant risk that base-case cost-effectiveness estimates are too optimistic		Uncertainty or overly unfavorable model assumptions creates significant risk that base-case cost-effectiveness estimates are too pessimistic
Very similar mechanism of action to that of other active treatments		New mechanism of action compared to that of other active treatments
Delivery mechanism or relative complexity of regimen likely to lead to much lower real-world adherence relative to the comparator		Delivery mechanism or relative complexity of regimen likely to result in much higher real-world adherence relative to the comparator
The intervention offers no special advantages to patients by virtue of presenting an option to patients with a notably different balance or timing of risks and benefits		The intervention offers special advantages to patients by virtue of presenting an option to patients with a notably different balance or timing of risks and benefits
Small health loss without this treatment as measured by proportional and/or absolute QALY shortfalls		Substantial health loss without this treatment as measured by proportional and/or absolute QALY shortfalls
Will not significantly reduce caregiver or broader family burden versus the comparator		Will significantly reduce caregiver or broader family burden versus the comparator
Will not have a significant impact on improving return to work and/or overall productivity versus the comparator		Will have a significant impact on improving return to work and/or overall productivity versus the comparator
Other		Other

Discussion

1. ICER will change the wording of all questions related to potential other benefits and contextual considerations to improve clarity and consistency of interpretation.

The current list of potential other benefits and contextual considerations was formally put into place with the adoption of the value assessment framework methods update in 2017. In the subsequent two years our experience has shown that some of the questions were difficult for the appraisal committees to interpret in the context of specific topics under review. We are therefore proposing to adapt most of the existing concepts represented in the current list of potential other benefits and contextual considerations into newly worded questions framed as a Likert scale between two ends of a spectrum. We have done preliminary pilot testing of this approach with our appraisal committees and we believe this framing of the questions will prove to be more a consistent and transparent guide to these issues than the current format.

2. ICER will add a first question related to whether appraisal committee members believe that uncertainty and/or model assumptions lead the base-case cost-effectiveness results to be overly optimistic or pessimistic.

The deliberation on the cost-effectiveness model is an important part of the public meetings of our independent appraisal committees. We believe it will aid decision-makers if we initiate a new question specifically related to whether they believe the model structure, assumptions, and relative level of uncertainty, makes it likely the base-case results are too pessimistic or too optimistic. This vote should help provide greater transparency and guidance to decision-makers seeking to apply the base-case results to medical policy.

3. ICER will add several new potential other benefits of a new intervention compared to the selected comparator:

a. For interventions that may offer special advantages by virtue of presenting an option to patients with a notably different balance or timing of risks and benefits versus other treatments.
b. For interventions that have a delivery mechanism or complexity of regimen that may improve or decrease real-world adherence relative to the comparator

As we have also mentioned in our proposed methods adaptations for single or short-term transformative therapies, we believe that the concept of "value of hope" is poorly named to convey the advantages that some treatments may offer if they have a distinctly different timing or balance of risks and benefits compared to other available treatments. The classic example is a treatment for cancer that may have, overall, the same total QALYs gained as existing options, but which has a higher risk of short-term death and a higher chance of longer-term survival. For risk-taking patients this treatment option, although its QALYs are identical to other options, offers a special advantage, and so we think this potential other benefit merits consideration given the heterogeneity of patients and the way they view the relative balance of risks and benefits of different treatment options.

We also note that there are some treatments that may, through stimulation of antibodies or other clinical effects, decrease the chance of benefit from future treatment options. Although this is infrequent, we feel it merits a place in the voting list.

4. ICER will provide empirical results for absolute QALY shortfall and proportional QALY shortfall to support deliberation and voting on a single question on "health loss without this treatment" as a contextual consideration. This question will take the place of two separate questions on severity of illness and lifetime burden of illness.

Ethicists, health economists, and health technology assessment groups have long recognized that pure QALY maximization does not incorporate all the values that societies wish to consider when making prioritization decisions for health care spending. One important social value is that which gives some preference to treatments for patients with more severe conditions.[8,44] "Severity of illness" has therefore been proposed as one element of value that should modulate applications of cost-effectiveness results to medical policy, but countries and health technology assessment groups have conceptualized this idea somewhat differently. Some have seen that giving some priority to treatments according to "lifetime burden of illness" or "need" may better represent the ethical instincts of a society or other decision-makers.[45,46]

Our current methods have asked appraisal committee members to vote separately on severity of illness and lifetime burden of illness without providing any specific conceptual or empirical guidance. We believe that we can gain greater clarity and consistency in consideration of these issues by changing the terms used and by providing empirical results for the absolute QALY shortfall and proportional QALY shortfall.

The absolute QALY shortfall is defined as the total absolute amount of future health patients with a condition are expected to lose without the treatment that is being assessed.[47] It can be measured over the entire lifetime of patients with a condition, but more often it is measured from the point at which patients are diagnosed with a condition. By capturing the magnitude of the number of QALYs lost, the absolute QALY shortfall reflects the aspect of severity of illness related to the idea that treatments for people who stand to lose the most absolute numbers of QALYs should merit some increased prioritization. The ethical consequences of using absolute QALY shortfall to prioritize treatments is that conditions that cause early death or that have very serious lifelong effects on quality of life receive the greatest prioritization. Thus, certain kinds of treatments, such as treatments for rapidly fatal conditions of children, or for lifelong disabling conditions, score highest on the scale of absolute QALY shortfall. The Norwegian health technology assessment program is perhaps the most notable organization currently using measures of absolute QALY shortfall as a component in their appraisal process.[48]

Absolute QALY shortfall is often viewed in contrast to another way to empirically measure a sense of severity of illness, or "need" as the Dutch have called it.[49] This alternative measure is called proportional QALY shortfall. The proportional QALY shortfall is measured by calculating the proportion of the total QALYs of remaining life expectancy that would be lost due to untreated illness.[50] The proportional QALY shortfall reflects the ethical instinct to prioritize treatments for patients whose illness would rob them of a large percentage of their expected remaining lifetime. As with absolute QALY shortfall, rapidly fatal conditions of childhood have high proportional QALY shortfalls, but the highest numbers can also often arise from severe conditions among the elderly who may have only a few years left of average life expectancy but would lose much of that to the illness without treatment.

Absolute QALY shortfall and proportional QALY shortfall are therefore empirical measurements that capture different aspects of society's instincts for prioritization related to the severity or burden of an illness. Because they can be viewed as complementary in some ways, we propose to calculate both measures for every intervention. We will include these results in our reports and highlight them when asking our independent appraisal committees to vote on relative health loss. In order to provide some anchoring to the deliberation, we will also present league tables of absolute and proportional QALY shortfalls from the academic literature.[51] We will also explore real-time use during meetings of a burden of disease calculator developed by Dutch investigators (see https://imta.shinyapps.io/iDBC/) that allows for calculation of absolute and proportional QALY shortfalls under different assumptions.

By changing the wording of this voting question and providing absolute and proportional QALY shortfall data, we believe we will be able to enhance the deliberation of our appraisal committees and, ultimately, improve the ability of decision-makers in the US health care system to integrate these important ethical dimensions in their decisions.

5. ICER will add one new potential disadvantage related to treatments that, if not entirely curative, could reduce or even preclude the potential effectiveness of future treatments.

In our discussions with patient groups we have learned that, on occasion, patients and clinicians must factor into their decision-making whether a treatment option may carry the risk of reducing the effectiveness of future treatment. Whether through the stimulation of antibodies to treatment vectors or other factors, this potential disadvantage seems important enough to warrant a position on our voting list.

6. ICER will change the voting structure for all questions from a yes/no format to a Likert scale from 1-3. The intent of the new voting structure is to enhance the application of these considerations by decision-makers within a cost-effectiveness range suggested by the base-case economic model. ICER will not adopt a formal multi-criteria decision analytic approach but retain this modified approach to integrating other factors into deliberation and decision making.

The current voting format was designed largely as a series of yes/no questions to whether a particular potential other benefit or contextual consideration was a "significant" factor in judgments of long-term value for money of the intervention under review. Our experience has been that this voting structure was superior to the less formal deliberative process we had used prior to 2017. Having votes on each item improved transparency and also served as a more explicit signal to decision-makers about how the appraisal committee viewed each individual item, helping to emphasize that potential other benefits and contextual considerations should always be considered in applying the results of cost-effectiveness analysis to medical policy.

We have also learned in the past two years that the dichotomous nature of the voting questions, often hinging on interpretation of the key word "significant," resulted in some cases in which the judgments of the appraisal committees were hard to interpret. It has also been clear that some of the voting questions were tailored to capture features of treatments that were only infrequently relevant to the topic at hand.

We are therefore proposing to move to a three-item Likert scale voting format. We feel this will help provide the appraisal committees with a clearer understanding of the ends of the spectrum within which they are expected to vote. We also think that a Likert scale approach will provide a more transparent record of how the appraisal committee feels that these considerations should be applied when integrated with the cost-effectiveness results in making decisions about pricing. It has always been our intention to use these votes as a way to signal to decision-makers that the "right" cost-effectiveness threshold to be applied in any individual situation should be a judgment that benefits from integration of cost-effectiveness results with an intervention's potential other benefits (or disadvantages) and broader contextual considerations that include ethical dimensions of priority setting. We believe that a Likert scale voting format will provide not only a record of individual votes but also an average score that will be a more powerful and transparent signal on the relative importance of broader factors that should guide decision-makers in applying the cost-effectiveness results.

We received multiple public comments that recommended that we quantitatively measure other aspects of value, including both health and non-health benefits such as value of hope, reduced uncertainty, insurance value, and achievement of public health goals, rather than only qualitatively incorporating them as potential other benefits or contextual criteria. Most of the commenters were concerned that the QALY alone does not adequately represent the other benefits and advantages associated with the intervention of interest. They noted that "other benefits" can be substantial even if the cost per QALY is very high. Some suggested the use of multi-criteria decision analysis (MCDA) as an alternative to traditional CEA. Some also hoped that applying MCDA would allow individual users of ICER's reports to assign weights to different elements of value and arrive at their own estimate of a treatment's value.

MCDA offers a framework that can capture a wider range of objectives, offer flexibility in the way trade-offs are made between competing objectives, and allow larger public participation in determining these trade-offs.[52] Proponents argue that it has the potential to make value assessments more customizable, transparent, and comprehensive, while incorporating other elements of value that patients care about beyond the QALY. In the MCDA approach, various qualitative measures are weighted and can be translated into one metric that allows for a comparison of different interventions. The weights are based on value judgments and assumptions. In other words, those weights depend on the priorities of the decision-maker.[53] However, the quality of MCDA is dependent on these weights and assumptions, and it may be difficult to determine these in a practical and consistent manner.

In 2009-2010 ICER attempted on several occasions to use a formal MCDA process in its appraisal committee deliberations. We found, as have others, that it was very difficult for participants to identify mutually independent factors in their decision-making, much less to give weights to them. We continue to monitor the academic and policy work in this field but do not feel that MCDA, given its procedural and conceptual limitations, offers advantages to our modified approach in which factors are voted upon but not weighted.

5. Potential Budget Impact Analysis

Proposed Changes

We received several general comments on the potential budget impact analysis, ranging from recommendations to exclude it entirely from our reviews and to use it as a more primary economic evaluation in all assessments. Below we detail several proposed changes.

1. ICER will extend the time period over which we average the annual number of drugs approved by the FDA from two years to five years.

   ICER recalculates the potential budget impact threshold each calendar year, using the most recent inputs available. In the recalculation of ICER's potential budget impact threshold for calendar year 2019, we have now extended the time period over which we average the annual number of drugs approved by the FDA from two to five years, to reduce fluctuations in the threshold due to this variable. See Table 5.1 for the updated calculations used to derive the threshold for 2019.

Table 5.1. Potential Budget Impact Threshold Calculations

Item	Parameter	Estimate	Source
1	Growth in US GDP, 2019 (est.) +1%	3.5%	World Bank, 2019
2	Total personal medical health care spending, 2018	$2.95 trillion	CMS National Health Expenditure, 2019
3	Contribution of drug spending to total health care spending (%) (Row 4 ÷ Row 2)	16.9%	Calculation
4	Contribution of drug spending to total health care spending, 2018	$498.6 billion	CMS National Health Expenditures, 2019 Altarum Institute, 2018
5	Annual threshold for net health care cost growth for ALL drugs (Row 1 x Row 4)	$17.4 billion	Calculation
6	Average annual number of new molecular entity approvals, 2014-2018	42.6	FDA, 2019
7	Annual threshold for average cost growth per individual new molecular entity (Row 5 ÷ Row 6)	$409.6 million	Calculation
8	Annual threshold for estimated potential budget impact for each individual new molecular entity (Doubling of Row 7)	$819 million	Calculation

2. ICER will add the following new language to our economic reference case providing greater detail regarding our methods of potential budget impact analysis:

"ICER uses the cost-effectiveness model in an economic evaluation to estimate the potential total budgetary impact of new treatments in the US, assuming different prices, including the treatment's list and net prices, and the three threshold prices to achieve cost effectiveness at $50,000, $100,000, and $150,000 per QALY. Results from the cost-effectiveness model are used to provide undiscounted net costs (including intervention/comparator costs, other health care costs, and total costs) broken out by year for years one through five, for use in the potential budget impact analyses. Potential budget impact is defined as the total differential cost of using each new therapy rather than relevant existing therapy for the treated population, calculated as differential health care costs (including drug costs) minus any offsets in these costs from averted health care events.

Potential budget impact analyses are based on net cost per patient and estimates of the proportion of the US population eligible for treatment with the new intervention. ICER uses epidemiologic and other data to estimate the size of the potential candidate population for each new treatment. We then assume that an equal proportion of patients (20%) would be treated with the new treatment each year over five years, arriving at a cumulative 100% uptake at five years.

The goal of ICER's potential budget impact analysis is to estimate the net cost per patient treated with new interventions so that decision-makers can use their own assumptions about uptake and pricing to determine their own estimate of potential budget impact. We also seek to produce calculations that will help policy makers identify situations in which the potential uptake of a new treatment, at various pricing levels, might exceed a budget impact threshold that signifies that the budget impact in the near term (over 5 years) would contribute to overall health care cost growth at a higher rate than growth in the national economy (plus 1%).

To accomplish these goals, ICER's potential budget impact analyses must evaluate whether a new drug would be likely to take market share from one or more drugs. ICER will continue to use clinical expert opinion regarding the treatments likely to be displaced by use of a new treatment within the eligible population. ICER will then follow one of the procedures listed below, dependent on whether existing treatments are being displaced. These are explicitly NOT meant to represent our assumptions of the budget impact of new interventions that are most likely in the real world. Our methods are intended to provide the calculations that can underpin a graphic figure that allows decision-makers and policy makers to make their own assumptions.

- No existing active treatment: If the intervention is for a condition which has no existing active treatment in the market (other than best supportive care), we will calculate

potential budget impact for 100% of the eligible population at the end of five years (20% marginal new uptake per year).

- Existing treatments launched within prior 2 years: If the intervention is for a condition with existing active treatment(s), one or more of which was launched within the last two years, equal proportions of the eligible population will be split among the intervention and the recently launched treatment(s), with 100% displacement of relevant treatments launched more than two years ago.
- Existing treatments all on market >2 years: If the intervention is for a condition with existing active treatment(s) all launched more than two years ago, we will calculate potential budget impact for 100% of the eligible population at the end of five years, with displacement of existing treatments.
- Multiple existing treatments: When there are multiple existing treatments on the market, clinical expert opinion will be used to estimate the percentage of patients converted from each existing treatment to the new treatment.
- Untreated patients: For all cases, we will include the untreated portion of the eligible population, as long as they are considered eligible for the new treatment.

3. ICER will present a cumulative per-patient potential budget impact. ICER will now present a cumulative per-patient potential budget impact for each year over the five-year time horizon, with results being presented graphically for each intervention assessed, and numerical data presented in tabular format in an appendix of the report. This graph will replace the prior tables that reported five-year annualized potential budget impact per patient.

Discussion

2. Treatments Potentially Displaced. ICER's potential budget impact analyses already follow the general procedures outlined in the language above, but the details of this process have not been publicly codified as part of our value framework. By adding these details to ICER's Reference Case specifications, we hope to provide greater clarity to users of our reports.

3. Per-Patient Potential Budget Impact. ICER's potential budget impact analyses currently include tables reporting the five-year annualized per-patient potential budget impact. However, the annualized per-patient potential budget impact as presented was dependent on the cohort sample size entering the potential budget impact model each year and was difficult to interpret as it could not be applied to individual patients. The new graph will allow readers to see the average potential budget impact for a single patient over various time horizons from one to five years. The new potential budget impact graph will help payers understand the estimated average net cost of treating a patient with an intervention relative to comparator(s) over the five years of the potential budget impact analysis. See Figure 5.1 and Table 5.2 for examples of this new approach.

Figure 5.1. Example Per-Patient Potential Budget Impact Figure (Spinraza for Presymptomatic Spinal Muscular Atrophy)

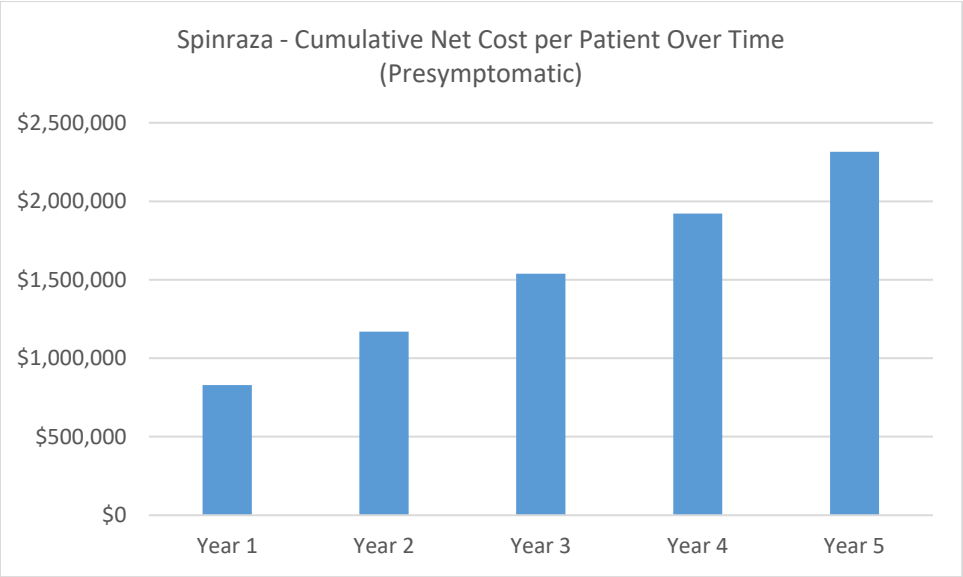


Table 5.2. Example Per-Patient Potential Budget Impact Table (Spinraza for Presymptomatic Spinal Muscular Atrophy)

	Cumulative Cost	Additional Costs Per Year (Non-Cumulative)
Year 1	$828,183	$828,183
Year 2	$1,170,133	$341,950
Year 3	$1,537,652	$367,519
Year 4	$1,922,075	$384,423
Year 5	$2,315,537	$393,462

6. Report Development and Public Meetings

6.1. Report Development

Proposed Changes

1. ICER will extend the timeline for large class reviews by nine weeks.

2. ICER will implement a formal process through which to reassess whether new evidence has emerged that should be included in an update to the report one year after the release of a Final Evidence Report.

3. ICER will make the following changes to public comment periods:
    a. Extend the draft report public comment period for class reviews by one week as part of the aforementioned timeline extension.
    b. Extend the word limit for written summaries of oral public comments included in the final report from 250 to 750 words.

4. ICER will create a new "Patient Perspectives" chapter for its reports that will describe the input we have received from patients, families, and patient organizations, as well as relevant sources of patient-generated evidence. We will also summarize relevant sources of patient-generated evidence that have been shared by patients and identified through our research process.

5. Methods Transparency. No changes – see discussion section.

6. Policy Guidance for Stakeholders. No changes – see discussion section.

Discussion

1. Review Timelines. ICER conducts reviews on a tight schedule in order to balance the timing of expected drug approvals with decision makers' needs for timely information to inform policy and practice, necessitating a rapid timeline. While our experience demonstrates that the standard eight-month timeline is appropriate for an average review, we believe that additional time is needed for large class reviews due to the larger evidence base and number of stakeholders involved in these assessments. As such, we propose to extend our standard timeline by nine weeks for large class reviews. Appendix Figures 1a and 1b describe ICER's standard review timeline and proposed modifications for large class reviews, respectively. Briefly, we propose to add time to 1) the scoping phase (one week), 2) the draft report phase (five weeks), 3) the draft report comment period (one week), 4) the Evidence Report drafting phase (one week), and 5) between the Evidence Report

posting and public meeting (one week). ICER will continue to provide stakeholders with timelines at the outset of each review so that stakeholders may plan their engagement accordingly.

2. Report Updates. As noted above, ICER aims to complete initial drug assessments near FDA decision dates whenever possible to ensure the information within each report is as timely as possible for stakeholders. We recognize, however, that the evidence base for new treatments may evolve rapidly in the months following market release and that this may cause our reports to become outdated. Our current practice, implemented as part of the previous framework revision cycle, is to update our assessments on an ad hoc basis when new evidence or treatments emerge that may meaningfully impact the conclusions of prior reviews (i.e., developments that would change clinical practice patterns, lead to different judgments regarding the net health benefit of treatment, that would substantially impact value-based prices, etc.). In addition, ICER includes a disclaimer at the beginning of each report noting that the findings are current as of its posting date.

Our experience since adopting the above approach suggests that stakeholders would benefit from a formal process to indicate whether report findings remain applicable or that new developments have occurred that could lead to different conclusions. As such, we propose to implement a process in which we will conduct a broad search for new developments in the treatment of the reviewed condition and for new evidence related to the included interventions. This review will be completed around the one-year anniversary of a final report and will be summarized in a public statement describing our rationale for why we will or will not update the assessment. We envision that this document may take one of three forms: 1) a statement that, given the magnitude and/or volume of new evidence, a full review be undertaken, 2) a brief narrative summary of the new evidence with a statement describing why ICER does not believe a full update is warranted, or 3) a statement that no new evidence is available and that the report remains current.

3. Public Comments. Due to the rapid timelines for ICER assessments, the length of comments and comment period must be limited to ensure that ICER staff has adequate time to review and incorporate suggestions. We reiterate our commitment to publicly posting review timelines, including public comment periods, at the beginning of each review so that stakeholders are able to plan for their engagement with us. We note that when submitting public comments, content such as data tables, figures, and reference lists may be included as an appendix that does not count toward the three- or five-page limit for draft scoping documents and draft reports, respectively, and that there are no page limits for the Open Input period that takes place during the first three weeks of a review.

However, as noted above, we recognize that large class reviews pose special challenges for stakeholders due to their length and complexity. Thus, as part of the timeline extension, we propose to add one week to the public comment period of draft reports for large class reviews.

In addition, ICER proposes to extend the word limit for written summaries of oral comments delivered during its public meetings from 250 to 750 words. This shift is intended to make it easier for commenters to submit summaries that capture the entirety of their remarks.

4. Patient Perspectives Chapter. ICER includes information on the patient perspective (i.e., input gathered through conversations with patients and patient organizations, summaries of existing literature on the patient experience and preferences, etc.) in the "Background" section of its reports. ICER recently expanded this section to include additional details about the methods used to gather patient input, how such input informed ICER's research, and to provide greater detail on the patient experience. Over the past year, several patient organizations recommended the creation of a separate chapter about patient perspectives, a suggestion that was echoed in several Open Input comments. We agree with this suggestion and propose to create a new section for this content that will follow the "Background" chapter and will precede the chapters on clinical effectiveness, cost-effectiveness, potential budget impact, and potential other benefits and disadvantages / contextual considerations. This sequence ensures that readers are presented with information on patient perspectives in the early pages of each assessment, allowing them to interpret the subsequent evidence and analyses through the lens of the patient experience.

5. Methods Transparency: It has long been ICER's practice to publicly release methods documentation related to its research and to update this documentation to reflect any modifications that occur during a review. This documentation includes draft and final scoping documents posted to the ICER website, research protocols and model analysis plans posted to the Open Science Framework website, research protocols registered with the PROSPERO database, modeling methods registered with the Tufts Center for the Evaluation of Value and Risk in Health (CEVR) registry of cost-effectiveness analyses. These practices meet or exceed established best practices and, as such, we do not believe any changes are warranted.

6. Policy Guidance for Stakeholders. A small number of Open Input comments requested that ICER provide guidance on how to interpret and apply the findings of each report. We reiterate that ICER has always included guidance on how to interpret results within each report version (draft, revised, and final), including discussion of the limitations of the evidence base and economic modeling. We believe it is important to reserve any policy recommendations for the Final Report so that stakeholders involved in the public meeting (patients, clinicians, manufacturers, and payers) may all participate in the development of these recommendations. As such, we do not believe any changes are necessary.

6.2. Public Meetings

Proposed Changes

1. Council Membership. No changes – see discussion section.

2. ICER will post annual COI disclosure statements to its website for each voting council.

3. ICER will adopt a code of conduct for public meetings.

Discussion

1. Council Membership. ICER voting councils are composed of a multidisciplinary set of practicing clinicians, health services researchers, and patient advocates. The councils are standing bodies (i.e., they do not change from one meeting to the next), and we seek members for their expertise in research methods, economic analysis, evidence-based practice, and patient advocacy, among other qualifications. All members meet strict conflict of interest requirements to limit any bias that may be introduced by the presence of certain personal or financial relationships. This means that, by design, ICER voting councils do not necessarily include those affected by the condition under review, whether they are individual patients or practicing clinicians, though this may occur from time to time (i.e., a neurologist may serve on a voting council for a neurology topic, provided he or she does not have any disqualifying conflicts). This approach aligns with that of many other organizations, including the United States Preventive Services Task Force (USPSTF) and all international HTA organizations.

ICER recognizes how vital the patient and clinical expert perspective is to our review process and public meeting, which is why we seek input from patient and clinical experts throughout the report development process, and by including several such experts as active participants as throughout our public meetings. We believe this approach provides members of ICER voting councils with sufficient insight into the patient experience and clinical practice, and do not propose any changes.

2. Conflict of Interest Statements. It has long been ICER's practice to include voting member's conflict of interest disclosures on the agenda for each public meeting, and for each member to confirm the lack of relevant conflicts at the beginning of each meeting. One commenter suggested that ICER post annual conflict of interest (COI) disclosure statements to its website from members of its voting councils. We agree with this suggestion and propose to adopt it to provide the public with greater confidence that voting members are free from undue conflict of interest.

3. Code of Conduct. ICER recently implemented a public meeting code of conduct to outline our expectations for all public meeting participants. This code is intended to facilitate respectful meetings that drive collaborative action from multiple stakeholder groups.

7. Stakeholder Engagement

7.1. Stakeholder Engagement

Changes

1. ICER will update the following patient engagement materials and approaches:
    a. Revise patient engagement materials to include examples of how patient input informed reviews.
    b. Revise the language of its patient input survey to include PICOTS language
    c. Continue to include suggestions that were adopted in the "Stakeholder Input" section of scoping documents, and will expand the section to include discussion of suggestions that were not adopted.

2. Economic Model Transparency. No changes – see discussion section.

3. ICER will formalize the practice of debriefing with patient groups after a review has concluded.

4. ICER will produce a series of lay-friendly seminars that will provide background on evidence-based medicine and its application to health technology assessment.

Discussion

1. Evidence Sought from Patients and Patient Advocates. Several patient organizations requested that ICER provide more detailed guidance on the types of evidence we seek from patients and patient organizations, and how that evidence has been used. We agree that such guidance is important to facilitate patients and patient groups' ability to effectively inform our research. Patient groups suggested several ways to provide this information, including by giving examples of valuable patient contributions to reviews and describing rationale for why suggestions were or were not incorporated. ICER's practice, which has been the same for many years, is to respond to draft report comments with this degree of detail and will continue to do so; scoping documents currently describe suggestions we have accepted under a "Stakeholder Input" heading, and we propose to include details of why some suggestions have not been adopted.

Commenters suggested that ICER seek patient input through individual patient interviews, focus groups, partnering with patient organizations to conduct surveys, and by requesting existing resources from patient groups. ICER already uses these approaches to gather patient input and will continue to do so.

Commenters also suggested that ICER solicit input from patients and patient groups about the PICOTS (Population, Intervention, Comparators, Outcomes, Timing, and Setting) framework that describes the research agenda for a given review. ICER seeks this information through calls with patients and patient groups during the Open Input and scoping periods of reports, through a patient input survey, and during the public comment period on draft scoping documents. We reaffirm our commitment to seek direct patient input on these elements of our research agenda. As part of a broader update to our engagement materials, we intend to revise the language of our patient input survey to directly reference PICOTS elements. We hope this will make it easier for stakeholders to track the impact of their feedback.

2. Economic Model Transparency. We received several public comments acknowledging our commitment to transparency while others requested greater transparency via access to fully executable models available to all interested stakeholders. ICER's process for conducting health technology assessments provides transparency in our methods to various stakeholders during each phase of a review. ICER presents preliminary methods, inputs, and assumptions for the clinical evidence review and economic modeling to manufacturers for feedback. Our evidence reports provide a detailed explanation of our economic models, methods, inputs, and assumptions. Additionally, involved manufacturers may obtain a working copy of the economic model for review prior to providing public comment on each draft report. This model sharing has been adopted to equip manufacturers with an in-depth knowledge of the methods used in our economic evaluation, so that they can provide more focused and robust comments on our economic modeling efforts.

3. Patient Group Debriefs. One commenter suggested that ICER hold debriefing sessions with patient groups after the conclusion of each review. Although the commenter suggested that ICER use these discussions to provide more insight into how submitted data were or were not useful, we believe a more transparent way to do so is through the first change proposed in this section. ICER began piloting a similar series of debriefing meetings early in 2019 to gather feedback on our processes could be improved. We thus propose to formalize these debriefs as part of our updated engagement process. The conversations we have held thus far have yielded valuable feedback on how ICER's processes and engagement materials can be improved to better facilitate patient engagement.

4. Methods Seminars. As part of ICER's commitment to facilitating effective stakeholder engagement, we propose to create a series of webinars that will describe the principles of health technology assessment and economic modeling for a lay audience. While we have yet to determine the specific content of these webinars, potential topics include an overview of the strengths and limitations of certain types of clinical evidence, an overview of health economic modeling concepts (e.g., the QALY, evLYG, health system vs. societal perspectives, willingness-to-pay thresholds), and how ICER combines these techniques in its reviews. ICER encourages patient groups and other stakeholders to provide suggestions as to which topics to include in these seminars.

References

1. Stern A, Pietrulla F, Herr A, Kesselheim A, Sarpatwari A. The Impact Of Price Regulation On The Availability Of New Drugs In Germany. *Health Affairs.* 2019;38(7):1182-1187.
2. Robinson J, Panteli D, Ex P. Reference pricing in Germany: implications for U.S. pharmaceutical purchasing. The Commonwealth Fund, 2019. https://www.commonwealthfund.org/publications/issue-briefs/2019/jan/reference-pricing-germany-implications. Accessed August 15, 2019.
3. Institute for Quality and Efficiency in Health Care (IQWiG). General Methods: Version 5.0. July 10, 2017; https://www.iqwig.de/en/methods/methods-paper.3020.html. Accessed August 15, 2019.
4. Tice J, Walsh J, Whittington M, et al. Chimeric Antigen Receptor T-Cell Therapy for B-Cell Cancers: Effectiveness and Value; Final Evidence Report. Institute for Clinical and Economic Review, March 23, 2018. https://icer-review.org/material/car-t-final-report/. Accessed August 14, 2019.
5. Rind D, Agboola F, Steuten L, et al. Emicizumab for Hemophilia A with Inhibitors: Effectiveness and Value; Final Evidence Report. Institute for Clinical and Economic Review, April 16, 2018. https://icer-review.org/material/hemophilia-a-final-evidence-report/. Accessed August 14, 2019.
6. Rind D, Ollendorf D, Guzauskas G, et al. Treatment Options for Advanced Non-Small Cell Lung Cancer: Effectiveness, Value and Value-Based Price Benchmarks; Final Evidence Report and Meeting Summary. Institute for Clinical and Economic Review, November 1, 2016. https://icer-review.org/material/nsclc-final-report/. Accessed August 14, 2019.
7. Ellis A, Mickle K, Thokala P, et al. Spinraza® and Zolgensma® for Spinal Muscular Atrophy: Effectiveness and Value. Institute for Clinical and Economic Review, May 24, 2019. https://icer-review.org/material/sma-final-evidence-report/. Accessed August 16, 2019.
8. Lakdawalla DN, Doshi JA, Garrison LP, Jr., Phelps CE, Basu A, Danzon PM. Defining Elements of Value in Health Care-A Health Economics Approach: An ISPOR Special Task Force Report. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2018;21(2):131-139.
9. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. *JAMA.* 2016;316(10):1093-1103.
10. Carias C, Chesson HW, Grosse SD, et al. Recommendations of the Second Panel on Cost Effectiveness in Health and Medicine: A Reference, Not a Rule Book. *American journal of preventive medicine.* 2018;54(4):600-602.
11. Institute for Clinical and Economic Review. Overview of the ICER value framework and proposals for an update for 2017-2018. February 1, 2017. http://icer-review.org/wp-content/uploads/2016/02/ICER-VAF-Update-Proposals-020117.pdf. Accessed August 19, 2019.
12. Linley WG, Hughes DA. Societal views on NICE, cancer drugs fund and value-based pricing criteria for prioritising medicines: a cross-sectional survey of 4118 adults in Great Britain. *Health Econ.* 2013;22(8):948-964.
13. Karlsberg Schaffer S, Cubi-Molla P, Devlin N, Towse A. Shaping the research agenda to estimate relevant cost-effectiveness thresholds for health technology assessment decision making: report for ABPI. Office of Health Economics Consulting, April, 2016.

https://www.ohe.org/publications/shaping-research-agenda-estimate-cost-effectiveness-thresholds-decision-making#. Accessed August 14, 2019.
14. Ryen L, Svensson M. The Willingness to Pay for a Quality Adjusted Life Year: A Review of the Empirical Literature. *Health Econ.* 2015;24(10):1289-1301.
15. Hutubessy R, Chisholm D, Edejer TT-T, Who C. Generalized cost-effectiveness analysis for national-level priority-setting in the health sector. *Cost Effectiveness and Resource Allocation.* 2003;1(1):8.
16. World Health Organization. WHO Commission on Macroeconomics and Health & World Health Organization. (2001). Macroeconomics and health : investing in health for economic development : executive summary / report of the Commission on Macroeconomics and Health. 2001; https://apps.who.int/iris/handle/10665/42463. Accessed August 19, 2019.
17. Marseille E, Larson B, Kazi DS, Kahn JG, Rosen S. Thresholds for the cost-effectiveness of interventions: alternative approaches. *Bulletin of the World Health Organization.* 2015;93(2):118-124.
18. Bertram MY, Lauer JA, De Joncheere K, et al. Cost-effectiveness thresholds: pros and cons. *Bulletin of the World Health Organization.* 2016;94(12):925-930.
19. Robinson LA, Hammitt JK, O'Keeffe L. Valuing Mortality Risk Reductions in Global Benefit-Cost Analysis. *Journal of Benefit-Cost Analysis.* 2019;10(S1):15-50.
20. Hirth RA, Chernew ME, Miller E, Fendrick AM, Weissert WG. Willingness to Pay for a Quality-adjusted Life Year:In Search of a Standard. *Medical Decision Making.* 2000;20(3):332-342.
21. Phelps CE. A New Method to Determine the Optimal Willingness to Pay in Cost-Effectiveness Analysis. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2019;22(7):785-791.
22. Garber AM, Phelps CE. Economic foundations of cost-effectiveness analysis. *Journal of health economics.* 1997;16(1):1-31.
23. Arrow K. *Essays in the Theory of Risk-Bearing.* 2016/08/17 ed. Amsterdam: North-Holland Publishing Company; 1971.
24. US Bureau of Economic Analysis. Personal Income Per Capita [A792RC0A052NBEA]. Federal Reserve Bank of St. Louis, 2018. https://fred.stlouisfed.org/series/A792RC0A052NBEA.
25. Pauly MV. The Questionable Economic Case for Value-Based Drug Pricing in Market Health Systems. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2017;20(2):278-282.
26. Woods B, Revill P, Sculpher M, Claxton K. Country-Level Cost-Effectiveness Thresholds: Initial Estimates and the Need for Further Research. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2016;19(8):929-935.
27. Lomas JRS, Martin S, Claxton KP. Estimating the marginal productivity of the English National Health Service from 2003/04 to 2012/13. Centre for Health Economics, University of York, 2018.
28. Claxton K, Lomas J, Martin S. The impact of NHS expenditure on health outcomes in England: Alternative approaches to identification in all-cause and disease specific models of mortality. *Health Econ.* 2018;27(6):1017-1023.
29. Vanness D. Deriving an opportunity cost-based threshold for CEA in the United States. Paper presented at: International Health Economic Association; July, 2017; Boston, MA.
30. Maryland Health Services Cost Review Commission. Maryland's Total Cost of Care Model: Update for Providers. https://hscrc.state.md.us/Pages/tcocmodel.aspx. Accessed August 20, 2019.
31. Massachusetts Health Policy Commission. Health Care Cost Growth Benchmark. 2019; https://www.mass.gov/info-details/health-care-cost-growth-benchmark. Accessed August 19, 2019.

32. National Institute for Health and Care Excellence (NICE). Developing NICE guidelines: the manual. 2014; https://www.nice.org.uk/process/pmg20/chapter/introduction-and-overview. Accessed August 14, 2019.
33. Paulden M, O'Mahony JF, McCabe C. Discounting the Recommendations of the Second Panel on Cost-Effectiveness in Health and Medicine. *PharmacoEconomics.* 2017;35(1):5-13.
34. Severens JL, Milne RJ. Discounting health outcomes in economic evaluation: the ongoing debate. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2004;7(4):397-401.
35. Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG. *Cost-effectiveness in health and medicine.* New York: Oxford University Press; 2016.
36. Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-effectiveness in health and medicine.* New York: Oxford University Press; 1996.
37. Davis S. NICE Decision Support Unit Methods Development. In: *Assessing Technologies That Are Not Cost-Effective at a Zero Price.* London: National Institute for Health and Care Excellence (NICE); 2014.
38. Pelletier EM, Smith PJ, Dembek CJ. Payer Costs of Autologous Stem Cell Transplant: Results from a U.S. Claims Data Analysis. *Blood.* 2008;112(11):2373-2373.
39. Grosse S. In:2018.
40. Basu A, Meltzer D. Implications of spillover effects within the family for medical cost-effectiveness analysis. *Journal of health economics.* 2005;24(4):751-773.
41. Pacheco Barzallo D. Spillover Effects of Long-Term Disabilities on Close Family Members. *Applied health economics and health policy.* 2018;16(3):347-355.
42. Lu Y, Penrod JR, Sood N, Woodby S, Philipson T. Dynamic cost-effectiveness of oncology drugs. *The American journal of managed care.* 2012;18(11 Suppl):S249-256.
43. Hoyle M. Future drug prices and cost-effectiveness analyses. *PharmacoEconomics.* 2008;26(7):589-602.
44. Clark S, Weale A. Social values in health priority setting: a conceptual framework. *Journal of health organization and management.* 2012;26(3):293-316.
45. National Institute for Health and Care Excellence (NICE). Consultation Paper: Value Based Assessment of Health Technologies. https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/VBA-TA-Methods-Guide-for-Consultation.pdf. Accessed August 19, 2019.
46. Reckers-Droog VT, van Exel NJA, Brouwer WBF. Looking back and moving forward: On the application of proportional shortfall in healthcare priority setting in the Netherlands. *Health policy (Amsterdam, Netherlands).* 2018;122(6):621-629.
47. Versteegh MM, Ramos IC, Buyukkaramikli NC, Ansaripour A, Reckers-Droog VT, Brouwer WBF. Severity-Adjusted Probability of Being Cost Effective. *PharmacoEconomics.* 2019.
48. Ottersen T, Forde R, Kakad M, et al. A new proposal for priority setting in Norway: Open and fair. *Health policy (Amsterdam, Netherlands).* 2016;120(3):246-251.
49. van de Wetering EJ, Stolk EA, van Exel NJ, Brouwer WB. Balancing equity and efficiency in the Dutch basic benefits package using the principle of proportional shortfall. *The European journal of health economics : HEPAC : health economics in prevention and care.* 2013;14(1):107-115.
50. Stolk EA, van Donselaar G, Brouwer WB, Busschbach JJ. Reconciliation of economic concerns and health policy: illustration of an equity adjustment procedure using proportional shortfall. *PharmacoEconomics.* 2004;22(17):1097-1107.
51. Claxton K, Sculpher M, Palmer S, Culyer AJ. Causes for concern: is NICE failing to uphold its responsibilities to all NHS patients? *Health Econ.* 2015;24(1):1-7.

52. Jit M. MCDA from a health economics perspective: opportunities and pitfalls of extending economic evaluation to incorporate broader outcomes. *Cost Eff Resour Alloc.* 2018;16(Suppl 1):45-45.
53. Phelps CE, Lakdawalla DN, Basu A, Drummond MF, Towse A, Danzon PM. Approaches to Aggregation and Decision Making-A Health Economics Approach: An ISPOR Special Task Force Report [5]. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2018;21(2):146-154.

Appendix

Appendix Figure 1a. Standard Review Timeline

ICER Process	Week	Milestones	Comments
Topic Announced	0	Topic Announcement Open Input Period Begins	ICER begins scoping calls with clinical experts and patient groups. Stakeholders may submit information through the open input period.
Draft Scope	1		
	2		
	3	Open Input Period Ends Draft Scoping Document Posted	Manufacturers and other stakeholders have 15 business days to comment on the draft scope.
Final Scope	4	Public Comment Period	ICER holds calls with manufacturers to discuss the draft scoping document
	5		
	6		
	7	Final Scoping Document Posted ICER Sends Request for Data	ICER sends formal requests for data to each manufacturer. Supplemental data requests may be sent on an ad hoc basis.
Draft Evidence Report	8		
	9		
	10		
	11	Mfr. Evidence Submissions Due	
	12	Research Protocol Posting	Posting of evidence review protocol
	13		
	14		
	15	Preliminary Model Presentation Posting of Model Analysis Plan	Individual discussion calls with manufacturers 2-3 days after the preliminary model presentation. After reviewing ICER's preliminary model presentation, manufacturers may send supplemental data.
	16		
	17	Supplemental Data Submission Due	Supplemental data sent in response to ICER's preliminary model presentation are due 11 business days after call.
	18		
	19		
	20		
	21	Draft Evidence Report Posted	
Evidence Report	22	Public Comment Period	Mfrs. and other stakeholders have 20 business days to comment on the Draft Evidence Report. When possible, economic models are available for review by manufacturers.
	23		
	24		
	25		
	26		
	27		
Public Meeting	28	Evidence Report Posted	The relevant program voting panel reads this version of the report.
	29		
	30	Public Meeting	
Final Report	31		
	32		
	33	Final Evidence Report Posted	

Legend: Document Release | Data Request | Input Opportunity

Appendix Figure 1b. Proposed Changes to Timeline for Large Class Reviews

ICER Process	Week	Milestones	Class Review Adaptation
Topic Announced	0	Topic Announcement Open Input Period Begins	
Draft Scope	1		
	2		
	3	Open Input Period Ends Draft Scoping Document Posted	
Final Scope	4	Public Comment Period	
	5		
	6		
	7		+1 week for additional scoping calls
	8	Final Scoping Document Posted ICER Sends Request for Data	
Draft Evidence Report	9		
	10		
	11		
	12	Mfr. Evidence Submissions Due	
	13	Research Protocol Posting	
	14		+3 weeks to systematic literature review and model development timelines
	15		
	16		
	17		
	18		
	19	Preliminary Model Presentation Posting of Model Analysis Plan	
	20		
	21	Supplemental Data Submission Due	
	22		
	23		+1 week to address feedback on preliminary model
	24		
	25		
	26		+1 week to facilitate revision of longer and more complex report
	27	Draft Evidence Report Posted	
Evidence Report	28	Public Comment Period	+1 week to public comment period
	29		
	30		
	31		
	32		
	33		
	34		
	35		+1 week to review a higher volume of stakeholder comments
Public Meeting	36	Evidence Report Posted	
	37		+1 week to allow CTAF/CEPACs sufficient time to review complex report
	38		
	39	Public Meeting	
Final Report	40		
	41		
	42	Final Evidence Report Posted	

Legend: Document Release | Data Request | Input Opportunity