Value Assessment Framework

Updated September 25, 2023

©Institute for Clinical and Economic Review, 2023

About ICER

The Institute for Clinical and Economic Review (ICER) is an independent non-profit research organization that evaluates medical evidence and convenes public deliberative bodies to help stakeholders interpret and apply evidence to improve patient outcomes and control costs. Through all our work, we seek to help create a future in which collaborative efforts to move evidence into action provide the foundation for a more effective, efficient, and just health care system. More information about ICER is available at http://www.icer.org.

Funding for ICER's review activities comes from government grants and non-profit foundations, with the largest single funder being Arnold Ventures. No funding for these activities comes from health insurers, pharmacy benefit managers, or life science companies. ICER receives approximately 22% of our overall revenue from these health industry organizations to run a separate Policy Summit program, with funding approximately equally split between insurers/PBMs and life science companies. For a complete list of funders and for more information on ICER's support, please visit https://icer.org/who-we-are/independent-funding/

About this Document

This paper presents final updates to the ICER Value Assessment Framework, including refinements of its conceptual structure and modifications to the specific methods used to gather and assess evidence of different types. Adaptations to this framework are summarized in Table C.1 in the Supplement. Separate documents describing adaptations to this framework for treatments for ultra-rare diseases and single- or short-term transformative therapies can be found on ICER's website.

This update to the ICER Value Assessment Framework builds upon ICER's experience using the 2017-2019 and 2020-2023 frameworks in the evaluation of drugs, devices, tests, and delivery system innovations, as well as earlier iterations of the framework. During that time we have actively sought the input of all stakeholders and made iterative changes to our methods and overall procedures to enhance transparency and to improve the ability of all parties to participate meaningfully in the process. We have also benefitted from public comment opportunities during each framework revision cycle, including two comment periods for the 2020-2023 framework; the first being a call for open public input to propose changes to the framework, the second providing an opportunity for stakeholders to comment on proposed changes. During the 2023 public comment period, we received feedback from 32 organizations. Their comments can be found here along with our summary response to comments here. We wish to thank all of these commenters for the time and effort they put into these comments, and the many thoughtful contributions they have made.

This paper reflects this combined experience, public input, and many additional discussions with stakeholders in various settings. This finalized update to the ICER value framework and associated methods will be in place to guide reports launched after October 2023.

Table of Contents

1. Introduction

1.1. Overview

This document contains an overview and discussion of the concepts that underpin ICER's Value Assessment Framework. The Framework describes ICER's philosophy and approach assessing the value of a medical intervention at the population level, as well as the implications of its findings for practice and policy. Detailed descriptions of the technical methods we use to conduct our assessments (e.g., the ICER Evidence-Based Medicine Rating Matrix, Reference Case for economic modeling, and stakeholder engagement guides) may be found on ICER's website and links to these materials are provided in related sections of this document.

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

For more than 15 years we have 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." ICER's first formal Value Assessment Framework was posted in 2015, and following two years of further experience, and several rounds of public comment, an update to the framework was posted in early 2017 as the guide to ICER's reviews for 2017-2019. In 2019 we updated our framework for 2020 – 2023. This most recent update has also benefited from extended discussions with stakeholders, experience over the past three years, and formal public comment. This version of the ICER Value Assessment Framework will serve as the standard for our methods for our reports launched after October 2023.

Ultimately, the purpose of ICER's 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 we are guided by several key underlying principles. One is that we act with respect for all, in concordance with a presumption of goodwill on the part of all participants and stakeholders in the health care system. We do 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. We believe 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, we believe 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.

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

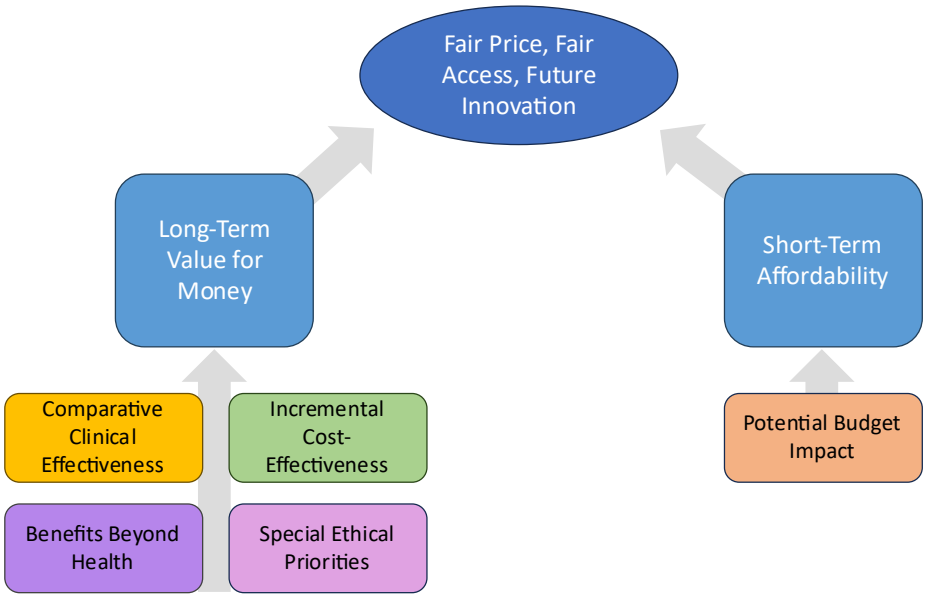
The ICER Value Assessment Framework describes the conceptual framework and set of associated methods that guide the development of ICER evidence reports. ICER reports are intended to support deliberation on medical policies related to health services (e.g., tests or treatments) and delivery system interventions (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.

1.4. Conceptual Structure of the ICER Value Assessment Framework

As shown in the structure of the ICER value framework, it seeks to inform decisions that are aimed at achieving sustainable access to high-value care for all patients (see Figure 1.1 below). This goal requires consideration of two general concepts: long-term value for money and short-term affordability.

Figure 1.1. Conceptual Structure of the ICER Value Assessment Framework



Long-term Value for Money

Long-term value for money serves as the primary anchor of the ICER value framework. It is itself a concept that is comprised of multiple domains: 1) comparative clinical effectiveness, 2) incremental cost-effectiveness, 3) benefits beyond health, and 4) special ethical priorities. A description of how these domains are measured and integrated into an ultimate judgment of long-term value for money is described in later sections of this paper. There are several high-level points about this element of the value framework that bear highlighting here:

Long-term perspective

Even though most of the clinical data available on health care services come from studies of relatively short duration, the grounding of any evaluation of value should recognize the long-term perspective on both outcomes for patients and costs. The ICER value framework recognizes this principle by grounding the methods of incremental cost-effectiveness analysis in simulations that estimate outcomes and costs at the longest feasible time horizon, usually the full lifetime of patients. Benefits for patients and potential cost offsets for new treatments that might take many years to be seen are therefore estimated and included as a core element of the value framework.

Foundation in the evaluation of evidence on comparative clinical effectiveness

The ICER value framework is rooted in an objective evaluation of the evidence on the comparative clinical effectiveness of different care or care delivery options. This element of the framework serves as the primary source of information to inform cost-effectiveness analysis and includes a systematic review of available evidence performed according to the highest academic methodological standards. As part of the evaluation of comparative clinical effectiveness, ICER reports include a clear description of the sources of evidence, the strengths and limitations of individual studies, and a summary judgment of the net health benefit of different care options along with a statement explaining the relative certainty that the body of evidence is able to provide. The methods used by ICER in our evaluation of comparative clinical effectiveness are discussed in Section 2 of this paper and described in more detail in documents available on the ICER website. An earlier incarnation of the ICER rating system for evidence of comparative clinical effectiveness was published in a peer-reviewed journal and was endorsed by the AMCP-NPC-ISPOR Comparative Effectiveness Research Collaborative.[1,2]

Acceptance of multiple forms of evidence

Patients, clinicians, and policymakers are most interested in evidence on the comparative clinical effectiveness of care options, but this does not mean that ICER's value framework limits the type of evidence to be considered to the results of randomized controlled trials (RCTs).

When available, high-quality RCTs and systematic reviews of RCTs provide evidence that is least susceptible to many scientific biases. However, head-to-head RCTs of active comparators are uncommon, especially for interventions near the time of regulatory approval. Without direct head-to-head evidence, insights into comparative clinical effectiveness may require indirect comparisons through formal network meta-analysis. Complementing these sources of information is evidence derived from many different analytic approaches and that are available from a wide range of sources. Although more vulnerable to some important confounding biases, observational methodologies such as cohort studies, case-control studies, and long-term disease and drug registries often provide helpful evidence, particularly on longer-term outcomes. As will be described in greater detail later in this document, we also have a commitment to explore how "real-world" observational evidence can contribute to a more comprehensive and accurate view of the risks, benefits, and costs associated with any intervention. This commitment extends not only from using available published sources, but includes the possibility of working with life science companies, patient groups, or data aggregator companies to develop and analyze new sources of real-world evidence in a way that will meet the evidentiary standards relevant to the questions being addressed.

In short, ICER has a flexible and inclusive approach to sources of evidence and, while stressing the importance of the rigor of clinical trial data in any assessment, the value framework and ICER's methods incorporate multiple sources and types of evidence, seeking the evidence that is most helpful in understanding the long-term net health benefits for patients of different care options.

Recognition that what matters to patients is not limited to measured "clinical" outcomes

The inclusion of an explicit domain of value labeled "benefits beyond health" demonstrates that the ICER value framework fully acknowledges that all too often what matters most to patients is poorly captured in the available clinical trial data. Sometimes this occurs because the clinical outcomes measured do not reflect what is most important to patients' day-to-day quality of life. Even when trials do capture the clinical outcomes that matter most to patients, there are other aspects of the treatment regimen that have a significant impact on the overall value of the treatment. This can be related to the complexity of the treatment regimen or the impact of care options on the ability of patients to return to work, on family and caregivers, on overall public health, or on other aspects of the healthcare system or society.[i] The ICER value framework identifies these "benefits beyond health" as important elements of any overall judgment on long-term value for money, and all ICER reports have separate sections in which evidence and information pertaining to these elements are presented. We describe in Section 4 of this paper a method for integrating these domains of value.

[i] For further insight and examples a useful resource is the FasterCures and Avalere Health work on "Integrating the Patient Perspective into the Development of Value Frameworks" available at http://www.fastercures.org/assets/Uploads/value-coverage-framework-March-2016.pdf

Acknowledgment of the role of special ethical priorities

Decisions about the value of care options do not happen in a vacuum. How to interpret and apply evidence in clinical care, insurance coverage, and pricing, involves a complex process of integrating information on risks and benefits of treatment within a broader set of special ethical priorities. The ICER value framework includes these elements and they are explored in a separate section of each ICER report. In addition, special ethical priorities often feature prominently in the deliberation on value between independent expert committees and all stakeholders and is a central feature of the public meetings convened by ICER on each report. Linked to the discussion of "benefits beyond health," we discuss the methods used to integrate special ethical priorities into ICER reports in Section 4.

Short-term Affordability

With long-term value for money being the dominant element in considerations of value, a complementary perspective is provided by including an evaluation of short-term affordability. The ICER value framework includes an explicit evaluation of the short-term affordability of different care options by analyzing the potential short-term budget impact of changes in health care expenditures with the introduction of a new test, treatment, or delivery system process. Detailed methods used to estimate potential budget impact are presented later in Section 5 of this paper.

Budget impact is a reasonable consideration within a Value Assessment Framework because insurers work in rapid cycles with purchasers and individual subscribers, translating short-term cost projections into planned insurance premiums for the coming year. Rapid cost growth in the short-term, especially when it increases beyond anticipated inflation rates, pushes quickly upstream to purchasers and policymakers who have to make their own short-term decisions about how to find the needed resources. This may lead to decisions to increase deductibles or otherwise reduce health care benefits for employees; for example, state governments might need to consider reducing next year's education budget to find the funds to keep a Medicaid program afloat.

In addition, for provider groups that bear financial risk, budget impact analyses inform very real short-term decisions about how to allocate resources to maximize the quality of health care within a given budget. A rapid increase in costs resulting from the significant budget impact of a new drug might lead to decisions to forgo hiring of needed new staff or delay the introduction of other new services. Quite simply: budget impact, and not long-term cost-effectiveness, determines how affordable health care insurance will be in coming years and shapes what health care can be provided with the resources available.

ICER's value framework represents the conviction that keeping budget impact considerations off the table, to be factored in only *post hoc* by insurers or provider groups in ways unknown, would be a mistake. It would rob our nation of the chance to bring the public directly into the critical

discussions about health care and health insurance that we need to have if we are going to achieve sustainable access for all patients to the kind of innovative new tests, treatments, and delivery system interventions that add value to our lives.

Potential budget impact analyses estimate the net budget impact across all elements of the health care system

ICER's methods have never sought to estimate the potential budget impact of treatments within "silos" of a payer budget, such as the expenses only on pharmaceuticals, devices, or hospital costs. It remains a core principle of ICER's value framework that we should evaluate both short- and long-term costs across the entire health system, so that care options that might increase spending for one type of service (e.g. drugs) while reducing other spending (e.g. hospital costs) receive full credit for cost offsets and are not penalized in any way.

At five years, the time frame for considering "short-term" affordability is stretched as far as possible without losing relevance for identifying new care options that may require special measures – in pricing, payment mechanisms, coverage criteria, or budgeting – to maintain patient access without serious financial strain throughout the health care system. Using a five-year time horizon may reduce the utility of the analysis for insurers focused on shorter budget timeframes but helps accommodate some of the important potential clinical benefits and cost offsets that may not occur immediately with the adoption of a new therapeutic option. With the primary anchor of the ICER value framework being the long-term perspective represented by long-term value for money, the time horizon for short-term affordability has been extended as far as it seems possible in order for it to serve the important purpose of informing discussions on whether special efforts need to be taken to manage the introduction of a new therapeutic option so that access and affordability can both be maintained.

Considerations for Assessments of Non-Drug Interventions

Devices

There are many important, unique aspects to the development, early evaluation, regulatory approval, and patterns of use and iterative evidence generation for devices. Therefore, although the conceptual elements of the ICER value framework remain the same for any health care intervention, the specific methods for incorporating and judging evidence will differ for devices. For example, ICER methods acknowledge the practical and ethical considerations that may make it impossible to use RCTs in the early evaluation of clinical effectiveness, while iterative changes to devices, along with the learning curve for practitioners, also raise special considerations about how to judge the available evidence. Evaluations of long-term cost- effectiveness are made challenging because of the potential for evolution of devices and the attendant changes in cost, effectiveness, and the types of patients who will be treated. These complexities are also relevant to estimations

of potential budget impact, and, as noted in sections below, it is very difficult to identify the current baseline costs of all device use in the US health care system in order to calculate a growth target for a budget impact threshold. For these reasons the conceptual elements of the ICER value framework remain relevant for devices but within that framework we will continue to incorporate specific approaches to evidence evaluation for devices that reflect their unique features.

Digital Health Technologies

Digital Health Technologies (DHTs) are a broad and rapidly innovating class of health technology with distinctive pathways for development, regulatory approval, uptake, and reimbursement. Given the unique nature of DHTs, existing value assessment frameworks and evidence standards for health technologies such as drugs and devices are not directly applicable. Working with the Peterson Health Technology Institute (PHTI), ICER developed a new value framework for the assessment of DHTs that was published in September 2023. The ICER-PHTI assessment framework positions evaluation of clinical effectiveness and economic impact as two primary domains whose outcomes are informed by other attributes of the technology, including user experience, impact on health equity, privacy, and data security. The ICER-PHTI assessment framework is anchored on comparative clinical effectiveness, but emphasizes budget impact as the primary measure of comparative value. We will use this alternative framework for assessments of DHTs for reviews starting in 2024.

Tests

Similarly, different approaches to evidence evaluation are required for diagnostic interventions and tests used to monitor patients or provide information on disease prognosis. For example, the general hierarchy in the types and strength of evidence for tests is different than that for therapeutic interventions.[ii] As with devices, tests will continue to be evaluated using the overall conceptual approach of the ICER value framework but there will be important modifications based on the distinctive nature of the evidence and the system for the development, evaluation, and use of diagnostic interventions. Further work will be needed to develop a method for estimating a threshold for potential budget impact that should trigger additional policymaker consideration of short-term affordability.

Delivery System Innovations

There are also many distinctive challenges to evaluating the evidence on the effectiveness and value of delivery system interventions. Chief among these is that in most cases a delivery system

[ii]See for example the discussion of the Fryback and Thornbury evidentiary model used as part of the ICER review on cardiac nuclear imaging, coronary computed tomographic angiography, CT colonography, breast cancer screening, and diagnostic tests for Alzheimer's disease (https://icer.org/explore-our-research/assessments/)

intervention will be highly variable in its implemented form across different settings, raising great questions about the generalizability of results from studies of one institution or one system of care. RCTs can be difficult to perform, increasing concerns about the internal validity of study findings. We will use the same general value assessment framework to guide its reviews of delivery system interventions, but as with devices and tests, some of the specific methods for judging evidence and for determining thresholds for potential budget impact analysis will reflect the unique nature of these kinds of health service innovations.

2. Comparative Clinical Effectiveness

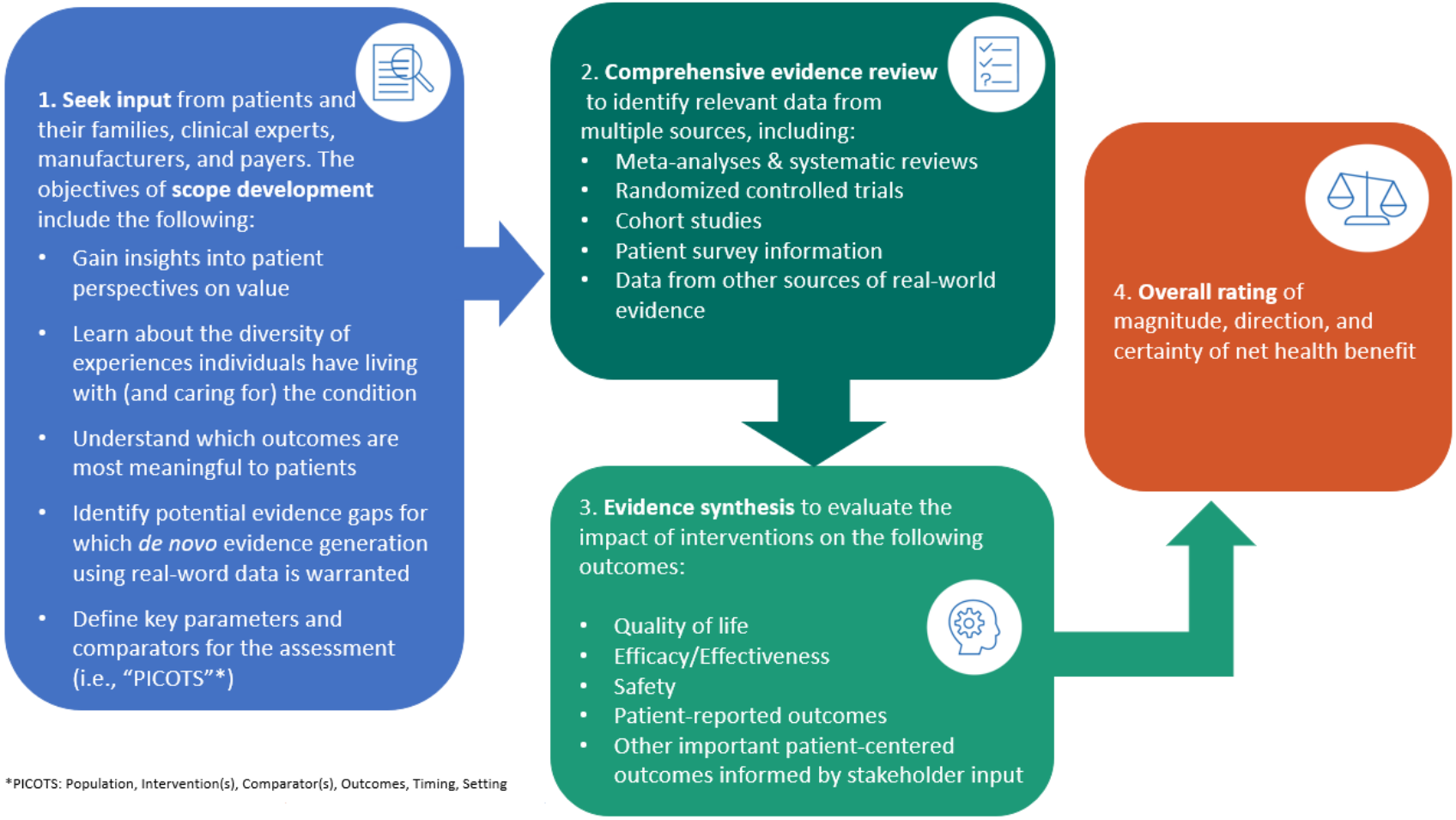
2.1. Overview

A central part of the ICER value framework is an objective evaluation of the evidence on comparative clinical effectiveness. Comparative clinical effectiveness involves weighing the benefits and harms/burdens of one treatment option versus another. The most important benefits and harms are those that are important to patients and their families/caregivers. As such, from the outset of a review we solicit input from patients, families, caregivers, and expert clinicians to understand the day-to-day experience of living with a condition and what outcomes it would be most important for a therapy to affect. Information on what has been learned from patient input is presented in the ICER report prior to the discussions of the evidence so that readers can interpret the evidence through the lens of patient experience.

Stakeholder input from clinicians, manufacturers, and payers is used in addition to input from patients and families to frame the questions that an ICER comparative effectiveness review attempts to answer. When evidence on patient-important outcomes is limited or unavailable,we will seek evidence on surrogate endpoints that might be associated with outcomes important to patients and families.

Once we have defined the scope of a review, we evaluate the available clinical evidence. We conduct a systematic review of the existing literature using established best practices for evidence synthesis. The findings of our evidence review are described in a publicly available report, which includes a description of the sources of evidence, the strengths and limitations of individual studies, the diversity of clinical trial participants, an assessment of the relevancy and generalizability of the published literature for patient and provider decision making, and a summary rating of the net health benefit of different care options. ICER's approach to evaluating the comparative clinical effectiveness is summarized in Figure 2.1 and discussed in the section that follows.

Figure 2.1. Summary Process for Assessment of Comparative Clinical Effectiveness



2.2. Scope of Clinical Effectiveness Evaluation

From the inception of the evaluation, we examine the contextual landscape of the topic under review. We compile data related to epidemiology, prognosis, standards of care, and natural history, while seeking to understand the lived experiences of patients affected by the condition. Insights from patient groups and other stakeholders, along with reviews of the evidence, inform definitions of the population, interventions, comparators, outcomes, timing, and setting (PICOTS) components that anchor ICER's evaluation of comparative clinical effectiveness. These components are described below.

- **Population:** The population that is eligible to use the intervention(s) under review. For certain topics, such as drug therapies, the population may be defined to align with current or anticipated FDA indications for that therapy. We also examine whether there are subpopulations for whom the relative effectiveness or safety of the intervention may vary or whether there are subpopulations for whom variations in baseline risk lead to higher or lower absolute benefits or harms. These different subgroups are highlighted in the ICER report.

- **Interventions**: Interventions may include drug therapies, medical tests, devices, and delivery system innovations, among others. When relevant, we may focus our review on specific attributes of an intervention (e.g., mode of administration, line of therapy, etc.).

- **Comparators:** Appropriate comparators represent alternative therapies used among the populations and settings of focus. Active comparators (i.e., non-placebo interventions) are prioritized when feasible. Relevant comparators are selected through a survey of clinical guidelines from professional societies, consultation with clinical experts and patients, and review of clinical trial designs.

- **Outcomes:** Critical to the evaluation of net health benefit of an intervention are the measures of potential benefit and harm. Health outcomes, i.e., changes in symptoms or conditions that people experience and that affect the quantity or quality of life (e.g., change in pain, quality of life, length of life) are given greater weight than intermediate outcomes (e.g., change in cholesterol). Patient-important outcomes are health outcomes that are central to ICER's judgements of benefit and harm. When appropriate, we also look for evidence on non-clinical outcomes such as resource utilization or measures of societal benefit.

- **Timing:** The minimum duration of study follow-up considered adequate to capture the outcomes of interest.

- **Setting:** The setting(s) of focus for a review may be specified (e.g., inpatient, emergency department, and/or outpatient) and we will state whether we will exclude certain study settings from consideration.

2.3. Sources of Evidence

ICER's evaluation of comparative clinical effectiveness is grounded in a systematic review of all available evidence. A systematic review identifies all relevant existing evidence using explicit, replicable methods in a way that minimizes the risk of biased selection of studies. Established best methods of systematic literature reviews are followed in order to foster transparency and facilitate reproduction of results.[3,4] Reviews are reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[5]

ICER's judgements around comparative clinical effectiveness are informed by evidence arising from multiple sources. When available, high-quality RCTs or their meta-analyses provide evidence that is least susceptible to certain scientific biases. When benefits and harms occur over the course of many years, or when harms are rare but clinically important or even catastrophic, evidence from high quality published peer-reviewed studies using observational data and methodologies such as cohort studies, case-control studies, and long-term disease and drug registries may be used. Furthermore, if important patient reported outcomes have not been collected as part of a manufacturer's clinical development program, we will again conduct a comprehensive literature review to identify any published, peer-reviewed observational studies providing this information.

Real-World Evidence

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. We have 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 meetings and appraisal committee 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 RCTs 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 "benefits beyond health" 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.

ICER's use of real-world data also may include *de novo* evidence generation under certain circumstances where critical data elements are lacking. Options for generating new RWE may include conducting a patient survey using a validated patient-reported outcome (PRO) instrument or using claims data to better understand concordance and persistence. 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.

As with all evidence, we 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. The process by which we will evaluate RWE will follow the general outline presented in our separate framework to guide the optimal development and use of real-world evidence for drug coverage and formulary decisions.[6] Efforts will be undertaken to assure that the data are curated with input from individuals with knowledge of the nuances of the data source. Methods for adjusting for known and potential unknown confounders will be assessed, and replication of results using different methods within the same data set and/or using different data sources will be pursued. We will also apply best practices in real-world data analysis as described in guidelines from ISPOR and other authoritative methods bodies.[7]

Grey Literature

ICER also includes evidence from the "grey literature" as per our criteria available at https://icer.org/policy-on-inclusion-of-grey-literature-in-evidence-reviews/. We supplement our reviews of studies from peer-reviewed publications with data from conference proceedings, regulatory documents, materials from other HTA groups, information submitted by manufacturers, and input gleaned from patients. Consideration of multiple sources of evidence helps us evaluate whether there is a biased representation of study results in the published literature and provides a panoramic understanding of net health benefit.

In summary, we have a flexible and inclusive approach to sources of evidence, which stresses the importance of the rigor of clinical trial while augmenting such evidence with data from other real-world or grey-literature sources.

Patient Input on Clinical Trial Design

The FDA's Patient-Focused Drug Development[8] (PFDD) program has been an important step in acknowledging the critical importance of capturing outcomes that matter most to patients in the clinical development program for new drugs. Many of the PFDD efforts have yet to be realized in evidence currently being produced for new drugs, as changes to the development of clinical trial programs takes time. However, some manufacturers have made important strides in incorporating the patient perspective into their drug development programs. To highlight industry-leading practices, we will invite the manufacturers of drugs under review to provide a written description of how patients were involved in the design of the clinical trial program. Manufacturers are encouraged to describe the methods used to collect

patient experience data, and how they identified the outcomes most important to patients in their clinical trials.[9] We will include any written descriptions received from manufacturers in the Evidence Report.

2.4. Appraisal and Synthesis of Evidence

Assessment of Quality of Individual Studies

We evaluate the methodological quality of individual studies in part by applying risk of bias tools deemed appropriate for the topic under review. The quality assessment tool developed by the US Preventive Services Task Force (USPSTF) for judging the quality of clinical trials and cohort studies is one of the tools we commonly adapt.[10] The Cochrane Risk of Bias Assessment Tool for randomized control trial is another published tool we commonly use. However, we believes that no single tool exists that is ideal for evaluating all possible studies included across reviews. Thus, for each review, we thoughtfully consider which quality assessment tools are most appropriate for the topic at hand and document the choice in our protocol.

When examining individual study quality, the main focus is on risk of bias and selective reporting rather than other aspects pertaining to study conduct (e.g., obtaining ethical approval or calculating sample size). Of note, ICER's assessment focuses on the internal validity of the study (i.e., how well the study is able to estimate what it set out to measure). Relevant quality issues evaluated in our assessment include selection bias (e.g. was allocation concealed?), performance bias (e.g., were patients blinded?), attrition bias (e.g., was intention to treat analysis used?), detection bias (e.g., was outcome assessment blinded?), and selective reporting (e.g., were the important outcomes measured and analyzed in the study fully reported?). ICER's assessment incorporates how particular aspects of a study may lead to biased results and states the likely direction of such bias.

For each review, the rationale for the assessments is explicitly determined *a priori* in our protocol, and the judgment on each study is provided in the appendix of each Evidence Report.

Evaluation of Clinical Trial Diversity

We capture the demographic diversity of participants in the clinical trials included in ICER assessments in a subsection titled "Clinical Trial Diversity." ICER's evaluation of clinical diversity is not meant to address the question of generalizability. Instead, it is designed to promote conversations around equity in clinical trials of new drugs. Each trial of the intervention under review will be rated on diversity based on the following demographic characteristics: race/ethnicity, sex, and age (older adults). In addition, information on other relevant demographic characteristics will be described in this section as appropriate.

To do this objectively and consistently across all ICER assessments, we have developed a framework for evaluating clinical trial diversity based on the potential best practices described in our white paper on Advancing Health Technology Assessment Methods that Support Health Equity.[11] This framework relies on evaluating clinical trial diversity quantitatively by comparing clinical trial participants to disease-specific

prevalence estimates and using the thresholds defined in Tables 2.1 and 2.2 below to judge representation. Specifically, a score that ranges from 0 to 3 is assigned to each demographic category based on the estimated participation-to-prevalence ratios. Then, based on the cumulative score and the pre-defined cut points for the demographic characteristic being evaluated, a rating of "good," "fair," or "poor" is used to communicate the demographic diversity of the participants in a clinical trial.

Table 2.1. Representation Score

PDRR (Participant to Disease Prevalent Ratio)	Representation Score
0 or not reported	0
>0 and Less than 0.5	1
0.5 to 0.8	2
≥0.8	3

Table 2.2. Rating Categories

Demographic Characteristics	Demographic Categories Included in Rating	Maximum Score	Rating Categories (Total Score)
Race and Ethnicity*	Asian Black or African American White Hispanic or Latino	12	Good (11-12) Fair (7-10) Poor (≤6)
Sex	Male Female	6	Good (6) Fair (5) Poor (≤4)
Age	Older adults (≥65 years)	3	Good (3) Fair (2) Poor (≤1)

*American Indian or Alaskan Native & Native Hawaiian or Other Pacific Islander are not factored into the overall racial and diversity rating. However, information on enrollment and PDRR estimates are reported when reliable prevalence estimates are available.

We review the following reliable sources for disease-specific prevalence estimates: the Centers for Disease Control and Prevention website and the Global Burden of Disease database, a comprehensive epidemiologic dataset by country supported by the World Health Organization. In addition, a literature search is conducted to obtain peer-reviewed journal articles that estimate the prevalence of US disease by sex, age, race, and ethnicity. We recognize that there may be a lack of reliable disease-specific prevalence estimates for some conditions, particularly rare diseases. Trials of rare diseases with no reliable disease-specific prevalence estimate will not be rated on clinical trial diversity. Instead, a qualitative description of the demographic characteristics of participants in the clinical trial will be presented in this section. For other conditions (not considered rare) with no reliable disease-specific prevalence estimates, when appropriate, consideration would be given to comparing clinical trial participants to population estimates (US census demographic breakdown) and interpreting the finding accordingly.

We recognize the potential barriers for clinical trials conducted in other countries to reflect the racial and ethnic diversity of the disease population in the US. As such, for multinational clinical trials, only the subpopulation of patients recruited in the US will be rated on racial and ethnic diversity. Information on the racial and ethnic diversity of the overall patient population will only be provided for context as needed. We know that, in most cases, the baseline characteristics of the US subpopulation will not be published; therefore, this information will be requested as part of the routine data request sent to manufacturers. ICER will not accept such data as academic-in-confidence. If these data are not published and not provided to ICER as non-confidential data, we will focus on the racial and ethnic diversity of the entire trial population. Trials conducted exclusively in other countries will not be rated on race and ethnicity, as they are unlikely to be representative of the racial and ethnic diversity of the US population.

Synthesis of Results

We employ a transparent approach to evidence synthesis. Evidence is synthesized to help provide single best estimates and ranges of confidence that can help in evaluation of the comparative clinical effectiveness of interventions of interest. Syntheses also assist in understanding the limitations and gaps in the evidence base.

Following the identification of studies that meet our PICOTS criteria for a given evidence review, data from the studies are abstracted, and summarized in the text and in evidence tables of the evidence report. This summary is key to understanding the existing evidence base pertaining to the interventions and comparators of interest. Any key differences between the studies in terms of study design, patient characteristics, interventions (including dosing and frequency), outcomes (including definitions and methods of assessments), patient subgroups, and study quality are evaluated and described.

We examine the clinical and methodological characteristics of the set of studies reporting data for each outcome of interest and for each subpopulation with the goal of aggregating the results from the studies. When there is insufficient data or studies are judged to be too dissimilar for quantitative meta-analysis, we describe the results qualitatively in our evidence report and provide the key considerations for interpreting the results from the studies within the context of the evidence base.

When studies are sufficiently similar and report data that are appropriate for analysis, we conduct quantitative synthesis of the results across studies. Quantitative synthesis (e.g., meta-analysis) involves the use of a statistical method to pool results across multiple studies to generate the best estimate of the effect of the intervention on the outcome. In the absence of head-to-head studies comparing two interventions of interest, we often derive comparative evidence through quantitative synthesis methods that uses indirect comparisons (e.g., network meta-analysis, matching adjusted indirect comparisons), which may rely on common comparators or common predictors to link data from trials of the various treatments of interest.

The choice of the synthesis method ICER uses on a given topic depends on the research questions and the available evidence. In all reports, we provide the rationale for the choice of the synthesis method used and explicitly describe our methods.

Heterogeneity and Subgroups

ICER's reviews are not intended to guide individual shared decision-making between clinicians and patients and are not able to focus on the sorts of individual patient characteristics, values, and preferences that a skilled clinician would assess in making recommendations for a specific patient.

At the population level, data often show a range of responses to therapy with various distributions, including smooth normal distributions and sharply dichotomous outcomes. Heterogeneity of this sort may be unpredictable for individual patients but will still be highlighted in ICER reports as it can affect the assessment of therapies. For instance, a treatment that leads to a six-month increase in survival for all patients with no heterogeneity has different implications from a treatment that leads to a two-week increase in survival for 90% of patients and a long-term cure for 10% of patients. This is true even if it is currently impossible to know which patients will achieve each outcome.

In other cases, there is heterogeneity that is knowable *a priori*, based on patient characteristics prior to treatment. This may come in the form of characteristics that are effect modifiers, such that patients respond differently to treatment based on these knowable characteristics, even when treatment has net benefits for all patients. In many other cases, differences in baseline risk lead to groups that will achieve larger or smaller absolute benefits from therapy, even though the relative effect of therapy is the same across risk groups.

To ensure that our reviews focus on evaluating the most relevant subpopulations, we will include an *a priori* list of the subpopulations of interest and the scientific rationale for evaluating these subpopulations in the scoping document and research protocol. At the start of a topic, we will evaluate the current evidence base and consult with clinical experts, patients, patient groups, manufacturers, and other stakeholders to identify the most relevant subpopulations for the topic under review. In addition, we will consider race, sex, and age as presumptive subpopulations for every review. Information gathered during scoping may lead us to conclude that further consideration of subpopulations defined by these characteristics is not warranted or that additional information is needed to proceed. In such cases, our scoping document and research protocol will describe our rationale for not including these subpopulations.

In cases where effect modifiers or substantial differences in baseline risk would be known to clinicians and patients prior to treatment decisions, we highlight these differences in its discussion of the evidence. Depending on the nature of the evidence, the treatments, and the structure of the report, subgroups may be discussed in the greatest detail in individual outcomes sections of an ICER report or in a subsection called "Heterogeneity and Subgroups." In all reports, the subsection "Heterogeneity and Subgroups" will

be included to present the primary discussion of subgroup effects or highlight the other sections of the evidence review that discuss subgroup effects.

In cases where possible subgroup effects are encountered, we will evaluate the credibility of subgroup effect modification using the Instrument for assessing the Credibility of Effect Modification Analyses (ICEMAN for RCT).[12] The Instrument considers the following five key questions to evaluate the overall credibility of a subgroup claim:

1. *Was the direction of the effect modification correctly hypothesized* a priori*?*
2. *Was the effect modification supported by prior evidence?*
3. *Does a test for interaction suggest that chance is an unlikely explanation of the apparent effect modification?*
4. *Did the authors test only a small number of effect modifiers or consider the number in their statistical analysis?*
5. *If the effect modifier is a continuous variable, were arbitrary cut points avoided?*

Based on responses to these questions, we will judge the overall credibility of subgroup effect modification as "very low credibility", "low credibility", "moderate credibility", or "high credibility."

Subgroup differences may, on occasion, result in different evidence ratings for different subgroups. We will consider issuing different evidence ratings for a single intervention if robust, high-quality evidence supports substantial differences in the net benefit of the intervention across different populations or subgroups.

2.5. Judgment of Level of Certainty and Magnitude of Net Health Benefit: the ICER Evidence Rating Matrix™

Following synthesis of the evidence by quantitative and qualitative techniques, we assign overall evidence ratings to each of the interventions evaluated in its appraisal. A single intervention may be given more than one evidence rating if there are multiple comparators or if, as discussed above, there are substantial differences in the evidence ratings for a particular comparison across different populations or subgroups. Ratings reflect a judgment made at a moment in time and may be updated as new or additional evidence becomes available.

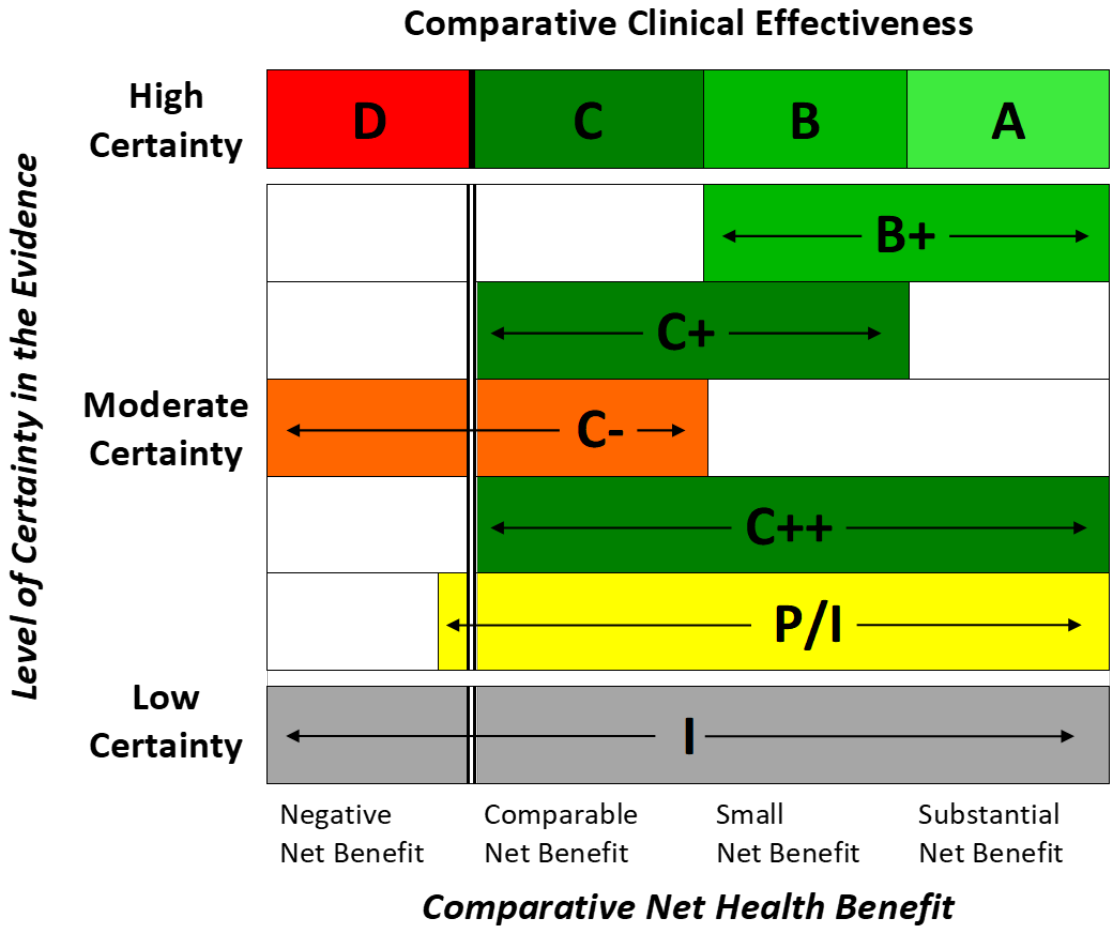
We developed the ICER Evidence Rating Matrix™ (see Figure 2.1) to evaluate the overall strength of evidence for a variety of outcomes. The evidence rating reflects a joint judgment of two critical components:

a) The **magnitude** of the difference between a therapeutic agent and its comparator in "net health benefit" – the balance between benefits and risks and/or adverse effects AND

b) The level of **certainty** in the best point estimate of net health benefit.[1,13]

The design of ICER'S Evidence Rating Matrix was informed by the approaches developed by the United States Preventive Services Task Force (USPSTF);[14] the international Grading of Recommendations Assessment, Development and Evaluation (GRADE) group;[15] and the Effective Healthcare Program of the Agency for Healthcare Research and Quality (AHRQ).[1,13,16] While each organization has developed unique criteria to rate the strength of evidence, each approach evaluates the entire body of evidence along a series of domains. The most important domains common to the four approaches include risk of bias, generalizability to "real-world" populations, consistency of findings across studies, directness (i.e., how closely the evidence measures the populations, interventions, and outcomes of interest), and precision.

Figure 2.2. 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 net health benefit is either comparable or inferior with high certainty of at best a comparable 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 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*

2.6. Appraisal Committee Voting on Comparative Clinical Effectiveness

Comparative clinical effectiveness votes are meant to capture the judgment of the appraisal committee on the adequacy of current evidence to demonstrate a *net health benefit* of a treatment versus a comparator for a population or subpopulation of patients. As part of the deliberation process, appraisal committee members formally weigh the relative magnitude of differences in risks and benefits, as well as the relative confidence that the body of evidence can provide regarding the accuracy of estimates of risks and benefits. The ICER report presents them with the detail on these elements as well as the summary rating in the ICER Evidence Rating Matrix. But this material is a prologue to the deliberation at the public meeting, and the subsequent vote taken by the appraisal committee.

Voting questions on comparative clinical effectiveness are typically framed using the following language:

> *For [patients with condition X], is the current evidence adequate to demonstrate that the net health benefit of [intervention A] is greater than that of [intervention B]?*
>
> *Yes / No*

The implication of a "yes" vote is relatively straightforward, indicating that there is adequate evidence to demonstrate one intervention is superior to another for the specified patient population. This can imply, among other possibilities, that clinicians may wish to consider using this therapy preferentially for such patients, and that insurers may wish to develop coverage policies that prioritize access for patients who are likely to benefit for intervention A over intervention B.

By contrast, the implications of a "no" vote may be less readily apparent. Such a vote does not necessarily mean that a treatment is ineffective relative to a given comparator; rather, it indicates that the current evidence base is insufficient to demonstrate incremental effectiveness in the specified patient population. In this way, a "no" vote can signal to stakeholders that additional research is necessary to determine an intervention's benefits and risks to patients.

3. Incremental Cost Effectiveness

To ensure consistency in analytic approaches across all of its reviews, we have defined a detailed Reference Case specifying the approach that ICER and our collaborators follow for cost-effectiveness analyses. The Reference Case details all the methods that ICER and our modeling collaborators follow when conducting the conventional base-case cost-effectiveness analysis. These methods generally follow the recommendations of the Second Panel on Cost-Effectiveness in Health and Medicine for the health care reference case,[17] and are also generally consistent with published guidance from international HTA organizations.[18,19] Following the Reference Case enables consistency in analytical approaches, but in specific cases, reasons may exist for deviating from the reference case. In such cases, the rationale for not fully applying the reference case methods will be clearly specified in the model analysis plan and Evidence Report.

Note that the description below provides guidance on ICER's Value Assessment Framework for health technology assessments (HTAs) in general. ICER's modifications to its methods for reviews of certain treatments for serious, ultra-rare disorders can be found here, and modifications for reviews of high-impact single or short-term therapies (SSTs) can be found here.

3.1. Overview

Cost-effectiveness analysis (also known as "economic modeling" or "decision analysis") helps to assess whether a technology is a good value for money in the long run by considering cost in relation to the clinical benefits provided and comparing one treatment and its associated care pathway to another. These comparisons are done through a simulated computer model of patient and cost outcomes of different care pathways.

The objective of the economic evaluation is to determine the incremental cost-effectiveness ratio, or the cost per unit of health benefit gained of one treatment over another. The unit of health gained can be a specific clinical outcome, such as an additional stroke prevented or a case of cancer diagnosed, or a more generalizable unit such as an additional year of life or an additional year of life adjusted for any changes in quality of life.

We use the equal value life year gained (evLYG)[20,21] as our usual measure of health gain due to its equal weighting of quality of life during life extension, a feature that eliminates concerns regarding the potential discriminatory effect of the cost per quality-adjusted life year (QALY). However, for benchmarking with U.S. and international academic standards, prior ICER reports, and broader international HTA work, we also will present results performed using cost per QALY gained as the primary measure of cost-effectiveness.[17] Lower incremental cost-effectiveness ratios represent better value for money. When the price of an intervention is known, the incremental cost-effectiveness ratio can be calculated. When the price is unknown (e.g. for an emerging treatment that has not yet received FDA approval), we will often use an

estimated price gained from analyst or other sources. We will also calculate the prices at which an intervention would hit certain cost-effectiveness threshold targets. For example, we calculate the prices needed to achieve $100,000 and $150,000 per additional evLYG and per QALY to provide different threshold prices for consideration by stakeholders.

All cost-effectiveness models must make some assumptions about how evidence on the short-term effects of care plays out in clinical and economic effects that happen many years in the future. We evaluate this uncertainty by varying the inputs to the model, first one at a time, and then systematically across all model inputs, to assess how robust the results are with different inputs.

3.2. Model Structure and Data Sources

We are committed to open and transparent engagement with stakeholders in the development of our economic models. To fulfill this commitment and explain the model approach in detail, we develops a model analysis plan following the publication of a revised scoping document. The model analysis plan outlines the methods the economic modeling team intends to employ, including information on the model structure and processes, all major inputs and sources of data, and key assumptions. In addition, the plan specifies whether the model is an adaptation of an existing model (with references as appropriate) or is being developed *de novo* for that HTA. The model analysis plan is published on the Open Science Framework (https://osf.io/7awvd/). The plan may be updated following review of additional data sources, discussions with stakeholders, and other activities.

In the model analysis plan and evidence report, the specific decision to be addressed by the analysis is specified in terms of the overall objective, the interventions and comparators, the relevant population groups and subgroups being considered, and the outcomes. Any differences in the population, intervention, or outcomes from the aims and structure of the clinical evidence review are documented with justifications. The analytic perspective (typically health care system) and time horizon (typically lifetime) used in primary analyses are also specified.

Following discussions with stakeholders and review of any additional data sources, the model analysis plan may be updated. The final version of the model used in conducting analyses is outlined in the Evidence Report, which is intended to provide enough information for an experienced researcher to be able to replicate the economic model and analyses.

Model Parameters and Data Sources

Model inputs, or "parameters," include those pertaining to intervention effectiveness, transition rates between health states, measurement and valuation of health states, resource use, and costs. Results from the evidence review, including the results from any meta-analyses, are used to inform input parameters when possible. All model parameters are described in the model analysis plan and evidence report, including risk equations as appropriate. We aim to use data from published or publicly available sources,

including peer-reviewed journals, supplementary appendices, briefing documents used by regulatory authorities, and conference proceedings. In specific instances, valid analyses may require the use of unpublished information, such as manufacturers' data on file.

Acceptance of Multiple Forms of Evidence

For comparative cost effectiveness, ICER's value framework does not limit the type of evidence to be considered to the results of randomized controlled trials (RCTs). When available, high-quality RCTs typically provide evidence on short to mid-term clinical benefits and more commonly occurring harms. When head-to-head trials have not been performed, indirect comparisons through formal network meta-analysis may be used as inputs for economic modeling. When benefits and harms occur over the course of many years, or when harms are rare but clinically important or even catastrophic, evidence from high quality published peer-reviewed studies using observational data and methodologies such as cohort studies, case-control studies, and long-term disease and drug registries may be used. Furthermore, if important patient reported outcomes have not been collected as part of a manufacturer's clinical development program, we will conduct a comprehensive literature review to identify published, peer-reviewed observational studies providing this information.

Real-World Evidence

Because inputs to economic models are often not included as outcomes in RCTS, the use and integration of evidence, based on observational or real-world data, has been an important source of model inputs and incorporated when appropriate in ICER cost-effectiveness analyses. RWE can be especially useful as a source of model inputs on transitional health states, concordance and persistence, costs, and health utilities, among others. The use of real-world data includes *de novo* evidence generation under certain circumstances where critical data elements for comparative cost effectiveness are lacking. This may include analyses of insurance claims data to better understand health states, resource utilization, and costs, or the analysis of new data from patient surveys to provide more direct information on health utilities.

Clinical Expert and Patient Input

For some economic models there will remain gaps in the available evidence despite review of published data and attempts to analyze or generate RWE. In these cases,we use input from clinical experts and/or patient groups to supply best estimates for the elements of a clinical care pathway, the likelihood of specific patient outcomes, and other inputs required to compare two or more treatments.

Data in Confidence

Because life science companies may have relevant information that is currently held in confidence, we have structured a process to accept and use such data. We allow manufacturers to submit data that is not yet in the public domain if the use of the information will be of help to the economic evaluation. ICER has specific protections in place for this confidential data, which are outlined at: https://icer.org/guidelines-on-icers-

acceptance-and-use-of-in-confidence-data-from-manufacturers-of-pharmaceuticals-devices-and-other-health-interventions/. In 2023, we announced changes to our academic-in-confidence policy. Going forward, academic-in-confidence data will be redacted from all external and public ICER documents until the earlier of: (a) publication or presentation of such data by the data owner or study investigators; (b) 12 months following the date of the public ICER meeting; or (c) for reports that are not subject to a public meeting, 12 months following report publication. Following any of these dates, ICER will unmask all redacted information from reports, presentations, and other public documents.

3.3. Measures of Health Gain

The sources and methods used for health preferences measurement are provided in the model analysis plan. These methods usually involve mapping health states in patients with a condition into a classification system with associated utility weights, such as the EQ-5D.[22,23] Generic classification systems such as the EQ-5D include measures of health state preferences that reflect those of the general US population, considered to be relevant to inform decisions at the population level (e.g., payer or health system formulary decisions) that involve individuals both with and without the condition of focus. Where general population estimates are not available or appropriate, utility estimates from different populations may be used, such as patients with the specific condition under study, those affected by similar symptoms, proxy respondents, or mixed samples. When there are challenges in translating outcome measures used in clinical trials or available patient-reported data into health states, the report discusses the rationale for choosing specific mapping algorithms.

Health effects are expressed in terms of total and incremental quality-adjusted life-years (QALYs), equal value life years gained (evLYG), life-years, and a condition-specific outcome achieved (e.g., treatment response, event avoided). We use the evLYG as our usual measure of health gain in calculations of our top health benefit price benchmark price due to its equal weighting of quality of life during life extension, a feature that eliminates concerns regarding the potential discriminatory effect of the QALY. 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 awards 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 population with a severe condition or disability, that treatment receives the same evLYG as a different treatment that adds a year of life for healthier members of the community. In certain situations, model structure may make the calculation of evLYG intractable, in which case we will report life years gained rather than evLYG. A broader discussion of the evLYG is now available in the peer-reviewed literature[21].

We will continue to include analyses in our reports using the QALY due to its position as a long-standing standard in the health economics landscape and because it may, on occasion, better reflect the preferences of patients themselves regarding the trade-off between quality of life and length of life presented by different treatments for patients with severe illness.

Presenting analyses using both the cost per evLYG and cost per QALY will enable policymakers to gain a broad overview of the cost-effectiveness of treatments while ensuring that results are available to demonstrate whether there is any impact of extended life at a lower quality of life. If ICER's analysis finds a major difference in these two measures, reports will include specific language describing the underlying characteristics of the treatment and the condition that lead to the difference. We participate in the global dialogue around the best methods for evaluating the value of health services and is always attuned to new developments that might provide a better and fairer system of measuring benefits across different kinds of interventions and patients.

Quantifying Additional Elements of Value

We have always presented a broad cost-effectiveness range for consideration by its appraisal committees and policymakers. The primary reason for presenting a range is to allow for formal consideration of potential benefits or disadvantages, as well as important ethical considerations, that cannot be fully captured in a quantitative manner in a cost-effectiveness model, or that could be captured but for which there are important reasons to consider these factors within a deliberative framework instead of "hard wiring" them into the cost-effectiveness results. Policy and ethical research continue on these questions of when and how to consider formal quantification of potential dimensions of value such as "value of hope," preferences for health gains among those with more severe illness, and scientific spillover effects.[24,25]

After ongoing consideration of the potential to perform quantitative analyses of these types of additional elements of value, we believe that there are still many methodological uncertainties and concerns that suggest it is more appropriate to consider these elements through deliberation focused on a broad cost-effectiveness threshold range. There are concerns related to potential double counting between these potential additional domains and the health gains captured by the evLYG/QALY. There is a lack of academic consensus on how to conceptualize or measure the value of hope, real option value, or scientific spillover effects. There are also concerns about whether these potential value domains, even if measurable, should be factored into the calculation of a fair price for drugs and other health care interventions, or whether these potential benefits should accrue to society as part of the broader social contract between society and health care innovators. Lastly, but very importantly, current opportunity cost thresholds are built on measurements of tradeoffs in health gain alone. It may be that a new health care intervention conveys additional benefits outside of health gain, but if these additional benefits are not also known for those services that would be lost due to opportunity cost effects, it seems most appropriate to assume that any additional benefits gained would be matched by those that are lost through opportunity cost effects.

With these considerations in mind, we believe it is most appropriate at the current time to continue to address additional potential value effects and modifiers explicitly through deliberation instead of trying to build in quantified approaches that would change the core health gain finding central to ICER's health benefit price benchmark. As discussed in greater detail later in this paper, quantified evLY or QALY shortfall calculations will be presented to the appraisal committees during deliberation and voting on the severity,

or “unmet need,” related to a condition, votes that feed into the ultimate appraisal committee votes on long-term value for money of treatments.

However, with this update to the Value Assessment Framework, we signal that we will begin a special focus in coming months on considering novel ways to quantify preferences related to severity, methods that often are framed as abandoning an assumption of a linear relationship between health gain value and replacing it with a formula that can capture risk aversion, severity, and the value of insurance. We will focus on exploring the Generalized Risk-Adjusted Cost-Effectiveness framework[26,27] and methods adopted by several international HTA programs that now weight health gains in relation to severity. In this effort to examine these methods, we will engage our Health Economics Council, Methods Advisory Group, and other researchers and stakeholders including international HTA bodies prior to testing the feasibility and impact of shifting to differentially weighting cost-effectiveness findings. We will also continue to monitor advances in methods as well as monitor changes made in the health technology assessment ecosystem on this topic. And, as a result of this special focus, ICER may entertain making an interim update to its Value Assessment Framework on this topic prior to the next overall update.

3.4. Impact on Distribution of Health Gains

Health inequality is an important concern for patients and policymakers in health systems across the globe. We have explored options for measuring the degree to which treatments may result in greater or lesser inequality across racial or socio-economic groups in the US. Data to support application of available methods are lacking in the US, and none of these methods have been adopted as standards within other HTA agencies such as the application of equity weights used in distributional cost-effectiveness analyses.

After piloting the Health Improvement Distribution Index (HIDI), we now routinely include the HIDI in reports and as a part of the public deliberation related to the potential benefit of some interventions to have a positive impact on health equity. As noted in our white paper on methods related to health equity[11], we will be careful to frame the HIDI as just one element in considering the potential impact on health equity. The HIDI is calculated as the disease prevalence in the subpopulation of interest divided by the disease prevalence in the overall population. A HIDI above 1.0 suggests that more health may be gained on the relative scale in the subpopulation of interest when compared to the population as a whole. Importantly, the HIDI will certainly not substitute the deliberative process that should integrate multiple important equity criteria in policy decisions, but it can serve as one example of a quantitative measure of the relative prevalence of the condition across key subpopulations. Additional evidence, qualitative or quantitative, on other health equity considerations (e.g., access, uptake, etc.) will remain important for deliberation.

3.5. Perspective

In each assessment, we continue to report cost-effectiveness results from both the health care system perspective as well as a modified societal perspective. Starting with the 2023 Value Assessment Framework, we will implement new methods to ensure that cost-effectiveness analyses done according to a modified societal perspective have "non-zero" inputs for impacts on productivity for the patient and caregivers, even when direct data are lacking.

We calculate incremental cost-effectiveness from the health care system perspective as our conventional base case, but also perform a modified societal perspective analysis that includes work productivity and other effects that may occur outside the health system. We use the health care system perspective as our conventional base case for several reasons. First, ICER's reports are primarily intended to inform population-based medical policy and pricing decisions within the US health care system. Employers, other plan sponsors, insurers, and risk-bearing provider groups in both private and public health insurance systems are not responsible for making trade-off decisions that involve broader societal resources, so the health care system perspective is the most directly relevant for decision-making. This is not to imply that plan sponsors, insurers, and others do not care about effects of health care interventions outside the health system. But their primary responsibility and the framework for the trade-offs they must manage rest within the health system.

Another reason that the health system perspective is favored for some consumers of ICER's research is that full consideration of the societal perspective often requires inclusion of broad and uncertain assumptions regarding the impact of health care not only on productivity, but on income tax generation, educational outcomes, the criminal justice system, and disability and social security benefits. Seeking to capture the full scope of these effects is practically almost impossible, and also raises the potential for unintended consequences, such as potentially favoring a selection of health care interventions that minimize the amount of time individuals spend receiving public financial support. A societal perspective raises several important ethical concerns of this nature, including whether interventions that support the health and productivity of younger – and healthier – individuals should be favored over interventions for those whose contributions to society cannot be equally measured through salaries, taxes paid, or independence from public services. We are sensitive to provide a framework for analyses that does not conflict with important ethical goals of US society.

The Second Panel on Cost-Effectiveness in Health and Medicine recommends reporting results from both the health care system perspective and the societal perspective, with an "impact inventory" used to make transparent which elements of a full societal perspective are included.[17] We follow this approach. To the extent feasible, the relative impacts of different care options on work productivity and other indirect impacts are estimated in the ICER report and are considered by ICER independent public appraisal committees as part of their weighing of "benefits beyond health," as described later in this paper. Each assessment will note by way of the "impact inventory," where direct evidence applicable to the modified societal perspective is available and is feasible to include. The additional direct evidence that is most

commonly assessed for inclusion in the modified societal perspective includes patient and transportation costs related to treatment, productivity for the patient and caregiver, and disutility for the caregiver.

There have been ICER reviews for which data unavailability has made it impossible to conduct an analysis from a modified societal perspective without the need to make unfounded assumptions or leave important elements as "zero." For example, it is not uncommon for there to be no data on the effects of new drugs on non-health sector costs such as patient and caregiver productivity and time seeking care. We recognize that these are only a subset of non-health impacts that could be considered in the societal perspective, but we have found them to be among the most influential in discussions about the broader value of new treatments.

When direct data are lacking, we use a method to capture the potential impacts of an intervention on patient productivity (formal and informal labor, household production, and time seeking care) and caregiver productivity time using an indirect approach. We believe that an indirect approach is appropriate and provides a reasonable balance between informing the potential impact of the treatment on broader outcomes and encouraging more research to be conducted to inform this broader perspective.

To inform estimates for the "non-zero" indirect approach, we use the published relationship between patient utility scores and US-based patient time use data[28] to derive the anticipated impacts of the treatment on time spent in each activity due to the disease and its management for the patient. The indirect approach values productivity time spent in a given health state, which is in contrast to the most typical approach of valuing productivity time lost, creating an opportunity to still capture productivity time lost during periods of non-life extension while estimating productivity time gained during periods of life extension. In these circumstances, and in line with the published literature,[28] we will include patient productivity time lost during non-life extension while estimating patient productivity time gained and patient consumption costs during periods of life extension. Since no parallel relationship between patient utility scores and caregiver time use data exists for the US setting, we assume that caregiver time spent is proportional to 75% of patient formal labor time lost. This estimate is based on the modeled relationship between caregiver time required[29] and patient time lost[30] according to patient utility scores in the United Kingdom setting. Although public feedback suggested concerns over the use of non-US evidence to estimate caregiver productivity impacts, we suggest that this approach may be updated to US evidence should it become available. Finally, the caregiver evidence remains consistent with the goals of the "non-zero" indirect approach. See the Reference Case for further details on the "non-zero" approach.

Modified Societal Perspective as Co-Base Case

To strike a balance between the ethical and other risks of a societal analysis and the potential interest of decision-makers in the results of analyses done in a modified societal perspective, we present a modified societal perspective as a co-base case for certain topics. When we judge that the societal costs of care for any disease are large relative to the direct health care costs, and that the impact of treatment on these costs is substantial (i.e., there are substantial differences in the cost-effectiveness findings between the two

perspectives), the modified societal perspective is included as a co-base case, so that threshold-based pricing estimates from both perspectives will be included in ICER's health benefit price benchmarks. This co-base case designation will most often occur in cases where the incremental cost-effectiveness ratio changes by greater than 20%, greater than $200,000 per evLYG or QALY, and/or when the result crosses thresholds of $100,000-$150,000 per evLYG or QALY based on direct evidence.

Due to added uncertainty and to promote generation of direct evidence to inform the modified societal perspective, in such instances where indirect evidence on patient or caregiver productivity time is used (by way of implementing the "non-zero" approach), we will not promote the modified societal perspective to be a co-base case.

3.6. Discounting

To account for time value and ensure comparability across studies, all economic models use constant-rate discounting of both costs and outcomes, at the rate of 3% per year. Discounting is a standard method in economic modeling, and in the US, the standard approach has been 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.[31] 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. The use of a single, uniform discount rate for all assessments allows for consistent comparisons across different or prior evaluations.

3.7. Patient Populations

To the extent possible, the patient populations included in ICER's economic evaluations are generally those for which the evaluated interventions are indicated. However, at the time of evaluation, the only available data on efficacy may come from trials that do not reflect the (likely) indicated population. In such cases, the discrepancy between the indicated population and the trial populations will be pointed out, along with discussion of the relevance of trial results to the larger population. While cohort models tend to reflect homogeneity in patient populations for whom health technologies are assessed, when relevant, ICER's evaluations include scenarios with different patient subgroups to account for the heterogeneity within patient groups within a specific disease area.

Heterogeneity and Subgroups

Evidence Reports include a sub-section on "Heterogeneity and Subgroups" in order to broaden discussion of heterogeneity and subgroups within the patient population. ICER's Reference Case calls for the inclusion of different subpopulations when analyzing the cost-effectiveness of health technologies, to the extent possible. Consistent with our consideration of subgroup analysis in ICER's white paper on health equity,[11] when subpopulations are clearly defined *a priori* by clinical characteristics, it is often an important goal to examine relative cost-effectiveness of treatment in these subpopulations. However, analyses focused on

subpopulations defined solely by race/ethnicity or socioeconomic status are vulnerable to confounding clinical variables, raising the risk of misinterpretation of results.[32] Therefore, in each report we will provide the rationale for why we performed or avoided cost-effectiveness analyses of subpopulations defined by characteristics other than appropriate clinical markers of risk or outcome.

Data permitting, subgroup analyses of appropriate clinical markers of risk or outcome will be considered for patient groups that could be of interest either clinically or economically. Such subgroup analyses have been and will continue to be undertaken when we believe 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.

3.8. Costs

Costs are reported in terms of total and incremental costs. When possible, we use estimates of prices net of discounts, rebates, and other price concessions as the conventional base-case input for prices used in cost-effectiveness and potential budget impact analyses. Analyses using wholesale acquisition cost (WAC) prices are also included for context. To provide pricing that can reliably and with relative transparency provide an estimate for net prices in the US market, we collaborate with SSR Health LLC, a consultancy which combines data on net US dollar sales with information on unit sales to derive net pricing at the unit level across all payer types. Further details on the mechanism used to estimate net prices are available in ICER's Reference Case.

ICER's cost-effectiveness analyses will not routinely make estimates of price changes across comparator treatments linked to patent and exclusivity time horizons, due to the uncertainty of predicting whether or when interventions will have generic competition, and the magnitude of any price change following generic introduction. However, when generic competition is imminent, and if we believe it is reasonable to assume a substantial price decrease within 12-24 months, this price decrease may be considered for the base case scenario or for a scenario analysis.

3.9. Conventional Base Case and Cost-Effectiveness Thresholds

In the presentation of the conventional base-case results of incremental cost-effectiveness analyses, health benefits and costs are summarized as incremental cost per evLYG, cost per QALY gained, cost per life-year gained, and cost per condition-specific measure of clinical benefit. ICER will provide cost-per-evLYG and per QALY results at $50,000, $100,000, $150,000 and $200,000 for all assessments, including those for treatments of ultra-rare disorders. The range for threshold prices used to determine ICER health benefit price benchmarks remains $100,000-$150,000 per evLYG and per QALY, reflecting ICER's judgment that recent research [33] confirms that both opportunity cost and willingness to pay paradigms produce estimates of an operational cost-effectiveness threshold at approximately $100,000 per QALY. As described later in this paper, we will usually select the broad range from $100,000 per QALY to $150,000 per evLYG for our health benefit price benchmark, but rarely we will only use evLYG threshold pricing if there are notable

differences due to acknowledged reduced valuation of life extension with QALY results. Because ICER's suggested health benefit price benchmarks are most often used as ceiling prices, we continue to use $150,000 per evLYG as a liberal upper bound to allow for ample integration of benefits beyond health and consideration of special ethical priorities (see Section 3.13). In 2019, following a webinar series featuring leading health economists, we developed a paper analyzing different approaches to determining operative cost-effectiveness threshold ranges for the U.S. This paper has not been updated but still reflects the basic perspective that ICER takes in using an opportunity cost approach.

ICER's Evidence Reports present a broader range of results symmetrically around this range, from $50,000-$200,000 per evLYG and per QALY. This range is meant to accommodate the needs of decision-makers in the US to think about their own desired interpretation of cost-effectiveness thresholds while considering uncertainty, benefits beyond health, and special ethical priorities.

3.10. Sensitivity and Scenario Analyses

As a method to evaluate uncertainty in the economic evaluation, the Evidence Report also includes one-way sensitivity analyses, presenting the results in "tornado diagrams" that display the findings across a feasible range for each input parameter estimate and a table containing the ranges and distributional assumptions around the input parameters varied. Expected values of costs and outcomes for each intervention are also estimated through probabilistic sensitivity analyses, which characterizes some of the uncertainty in the input parameter estimates. This type of analysis takes repeated samples, typically 1,000 or more, from the simultaneous distribution of all key model input parameters; results are presented tabularly in terms of the percentage of simulations that achieve $50,000, $100,000, $150,000, and $200,000 per evLYG thresholds, and graphically using scatter plots or cost-effectiveness acceptability curves (CEAC) which reflects the percentage of simulations that result in incremental cost-effectiveness ratios that fall at or under various cost-effectiveness thresholds.

Unless otherwise noted, all sensitivity analyses and scenario analyses use the health care system perspective conventional base-case findings as the starting point.

Scenario Analyses

Specific scenario analyses (including one using a modified societal perspective that incorporates estimates such as productivity losses, caregiver burden, and other indirect costs) and subgroup analyses are conducted when appropriate. In addition, the report presents results from threshold analyses which estimate the intervention prices that correspond to cost-effectiveness thresholds extending from $50,000 per evLYG gained to $200,000 per evLYG gained.

Shared Savings Scenarios

The below language is also found in Section 5.1 of the Adapted Value Assessment Methods for High-Impact "Single and Short-Term Therapies" (SSTs) (updated December 2022).

To stimulate further consideration of how the cost offsets generated by new treatments should be incorporated in calculations of the value and threshold-based price for a new treatment, we will present two new economic analyses that evaluate cost-effectiveness outcomes with a different approach to the cost offsets from a new treatment. These two analyses will be considered for all high-impact SSTs under review, as well as other (non-SST) treatments with relevant and substantial potential cost-offsets:

1. A 50/50 shared savings model in which 50% of the lifetime health system cost offsets from a new treatment are "assigned" to the health system instead of being assigned entirely to the new treatment; and

2. A cost-offset cap model in which the health system cost offsets generated by a new treatment are capped at $150,000 per year but are otherwise assigned entirely to the new treatment.

Threshold analyses for treatment price may be presented and may be considered as guides to ICER's pricing if the following two criteria are met:

i. a large percentage of the traditional value-based price comes from cost offsets of comparator (e.g. standard of care) therapy
ii. comparator therapy price is not known to meet common cost-effectiveness thresholds.

Under circumstances where the above two pricing criteria (I. and II.) are satisfied, we may present ranges from one of the SST shared savings calculations as the most policy-relevant for the recommended health benefit price benchmark range (Section 3.13).

Outcome-Based Payment Arrangements

When relevant, Evidence Reports include information from manufacturers and payers to model a scenario analysis including a limited number of outcome-based payment arrangements for the intervention under review. In some cases, these payment arrangements can be a useful tool in managing uncertainty and increasing the ultimate cost-effectiveness of treatment. We actively seek information from manufacturers and payers about the potential outline of outcomes-based contracts for scenario analyses in our reports. In cases where the list price of the treatment is known but there is no guidance from stakeholders, we will perform an exploratory scenario analysis using outcomes and levels of financial risk-sharing that could help manage uncertainty.

Exclusion of Non-Intervention Costs

In cases where an intervention that increases health gains would not be found to be cost effective, even with a zero-dollar price, a separate scenario analysis excluding non-intervention health care costs will be presented. By non-intervention health care costs, we mean all costs except those directly tied to administering the intervention or other modeled treatment options (i.e. comparator costs). We have encountered specific situations, rarely, 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 such 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. In such cases a scenario analysis excluding health state costs that are not related to the intervention *per se,* may be informative.

Dynamic Pricing Scenario

If policies in the Inflation Reduction Act are implemented consistent with current law, we believe it is reasonable to assume that the net price increase for Medicare for many drugs will not be above inflation after launch. In addition, many of these drugs are likely to fall under the provisions for Medicare price negotiation after nine years post-approval for small molecule products and after 13 years post-approval for biological products. For example, for many small molecule drugs, the price after nine years may be assumed to drop to at least 75% of the launch net price. The price of any active comparator may also become subject to Medicare negotiation as it hits nine or 13 years after approval, requiring attention to the possibility of mandatory price decreases at a future time point for not only interventions but also active comparators in an ICER review.

Prior to changing our approach to include price dynamics within a mandated scenario analysis, we commit to engaging our Health Economics Council, Methods Advisory Group, and other researchers and stakeholders including international HTA bodies to test the feasibility and impact of how best to include pricing dynamics within cost-effectiveness analyses. Although academic contributions are emerging in the dynamic pricing arena including those published as a themed section on prescription drug pricing at Value in Health,[34] best practices across health technology assessment entities do not exist. Further, public comments received on this topic supported additional deliberation on the methods prior to implementing them in ICER's Value Assessment Framework. We are willing to make updates to ICER's Value Assessment Framework on this topic if and when engagement and testing support making a change.

3.11. Validation and Calibration

All economic models are validated prior to conducting analyses, as well as during the production of the Evidence Report. Validation entails assessing whether a model has been implemented correctly (internal validation) and if its assumptions and results are in line with the current evidence and expectations (face validity, external validation). The specific approach to internally validate the model during development is detailed in the Model Analysis Plan. After the posting of the Model Analysis Plan and a presentation of model structure, assumptions, and inputs, key stakeholders also provide feedback on the model assumptions, parameters, structure, and overall face validity. In addition, we release economic model files and code to manufacturer stakeholders willing to agree to confidentiality and privacy restrictions, allowing participating stakeholders to include detailed critique of the model in public comments submitted on the Draft Evidence Report. Further, all models are published in the ICER Analytics Interactive Modeler following publication of the final ICER report.

Calibration entails assessing if the model inputs and outputs are consistent with known scenarios. Any calibration procedures used during model development are proposed in the Model Analysis Plan, including the calibration target (and source), the goodness-of-fit metric, and criteria for judging fit. Results from the calibration procedure are presented in the Evidence Report.

3.12. Uncertainty and Controversies

Evidence Reports include a sub-section on "Uncertainty and Controversies" in order to broaden discussion of alternative model structures and assumptions suggested by manufacturers or other stakeholders. One important goal of this section is to provide further elaboration of the rationale behind methodological decisions that underpin the conventional base case. This sub-section also serves 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.

To accomplish this goal, the sub-section provides discussion of different model variations that could be viewed as more conservative or optimistic. In particular, this sub-section addresses alternative model structures or inputs suggested by manufacturers or other stakeholders that differ importantly from the conventional base case. This sub-section also consolidates and expands 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 are also included in this sub-section, pointing out differences in model structure, inputs and assumptions, and the impact of these differences on model results. We review and compare the current model to published models that included the same interventions or comparators of interest, were developed in the last 10 years, and were similar to the current model from a setting and population perspective.

3.13. Health Benefit Price Benchmarks

For all assessments, an ICER "health benefit price benchmark" is developed for the new intervention, which reflects prices aligned with commonly-cited long-term cost-effectiveness thresholds ranging from $100,000 to $150,000 per evLY and per QALY gained. The prices represent the prices paid by insurers, net of rebates and other concessions, that would be required to reach these cost-effectiveness thresholds. Further information on the justification for the cost-effectiveness threshold range used for the health benefit price benchmarks is given below.

ICER's health benefit price benchmarks suggest a price range that aligns with a treatment's added benefits for patients over their lifetime. Prices at or below these thresholds meet an "opportunity cost" requirement that the health benefits gained by patients using new treatments are not outweighed by

health losses due to long-term cost pressures that lead individuals to delay care, abandon care, or lose health insurance.

We remain committed to emphasizing an opportunity cost perspective in determining appropriate cost-effectiveness thresholds for decision-making. Within this paradigm, academic work[33] suggests a top threshold at approximately $104,000 per QALY based on direct health losses within the health system perspective. The calculations underlying this estimate do not include consideration of the significant negative effects of self-rationing caused by increasing insurance premiums and health care costs for those who retain insurance coverage. If additional elements of value are quantified as benefits of new interventions, there is also the risk of unmeasured equal or greater losses within these same elements among those individuals who drop insurance coverage. We also note that consideration of health equity would suggest that individuals with lower incomes experience a disproportionate share of the harms from the opportunity costs imposed by increasing health insurance premiums. Working from these insights, although we are not changing our effective threshold range for price benchmarks at the current time, we will pursue further discussion with academic experts and stakeholders to consider the impacts on opportunity-cost thresholds of including additional elements of value within the cost-effectiveness findings and/or including severity or risk-adjustments to the cost-effectiveness findings. Other HTA groups and academic research[35] shares the view that with added elements of value comes the need to consider lowering the opportunity-cost threshold for decision-making. Finally, if severity or risk-adjustments are made within future versions of ICER's cost-effectiveness findings, then we will pursue feedback on the use of one primary threshold (e.g. $100,000 per evLYG) rather than providing a broad range as a price benchmark.

We believe that there is a confluence of results between research exploring opportunity cost thresholds and willingness to pay thresholds in the US setting. For conceptual and ethical reasons we favor a view of thresholds based in an opportunity cost paradigm. Claxton has presented data analyses supporting the adoption of cost-effectiveness thresholds in the UK, US, and other countries that are far lower than traditional thresholds, given the marginal productivity of the respective health care systems.[36-38] For the US, Claxton estimates an opportunity cost threshold of approximately $30,000-$40,000 per QALY.[38] More recently, Vanness has estimated health opportunity costs for direct purchase private plans in the US,[33] and produced an estimate of $104,000 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 $104,000 per QALY produces a net loss of health due to its impact on premium increases and thereby loss of insurance and the attendant negative health effects, especially among lower income members of the insurance pool.

In the US market-based system with multiple payers, there is a case for multiple thresholds based on willingness-to-pay which may differ by payer type (e.g., government vs. commercial insurance).[39] However, there are broad requirements across the US health care system to fund all "medically necessary" care. There is also 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.

Despite the lack of an explicit overall budget for health care in the US, the current environment of the US health care system is one in which policy-makers sense that the opportunity cost for current spending is already substantial, and that real harm is being done as health care costs continue to rise. We believe that anecdotal evidence and testimony from these policymakers further supports ICER's decision to apply an opportunity cost approach to a threshold range, the goal being to ensure that the prices paid for health gains from effective new treatments are aligned with the magnitude of those health gains such that greater health is not lost through the effects of rising health costs at the system and societal level.

Reflecting on the most recent conceptual and empirical research, a case could be made for reducing our health benefit price benchmark range to $50,000-$100,000 per evLYG. However, the top end of the price benchmark range is usually interpreted as a "ceiling" price beyond which a treatment will be viewed as not cost-effective. There is also value in retaining a consistent threshold range as a level playing field for all stakeholders, especially as an incentive for future innovation. Therefore, we continue to use the cost-effectiveness range of $100,000 to $150,000 to support health benefit 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 benefits beyond health and special ethical priorities.

We will continue to evaluate which cost-effectiveness thresholds should be used to generate health benefit price benchmarks to reflect ongoing academic work that may support a different threshold range and may update these thresholds prior to the next Value Assessment Framework update.

4. Benefits Beyond Health and Special Ethical Priorities

4.1 Conceptual Overview of Benefits Beyond Health and Special Ethical Priorities in Deliberation on Long-Term Value for Money

The inclusion of explicit domains of value labeled "benefits beyond health" and "special ethical priorities" are critical features of the ICER Value Assessment Framework. These elements of the framework force the ICER appraisal committees and all external stakeholders to consider broader domains of value than those that are core to clinical trial evidence and cost-effectiveness modeling. All too often what matters most to patients is poorly captured in the available clinical trial data. Sometimes this occurs because surrogate outcome measures do not reflect true patient-centered outcomes; but even when trials do capture the clinical outcomes that matter most to patients, there are other aspects of value related to the complexity of the treatment regimen or the impact of care options on the ability to return to work, on the negative impact of the condition on family and caregivers, on public health, or on other aspects of the health system or society. The ICER value framework identifies these "benefits beyond health" as important elements of any overall judgment on long-term value for money, and all ICER reports have separate sections in which evidence and information pertaining to these elements are presented.

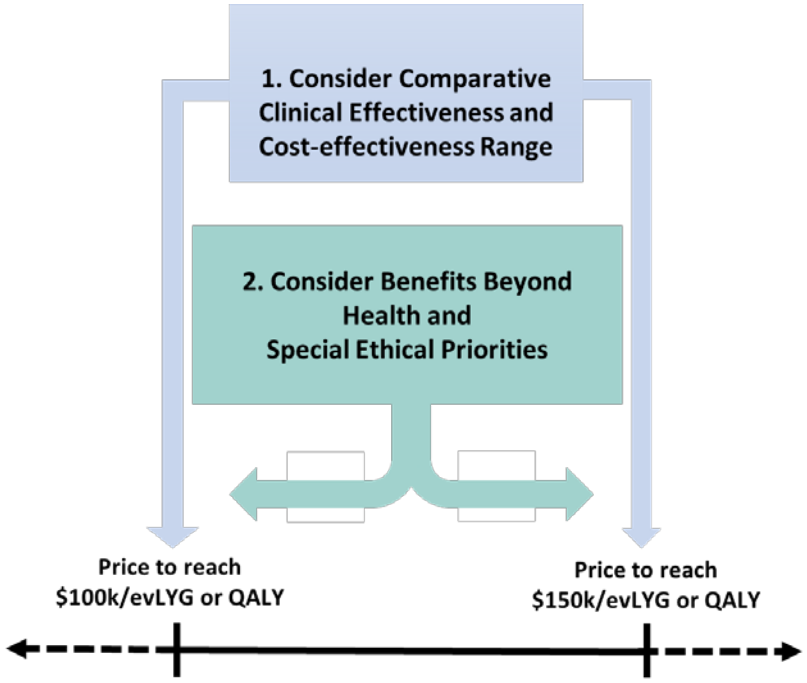
Similarly, decisions about value do not happen in a vacuum. Stakeholders may have special ethical priorities related to the severity, or "unmet need" related to the condition. Similarly, societies have an ethical priority to give some degree of preference to interventions that can provide health gains that reduce historical disparities in outcomes often due to discrimination of one kind or another. The ICER value framework includes this domain of value and it is explored in a separate section of each ICER report.

Researchers and policymakers continue to explore different ways to identify specific benefits beyond health and special ethical priorities and apply them in a formal quantitative fashion to weight health gains or to adjust cost-effectiveness thresholds. However, current methods for algorithmic integration of these factors carry important risks. Attempts to measure benefits that accrue to patients in their ability to fulfill their life goals, or to achieve greater personal dignity, are likely to represent some degree of double counting of the benefits captured by conventional measures of health gain, such as the evLYG or QALY. Some potential benefits beyond health, such as the scientific "spillover" of new science to other treatment areas, cannot easily be distinguished from the spillover of investments in other areas of health or societal wellbeing. And routine quantitative inclusion of productivity gains raise the specter of discrimination against people with chronic disabilities who may never achieve a health status that allows them to contribute as much traditional economic output as others. These examples demonstrate that the general impulse to recognize and account for benefits beyond health and special ethical priorities needs to be tempered by methodological and ethical concerns.

Thus, it is not surprising that most health technology assessment groups around the world do not attempt to quantify these domains of value, believing that consideration of these domains is essential in a judgment of value, and yet should be left qualitative and integrated into decision-making through public deliberation.

We have considered over many years the evolving methodological options for quantitative or mixed-methods approaches to enhance the explicit integration of these considerations in value assessment. Formal multi-criteria decision analysis (MCDA) has been considered but rejected because we do not believe that the methods for weighting individual elements are robust enough to add to reliability of value judgments. We have attempted formal MCDA with ICER's independent committees on several occasions in the past and found the technique too complicated for reliable use. Based on discussions with stakeholders, benchmarking other value frameworks around the world, and the input of public comment, ICER reports will continue to use a variation on MCDA that makes benefits beyond health and special ethical priorities explicit and gives clear guidance on their relevance to judgments of value, but that does not attempt an overly facile quantification. Decision-makers will be given guidance, however, that consideration of these factors should guide part of their thinking about how to use the cost-effectiveness threshold range, with higher ends of the range more applicable when there are important positive contributions related to these factors, and lower ends of the range reflecting relatively less consequential added value considerations. Figure 4.1 below summarizes this conceptual approach to integration of benefits beyond health and special ethical priorities into considerations of long-term value for money of a health intervention.

Figure 4.1. Conceptual Guide to Application of "Benefits Beyond Health" and "Special Ethical Priorities" to Judgements of Value



evLYG: equal value life-years gained, QALY: quality-adjusted life years

Categories of Benefits Beyond Health and Special Ethical Priorities

We will include information in every report relevant to the four potential benefits beyond health and special ethical priorities described below.

Patients: There is substantial unmet need despite currently available treatments.

Prior versions of the ICER Value Assessment Framework have attempted different approaches to capturing the relative "severity" of the condition. Some consideration of the severity of the condition is viewed by many academics and stakeholders as an important consideration in judgments of the value of treatment, but international HTA agencies have conceptualized this idea 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 society or decision-makers.

We believe that we can gain greater clarity and consistency in consideration of this special ethical priority by moving to consider whether there is substantial "unmet need" despite currently available treatments. In pilot testing this question with a subgroup of appraisal committee members we found that this framing was the easiest way for them to express their view of how severity should influence the thinking about the value of a new treatment. To inform the appraisal committee vote on this issue we will have perspectives from patients and patient groups in the ICER report, and we will also calculate absolute and proportional health shortfalls.

The absolute 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.[40] 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 evLYs lost, the absolute evLY 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 evLYs should merit some increased prioritization. The ethical consequences of using absolute evLY 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, will score highest on the scale of absolute evLY shortfall. The Norwegian health technology assessment program is perhaps the most notable organization currently using measures of absolute shortfall as a component in their appraisal process.[41]

Absolute shortfalls are often viewed in contrast to another way to empirically measure a sense of severity of illness, or "need" as the Dutch have called it.[42] This alternative measure is called a proportional shortfall. The proportional evLY shortfall is measured by calculating the proportion of the total evLYs of remaining life expectancy that would be lost due to untreated illness.[43] The proportional evLY 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 evLY shortfall, rapidly fatal conditions of childhood will have

high proportional evLY shortfalls, but high numbers can also often arise from severe conditions among older adults who may have only a few years left of average life expectancy but would lose much of that to the illness without treatment.

Absolute and proportional shortfalls 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 unmet need despite current treatment options.

Caregivers: The treatment is likely to produce substantial improvement in caregivers' quality of life and/or ability to pursue their own education, work, and family life.

The effects of treatment on caregivers are modeled in the modified societal perspective analysis that is done as part of every ICER report. These results can provide a quantitative estimate of the impact of treatment but we believe it is also helpful to include qualitative statements on these effects from patients and their families. In addition, to reinforce the importance of these effects on broader judgments on value, we will have a separate vote on this "benefit beyond health."

Health Equity: The condition is of substantial relevance for disadvantaged communities.

Health Equity: The treatment offers a substantial opportunity to improve access to effective treatment by means of its mechanism of action or method of delivery.

We will address two different questions related to health equity, the first of which asks whether the condition is of particular relevance to one or more disadvantaged communities. ICER reports will include information on this question gathered from patients, and the Final Report will include further testimony given at the public meeting by patient groups and clinical experts. The quantitative HIDI score will also be featured as one element of understanding whether the prevalence of a particular condition within a particular subpopulation of concern is far higher than the prevalence in the overall population.

The second question related to health equity will address whether a treatment's mechanism of action or method of delivery offers a "substantial opportunity" to improve equity through improved access. ICER reports will include information to evaluate whether treatments could offer improved access compared to other available treatments due to less complex regimens, including new treatments that offer options for at-home treatment as opposed to treatment in provider facilities.

Other: As determined by the ICER team based on input from patients, clinical experts, and appraisal committee members.

We may, on occasion, add additional voting questions under the Benefits Beyond Health and Special Ethical Priorities section based on input from:

- All stakeholders: During the public comment period that follows after publishing the Draft Evidence Report, we also publish a draft version of the voting questions that are being prepared for the Public Meeting. Stakeholders may include suggestions for additional voting questions in their public comment response (in addition to commenting on the proposed questions).
- Appraisal committee members: After publishing the Draft Evidence Report, the ICER research team meets with the Chair of the independent appraisal committee to review the voting questions. After that meeting, the Chair reaches out to the committee to elicit their feedback on the voting questions, including whether they would like to suggest adding any new questions.

We may also add an additional voting question specifically for situations involving ultra-rare diseases. Per our framework adaptation for treatments for ultra-rare diseases, we may include a question to seek evidence and perspective on the potential impact of the new treatment on the entire "infrastructure" of care, including effects on screening for affected patients, on the sensitization of clinicians, and on the dissemination of understanding about the condition that may revolutionize how patients are cared for in many ways that extend beyond the treatment itself. A recent example of this addition can be found in the Metachromatic Leukodystrophy Public Meeting and Final Report[44].

Appraisal Committee Voting on Benefits Beyond Health and Special Ethical Priorities

Appraisal committee voting on benefits beyond health and special ethical priorities will be structured as a Likert scale on "agreement" with a statement on whether the treatment has a substantial impact on each element. For each question the appraisal committee will be asked to vote from 1-5, with the following structure: 1=Strongly Disagree; 2=Disagree; 3=Neutral; 4=Agree; 5=Strongly Agree.

We believe that this voting structure will create results that are easy for all stakeholders to interpret and will enhance their ability to integrate these considerations into final decisions regarding value for money.

It is important to note that many key aspects of our approach to integrating key elements beyond the clinical and cost-effectiveness data into our reports and public meetings will remain the same. We will continue to provide guidance to our appraisal committees and to health care decision-makers that consideration of benefits beyond health and special ethical priorities should guide part of their thinking about how to use the cost-effectiveness threshold range, with higher ends of the range more applicable when there are important positive benefits and/or priorities, and lower ends of the range reflecting relatively less consequential added value considerations. In ICER public deliberation meetings,

independent appraisal committees will continue to take votes on each specific element so they can be highlighted for decision-makers.

Methods to integrate these elements into HTA reports and public deliberation are one of the most active areas of research in applied health economics, and we will continue to monitor this literature, participate and lead research when feasible, and remain ready to seek improvements to our methods in an iterative fashion. In particular, we are monitoring the research and conceptual analyses on modifying health gains quantitatively in relation to some measure of severity and risk aversion. As noted earlier, we do not feel that this research is mature enough to suggest adoption of a formal quantitative metric by which the health gains of some conditions would be upgraded – or downgraded – depending on the level of severity. As we continue to explore these approaches, however, we remain firm in our commitment to make deliberative consideration of these aspects of value judgments transparent and consequential to decision-making.

5. Long-Term Value for Money

Each appraisal committee will be asked to integrate all value dimensions into its vote on long-term value for money at current pricing. As noted in Section 1.4, the ICER Value Assessment Framework asks appraisal committees and health care decision-makers to integrate comparative clinical effectiveness, incremental cost-effectiveness, benefits beyond health, and special ethical priorities into overall judgments of value, consistent with recommendations by the Second Panel on Cost-Effectiveness in Health and Medicine.

When appraisal committees are asked to vote on long-term value for money at current pricing, they will be reminded of this mental task of integrating all value dimensions. Cost-effectiveness analyses in the ICER report will have been done using established prices for interventions if those prices have been announced by the developer, or estimated prices (a "placeholder" price) based on analyst forecasts and/or guidance from other sources. If the price for an intervention has not yet been announced but estimated prices are felt to be reasonably accurate and reliable, we will ask appraisal committees to vote on the intervention's long-term value for money assuming the estimated price.

When voting on the long-term value for money for therapies, the voting question to appraisal committees will be posed as follows:

> *For a specified patient population, given the available evidence on comparative clinical effectiveness and incremental cost effectiveness, and considering benefits beyond health and special ethical priorities, what is the long-term value for money of [intervention A] compared to [intervention B] at current (or assumed) pricing?*
>
> *High… / Intermediate… / Low Long-Term Value for Money at Current or Assumed Prices*

6. Potential Budget Impact Analysis

6.1. Overview

While it is important to understand how expensive a new technology is for a given unit of benefit, it is also important to look at the technology in terms of short-term financial impact to the overall health care system. We analyze the short-term potential budget impact of changes in health expenditures with the introduction of a new test, treatment, or delivery system process. The potential budget impact is an estimate of the projected cumulative resource expenditure across all elements of the health care system for a specific intervention in a specific population over a period of time. We use a five-year timeframe for its potential budget impact analysis to capture important potential clinical benefits and cost offsets provided by newer care options. Potential budget impact models aim to quantify the net cost over a short period of time for all eligible patients to receive the new technology.

For pharmaceuticals, the results of the budget impact analysis are compared to a national annual threshold for a new drug that is tied to growth in the overall US economy. This threshold, calculated by ICER, is updated each calendar year using the most recent inputs available to reflect changes in US gross domestic product, medical and pharmaceutical spending, and the average annual number of drugs approved by the FDA over the last five years. The current potential budget impact threshold calculations are detailed at https://icer-review.org/methodology/icers-methods/icer-value-assessment-framework-2/.

This comparison is intended to signal to stakeholders and policymakers when a new treatment, even one priced at a level commensurate with good long-term value, may add short-term health care costs that are so substantial that they would be difficult for the health care system to absorb over the short term without displacing other needed services or contributing to rapid growth in insurance costs that could threaten sustainable access to high-value care for all patients. We seek to include information for estimating short-term potential budget impact but also to use clinical expert testimony to identify when intended clinical use of a new treatment may be at a scale that would trigger access and affordability concerns. In such cases, the goal is to prompt discussions of possible policy steps to alleviate potential access restrictions or sudden sharp increases in health insurance premiums. The role of the potential budget impact analysis is not to suggest a cap on spending, but to signal to the health care system that special arrangements, such as lower prices, enhanced efforts to eliminate waste, or prioritizing treatment for the sickest, may be needed to ensure availability of the new drug without short-term adverse effects on patients and families seeking to pay for affordable health insurance.

6.2. Methods

The cost-effectiveness model in each economic evaluation is used 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 four threshold prices to achieve cost effectiveness at $50,000, $100,000, $150,000, and $200,000

per QALY. 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 or other aspects of treatment. The potential health care system budgetary impact of the intervention is explored over a five-year time horizon.

Potential budget impact analyses are based on net cost per patient across all sectors of health care spending, not just drugs. We use epidemiologic and other data to estimate the size of the potential candidate population for each new treatment. For each threshold price, 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 analysis indicates when the potential budget impact threshold is reached at each combination of price and percent uptake among eligible patients at five years. This analysis does not attempt to estimate the uptake of a new intervention. Rather than try to estimate real-world uptake, the analysis presents information on a national level that allows stakeholders to ascertain the potential budget impact of a new service given a range of prices. 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.

Evidence Reports note the percent uptake of a new intervention, at its net price level, that would produce a potential budget impact that exceeds this threshold, or that a new intervention will not exceed the threshold regardless of uptake level. Results of the analysis are presented as 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. The graph allows readers to see the average potential budget impact for a single patient over various time horizons from one to five years, and 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. We also seek to produce calculations that will help policymakers 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 five 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. The analysis uses clinical expert opinion regarding the treatments likely to be displaced by use of a new treatment within the eligible population. The procedures used in the analysis vary depending on whether and how many existing treatments are being displaced, with more details provided in ICER's Reference Case document. 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 policymakers to make their own assumptions. In addition, the budget impact

model is published as part of ICER Analytics' Interactive Modeler, which allows individual users to compare different uptake scenarios, at the conclusion of every review.

The potential budget impact threshold for new drugs is calculated as double the average net budget impact for new drugs that would contribute to overall health care cost growth beyond the anticipated growth in national GDP plus an additional 1%. See Table 6.1. for the template for deriving the annual potential budget impact threshold. For services other than new drugs, potential budget impact is estimated but not compared to a potential budget impact threshold.

Table 6.1. Template for Annual Potential Budget Impact Threshold Calculation

Item	Parameter	Source
1	Growth in US GDP + 1%	World Bank
2	Total personal medical care spending	CMS National Health Expenditures
3	Contribution of drug spending to total health care spending (%) (Row 4 ÷ Row 2)	Calculation
4	Contribution of drug spending to total health care spending	CMS National Health Expenditures, Altarum Institute
5	Annual threshold for net health care cost growth for ALL drugs (Row 1 x Row 4)	Calculation
6	Average annual number of new molecular entity approvals over 5 years	FDA
7	Annual threshold for average cost growth per individual new molecular entity (Row 5 ÷ Row 6)	Calculation
8	Annual threshold for estimated potential budget impact for each individual new molecular entity (Doubling of Row 7)	Calculation

6.3 Access and Affordability Alert

Within the potential budget impact analysis section of each final report, we will include a subsection discussing whether ICER is issuing an "Access and Affordability alert". The goal of the alert is to signal that the additional health care costs introduced by a new intervention may be difficult for the health care system to absorb over the short term. In this situation, other needed services may be displaced, or health care insurance costs may rapidly rise, which would threaten sustainable access to high-value care for all patients.

In addition to highlighting the potential budget impact of the intervention when using the upper end of ICER's Health Benefit Price Benchmark range (HBPB) ($150,000 per evLY), we will also report the budget impact of the therapy at either estimated net pricing, or the placeholder price if a net price estimate is not yet available. Our goal in reporting both estimates (the placeholder or net price, as well as the threshold price) is to determine whether there would be significantly different impacts on affordability at these various prices. For example, we will call attention to situations where the alert would be triggered by the placeholder price or net price, but not by the threshold price. In other situations, the alert may be triggered (or not) regardless of which price is used, and this will be described as well.

References

1. Ollendorf DA, Pearson SD. An integrated evidence rating to frame comparative effectiveness assessments for decision makers. *Medical care.* 2010;48(6 Suppl):S145-152.
2. CER Collaborative, Academy of Managed Care Pharmacy (AMCP), National Pharmaceutical Council (NPC), International Society for Pharmacoeconomics Research (ISPOR). Comparative Effectiveness Research Tool. https://www.cercollaborative.org/global/default.aspx?RedirectURL=%2fhome%2fdefault.aspx. Accessed October 11, 2019.
3. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med.* 1997;126(5):376-380.
4. Higgins J, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions.* 2nd ed. Chichester (UK): John Wiley & Sons; 2019.
5. Moher D, Liberati A, Tetzlaff J, DG A. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine.* 2010;8(5):336-341.
6. Pearson SD, Dreitlein WB, Towse A, Hampson G, Henshall C. A framework to guide the optimal development and use of real-world evidence for drug coverage and formulary decisions. *Journal of Comparative Effectiveness Research.* 2018;7(12):1145-1152.
7. Garrison LP, Jr., Neumann PJ, Erickson P, Marshall D, Mullins CD. Using real-world data for coverage and payment decisions: the ISPOR Real-World Data Task Force report. *Value Health.* 2007;10(5):326-335.
8. Administration USFaD. FDA Patient-Focused Drug Development Guidance Series for Enhancing the Incorporation of the Patient's Voice in Medical Product Development and Regulatory Decision Making. https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical. Published 2023. Updated 04/06/2023. Accessed2023.
9. Administration USFaD. Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints For Regulatory Decision-Making. In:2023.
10. US Preventive Services Task Force. *Procedure Manual.* 2008.
11. Agboola F, Whittington MD, Pearson SD. Advancing Health Technology Assessment Methods that Support Health Equity. 2023.
12. Schandelmaier S, Briel M, Varadhan R, et al. Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses. *Cmaj.* 2020;192(32):E901-E906.
13. Ollendorf DA, Pearson SD. ICER Evidence Rating Matrix: A User's Guide. Institute for Clinical and Economic Review. https://icer-review.org/methodology/icers-methods/icer-evidence-rating-matrix/. Published 2020. Updated January 31, 2020. Accessed January 31, 2020.
14. Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D. Update on the methods of the U.S. Preventive Services Task Force: estimating certainty and magnitude of net benefit. *Ann Intern Med.* 2007;147(12):871-875.
15. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ (Clinical research ed).* 2008;336(7650):924-926.
16. Agency for Healthcare Research and Quality. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/cer-methods-guide_overview.pdf. Published 2014. Accessed.

17. 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.
18. Guidelines for the economic evaluation of health technologies: Canada. In: 4th ed. Ottawa: CADTH; 2017: https://www.cadth.ca/sites/default/files/pdf/guidelines_for_the_economic_evaluation_of_health_technologies_canada_4th_ed.pdf.
19. National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal 2013. nice.org.uk/process/pmg9. Published 2013. Accessed January 27, 2020.
20. Institute for Clinical and Economic Review. The QALY: Rewarding the Care That Most Improves Patients' Lives. ICER. https://icer-review.org/wp-content/uploads/2018/12/QALY_evLYG_FINAL.pdf. Published 2018. Accessed January 27, 2020.
21. JD C, MD W, SD P. An Alternative Measure of Health for Value Assessment: The Equal Value Life-Year. *Pharmacoeconomics.* 2023;Epub ahead of print.
22. Pickard AS, Law EH, Jiang R, et al. United States Valuation of EQ-5D-5L Health States Using an International Protocol. *Value Health.* 2019;22(8):931-941.
23. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Medical care.* 2005;43(3):203-220.
24. 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 [3]. *Value Health.* 2018;21(2):131-139.
25. Lakdawalla DN, Doshi JA, Garrison Jr LP, Phelps CE, Basu A, Danzon PM. Defining elements of value in health care—a health economics approach: an ISPOR Special Task Force report [3]. *Value in Health.* 2018;21(2):131-139.
26. Lakdawalla DN, Phelps CE. The Generalized Risk-Adjusted Cost-Effectiveness (GRACE) Model for Measuring the Value of Gains in Health: An Exact Formulation. *Journal of Benefit-Cost Analysis.* 2023:1-24.
27. Lakdawalla DN, Phelps CE. Health technology assessment with risk aversion in health. *Journal of health economics.* 2020;72:102346.
28. Jiao B, Basu A. Associating Health-Related Quality-of-Life Score with Time Uses to Inform Productivity Measures in Cost-Effectiveness Analysis. *PharmacoEconomics.* 2023:1-13.
29. Rowen D, Dixon S, Hernández-Alava M, Mukuria C. Estimating informal care inputs associated with EQ-5D for use in economic evaluation. *The European Journal of Health Economics.* 2016;17:733-744.
30. Mukuria C, Rowen D, Hernández-Alava M, Dixon S, Ara R. Predicting productivity losses from health-related quality of life using patient data. *Applied Health Economics and Health Policy.* 2017;15:597-614.
31. Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG. *Cost-effectiveness in health and medicine.* New York: Oxford University Press; 2016.
32. Avanceña AL, Prosser LA. Examining equity effects of health interventions in cost-effectiveness analysis: a systematic review. *Value in Health.* 2021;24(1):136-143.
33. Vanness DJ, Lomas J, Ahn H. A health opportunity cost threshold for cost-effectiveness analysis in the United States. *Annals of internal medicine.* 2021;174(1):25-32.
34. Themed Section: Prescription Drug Pricing. *Value in Health* 2023;26(3):317-391.
35. Vallejo-Torres L. The Broader Opportunity Costs in the Broader Cost-Effectiveness Analysis Framework. *Applied Health Economics and Health Policy.* 2023:1-12.
36. 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.

37. Lomas JRS, Martin S, Claxton KP. Estimating the marginal productivity of the English National Health Service from 2003/04 to 2012/13. In. York, UK: Centre for Health Economics, University of York; 2018.
38. 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 economics.* 2018;27(6):1017-1023.
39. Pauly MV. The Questionable Economic Case for Value-Based Drug Pricing in Market Health Systems. *Value Health.* 2017;20(2):278-282.
40. Arrow KJ. *Essays in the theory of risk-bearing.* Vol 121: North-Holland Amsterdam; 1974.
41. Analysis. UBoE. Personal Income per Capita [A792RC0A052NBEA]. *In.* 2018;Vol 2019: Federal Reserve Bank of St. Louis.
42. Commission. MHSCR. Maryland's Total Cost of Care Model: Update for Providers.
43. Commission MHP. Health Care Cost Growth Benchmark. 2019.
44. Lin G SK, Fahim SM, Carlson JJ, Richardson M, Herce-Hagiwara B, Raymond F, Dickerson R, Pearson SD, Rind DM. Atidarsagene Autotemcel for Metachromatic Leukodystrophy. Institute for Clinical and Economic Review. https://icer.org/assessment/metachromatic-leukodystrophy-2023/#timeline. Published 2023. Updated October 30, 2023. Accessed.

Appendices

A. Glossary

Appraisal committee: ICER convenes public meetings of three regionally-focused appraisal committee (New England Comparative Effectiveness Public Advisory Council [CEPAC], Midwest CEPAC, and California Technology Assessment Forum [CTAF]) to review objective evidence reports and develop recommendations for how stakeholders can apply evidence to improve the quality and value of health care. The mission, processes, and role of the CEPAC and CTAF programs are the same, despite a different naming convention.

Conventional base-case analysis – the analysis using the initial set of assumptions and input parameter values, detailed in the Reference Case.

Budget impact – an estimate of the projected cumulative resource expenditure for a particular intervention in a specific population over a period of time.

Cost-effectiveness acceptability curve – a graph that plots the percentage of simulations that result in incremental cost-effectiveness ratios that fall at or under different cost-effectiveness thresholds.

Clinical effectiveness – the degree of health benefit produced by an intervention.

Comparator – an alternative health technology against which an intervention is evaluated.

Cost-effectiveness analysis – a type of economic evaluation in which an outcome is measured in incremental costs per incremental health unit, such as life years gained, or clinical event avoided.

Cost-effectiveness threshold – the maximum amount of money a decision-maker is willing to pay to ensure that the health benefits gained by patients using new treatments are not outweighed by health losses due to long-term cost pressures.

Direct comparison – An evaluation of two interventions that have been assessed head-to-head.

evLYG analysis – An analysis that 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 awards 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.

Health benefit price benchmarks – the treatment prices that would achieve incremental cost-effectiveness ratios of $100,000 and $150,000 per QALY or evLY gained.

Health technology assessment (HTA) – the systematic evaluation of evidence related to any healthcare intervention that can be used to improve health and prevent and treat disease; HTAs inform policy- and decision-making surrounding the use of such interventions.

Incremental cost-effectiveness ratio – the ratio of the difference in costs between two possible interventions, divided by the differences in their effectiveness.

Indirect comparison – an evaluation of two interventions via one or more common comparators.

Meta-analysis – a type of statistical analysis that combines data from multiple studies assessing the same two interventions and generates a pooled, summary estimate of the relative effect of one treatment versus a comparator.

Net health benefit – the balance between benefits and risks and/or adverse effects.

Network meta-analysis – an extension of pairwise meta-analyses to include many interventions and generate a series of pooled, summary estimates of the relative effect of each treatment versus each comparator.

Observational study – a non-experimental study in which investigators draw inferences about what is observed without trying to influence the outcome of the study; types of observational studies include cohort, cross-sectional, case-control and ecological studies.

One-way sensitivity analysis – a method of analysis in which the value of one model input parameter is varied at a time to assess the effect of the parameter on results.

Opportunity cost – the value of something that must be foregone in order to acquire something else.

Parameter – a characteristic that influences the output of a model.

PICOTS – population, intervention, comparator, outcomes, timing, and setting; ICER uses these items as a framework for defining the scope of its appraisals.

PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PRISMA is a set of criteria that guide the conduct and reporting and systematic reviews and meta-analyses.

Probabilistic sensitivity analysis – a method of analysis used to account for parameter uncertainty in which values for input parameters are sampled based on pre-specified probability distributions.

Quality-adjusted life-year (QALY) – a measure of health benefit that accounts for changes in both quantity (e.g., mortality) and quality of life.

Randomized controlled trial – a type of study design in which participants are allocated at random into intervention and control groups.

Reference case – the framework of methods that ICER follows when conducting the conventional base-case cost-effectiveness analysis.

Scenario analysis – a type of analysis that estimates results using alternative model assumptions.

Sensitivity analysis – a method of analysis in which model inputs are varied in order to determine how such changes affect the results.

Subpopulation – a subset of a larger population.

Systematic review – a literature review that identifies and summarizes the results of all empirical studies meeting pre-defined eligibility criteria.

Threshold analysis – a type of sensitivity analysis in which the values of model input parameters are varied in order to determine the value that produces a specific result (e.g., a given cost-effectiveness value).

Time horizon – the period of time over which outcomes are evaluated.

Tornado diagram – a graphical depiction of the results of one-way sensitivity analyses in which the analyses with the greatest impact on model results are displayed with the largest bars and are stacked at the top of the chart.

Utility – a measure of preference for a health outcome.

Validity – the assessment of whether a model has been implemented correctly (internal validity) and if its assumptions and results are in line with the current evidence and expectations (face validity).

Value assessment framework – a decision support tool intended to guide stakeholders in making decisions that will promote sustainable access to high-value care for all patients.

B. List of Major Revisions to 2020 Framework

Comparative Clinical Effectiveness
ICER's reports will include a new subsection called "Clinical Trial Diversity". This section will present information on the demographic diversity of participants in the clinical trials and rate trials on overall diversity based on the following demographic characteristics: race/ethnicity, sex, and age, specifically, adults aged 65 and older.
To ensure that our reviews focus on evaluating the most relevant subpopulations, ICER will include an *a priori* list of the subpopulation(s) of interest and the scientific rationale for evaluating these subpopulations in the scoping document and research protocol.
ICER will use a formal credibility assessment tool (ICEMAN for RCT: Instrument for Assessing the Credibility of Effect Modification Analyses) to evaluate and present information on the credibility of subgroup findings.
Long-Term Cost Effectiveness
ICER will continue to report cost-effectiveness results from both the health care system perspective as well as a modified societal perspective. ICER will implement new methods to ensure that cost-effectiveness analyses done according to a modified societal perspective have "non-zero" inputs for impacts on productivity for the patient and caregiver(s), even when direct data are lacking. In such instances where indirect evidence on patient or caregiver productivity time is used (by way of implementing the "non-zero" approach), ICER will not promote the modified societal perspective to be a co-base case in terms of its policy relevance or its inclusion in ICER's health benefit price benchmarks.
To support tangible consideration of severity as a potential modifier of the value of health gains, ICER will regularly calculate QALY and evLY shortfall measures to accompany primary cost-effectiveness analysis results and will include these findings in material presented during public deliberation by appraisal committees on the long-term value for money of treatments.
After piloting the Health Improvement Distribution Index (HIDI), ICER will continue to calculate this measure and include it in reports and as a part of the public deliberation related to the potential benefit of some interventions to have a positive impact on health equity.
ICER will continue to frame our health benefit price benchmarks based on analyses using the QALY at the $100,000 threshold and the evLYG at the $150,000 threshold. However, ICER will emphasize that policymakers who prefer or who may be mandated to consider only measures of health gain other than the quality-adjusted life year (QALY) can find results at every threshold based solely on the equal value of life-years gained (evLYG).
ICER's Reference Case will be revised to reflect the proposed and adopted revisions.
Potential Budget Impact Analysis
Starting in June 2025, ICER added a fourth estimate of a threshold price ($200,000 per QALY or evLY) to the potential total budgetary impact analysis results.
ICER revised and updated the language for the Access and Affordability Alert in June 2025.
Voting Questions
ICER will change the terms used to describe elements previously called "Potential Other Benefits" and "Contextual Considerations" to "Benefits Beyond Health" and "Special Ethical Priorities." The structure of the ICER report will continue to highlight these elements in a separate section.
ICER will change the voting categories and framework for "Benefits Beyond Health" and "Special Ethical Priorities" to address overlap and misinterpretation in certain areas and to seek to create a more actionable set of voting results on key elements. The voting structure will be changed to a "Strongly Disagree" to "Strongly Agree" five-point Likert scale on four elements. Votes in the public meeting will be presented as average scores across all voting appraisal committee members.
On January 23, 2025 additional language was added to Section 4 of this version to describe how additional voting questions may be included under the Benefits Beyond Health and Special Ethical Priorities categories. These voting questions may be added as determined by the ICER team based on input from patients, clinical experts, and appraisal committee members.

C. Summary of Ultra-Rare Diseases (URD) and Single and Short-Term Therapies (SST) Adaptations to Value Assessments

Table C1. Summary of URD and SST Adaptations.

Review Materials	URD	SST
Draft and Revised Scope	We propose to assess a treatment under an adaptation of the ICER Value Framework for treatments of serious, ultra-rare conditions if we believe they meet the following criteria: • The eligible patient populations for the treatment indication(s) included in the scope of the ICER review is estimated at fewer than approximately 10,000 individuals. • There are no ongoing or planned clinical trials of the treatment for a patient population greater than approximately 10,000 individuals. Following formal public comment and discussions with stakeholders, ICER will make a final decision on whether the therapy meets these criteria and will be assessed using an adapted approach.	We propose to assess a treatment under an adaptation of the ICER Value Framework for treatments of high-impact "single and short-term therapies" (SSTs), if we believe they meet the following criteria defined as: • The therapy is delivered through a single intervention or a short-term course (less than one year) of treatment that offers a significant potential for substantial and sustained health benefits extending throughout patients' lifetimes. • The therapy can eradicate a disease or condition, or produce sustained major health gains that can halt the progression of significant illnesses. Following formal public comment and discussions with stakeholders, ICER will make a final decision on whether the therapy meets these criteria and will be assessed using an adapted approach.
Data Request	Data request to include a table requesting information on important manufacturing and/or research and development costs.	N/A
Research Protocol	N/A	N/A
Model Analysis Plan	N/A	Conduct the following scenario analyses:

		• 50/50 shared savings in which 50% of lifetime health care cost offsets from a new treatment are assigned to the health care system instead of being assigned entirely to the new treatment • Cost-offset cap in which health care cost offsets generated by a new treatment are capped at $150,000 per year but are otherwise assigned entirely to the new treatment • A) Optimistic and B) conservative assumptions regarding the benefit of treatment, to be presented as highlighted scenarios • When the SST price is known or can be estimated, assessments of SSTs will also include a scenario with a threshold analysis determining the duration of beneficial effect (e.g. cure) for those patients receiving short-term benefit that would be needed to achieve standard cost-effectiveness thresholds (e.g., $150,000/evLYG).
Section 3.3. (Evidence Matrix)	ICER will provide specific context regarding the potential challenges of generating evidence for these treatments, including considerations of challenges to conducting RCTs, to validating surrogate outcome measures, and for obtaining long-term data on safety and on the durability of	N/A

	clinical benefit. Evaluating treatments for ultra-rare diseases against historical controls will be highlighted.	
Section 4. (Long Term Cost-Effectiveness)	• For assessment of cost-effectiveness of a treatment for ultra-rare diseases, ICER will seek to produce a cost-effectiveness model for every new treatment, acknowledging and highlighting additional uncertainty in translating patient outcomes into quality-adjusted life year (QALY) or equal value of life year gained (evLYG) measures. • When there are challenges translating the outcome measures used in clinical trials and available patient-reported data into QALYs, ICER will conduct a search for "mapping" studies that may allow translation of surrogate outcomes into quality of life measures. The validity of these mapping studies will be discussed with manufacturers, clinical experts, the patient community, and other stakeholders in order to get their input on the most feasible way to translate these other measures of patient outcome into evLYGs or QALYs.	ICER will make cure proportion modeling its standard reference case for high-impact SSTs whenever relevant, but to address uncertainty we will also provide survival analysis based on other modeling approaches when feasible.
Section 4. (WTP thresholds)	For all treatments, including those for ultra-rare diseases, ICER will	N/A

	provide willingness-to-pay threshold results from $50,000 per QALY/evLYG to $200,000 per QALY/evLYG. No special quantitative weighting system will be applied to different magnitudes of evLY/QALY gains or to baseline severity of the condition.	
Section 4.3. (Base-Case Results)	When the impact of treatment on patient and caregiver productivity, education, disability, and nursing home costs is substantial and these costs are large in relation to health care costs, ICER will present its base case health system perspective model results in tandem with the results of a scenario analysis inclusive of broader societal costs. This will most often occur in cases where the incremental cost-effectiveness ratio changes by greater than 20%, greater than $200,000 per evLYG/QALY, and/or when the result crosses thresholds of $100,000-$150,000 per evLYG/QALY.	N/A
Section 4.3. (Scenario Analyses) and E5 in the Supplement.	N/A	Conduct the following scenario analyses: • 50/50 shared savings in which 50% of lifetime health care cost offsets from a new treatment are assigned to the health care system instead of being assigned entirely to the new treatment • Cost-offset cap in which health care cost offsets generated by a new treatment are capped at $150,000 per year but are otherwise assigned entirely to the new treatment

		• A) Optimistic and B) conservative assumptions regarding the benefit of treatment, to be presented in conjunction with the base case • When the SST price is known or can be estimated, assessments of SSTs will also include a scenario with a threshold analysis determining the duration of beneficial effect (e.g. cure) for those patients receiving short-term benefit that would be needed to achieve standard cost-effectiveness thresholds (e.g., $150,000/evLYG or QALY).
Section 6. (Health Benefit Price Benchmarks)	ICER will calculate a health benefit price benchmark for these treatments using the standard range from $100,000 to $150,000 per QALY/evLYG, but will add language in all report formats indicating that decision-makers in the US and in international settings often give special weighting to other benefits and to contextual considerations that lead to coverage and funding decisions at higher prices, and thus higher cost-effectiveness ratios, than applied to decisions about other treatments.	Threshold analyses for treatment price may be presented and may be considered as guides to ICER's pricing if the following two criteria are met: I. large percentage of the traditional health benefit price comes from cost offsets of comparator (e.g. standard of care) therapy II. comparator therapy price is not known to meet common cost-effectiveness thresholds. Under circumstances where the above two pricing criteria (I. and II.) are satisfied, ICER may present ranges from one of the SST shared savings calculations as the most policy-relevant for the recommended health benefit price benchmark range.

Voting Questions	During public meetings of ICER's independent appraisal committees, votes on the "long-term value for money" of treatments for serious ultra-rare conditions will follow the same approach as other interventions by having appraisal committee votes on value regardless of the base-case results (i.e., even if results exceed $200,000 per QALY/evLYG).	N/A
Public Meeting Evidence Presentation	List applicable adaptation on Methods Overview slide	• List applicable adaptation on Methods Overview slide • Present applicable scenario analyses on Scenario Analyses slide
Policy Recommendations	N/A	N/A
ICER Analytics Materials	N/A	N/A

evLYG: equal value of life years gained, ICER: Institute for Clinical and Economic Review, N/A: not applicable, QALY: quality-adjusted life year, SST: single and short-term therapies, URD: ultra-rare disease, WTP: willingness to pay