

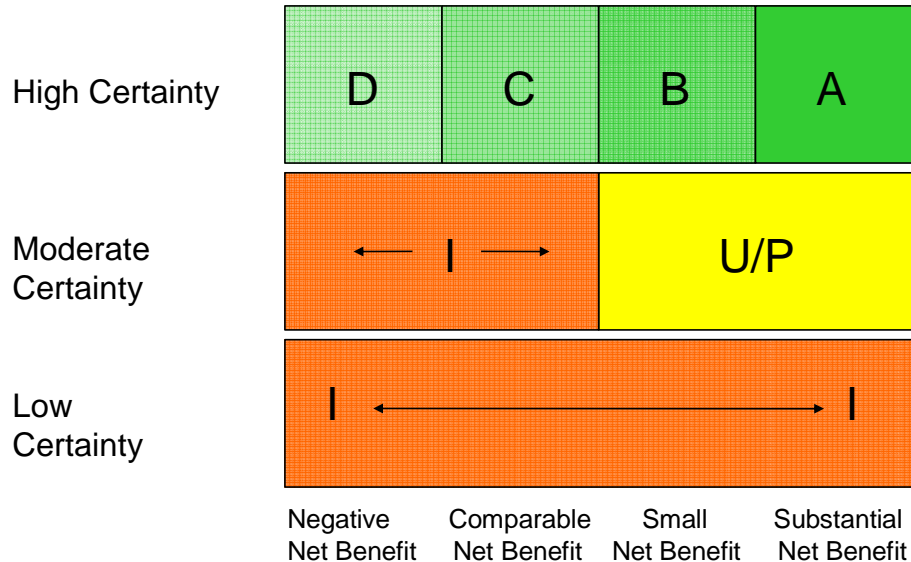


Methodology: ICER Integrated Evidence Rating™

Comparative Clinical Effectiveness

The ICER Integrated Evidence Rating™ combines a rating for comparative clinical effectiveness and a rating for comparative value. The clinical effectiveness rating arises from a joint judgment of the level of certainty provided by the body of evidence and the magnitude of the net health benefit – the overall balance between benefits and harms. This method for rating the clinical effectiveness is modeled on the “Evidence- Based Medicine (EBM) matrix” developed by a multi-stakeholder group convened by America’s Health Insurance Plans. This matrix is depicted below:

Comparative Clinical Effectiveness Comparing tech _____ vs. _____



A = “Superior” [High certainty of a 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 = “Inferior” [High certainty of an inferior net health benefit]

U/P = “Unproven with Potential” [Moderate certainty of a small or moderate-large net health benefit]

This category is meant to reflect technologies whose evidence provides:

- 1) High certainty of *at least* comparable net health benefit
- 2) Moderate certainty suggesting a small or moderate-large net health benefit

I = “Insufficient” The evidence does not provide high certainty that the net health benefit of the technology is at least comparable to that provided by the comparator(s).

Certainty

The vertical axis of the matrix is labeled as a degree of certainty with which the magnitude of a technology’s comparative net health benefit can be determined. This operational definition of certainty thus is linked to but is not synonymous with the overall validity, consistency, and directness of the body of evidence available for the assessment. ICER establishes its rating of level of certainty after deliberation by the Evidence Review Group, and throughout ICER follows closely the considerations of evidentiary strength suggested by the Effective Health Care program of the Agency for Health Research and Quality (AHRQ) (www.effectivehealthcare.org) and the GRADE working group (www.gradeworkinggroup.org).

High Certainty:

An assessment of the evidence provides high certainty in the relative magnitude of the net health benefit of the technology compared to its comparator(s).

Moderate Certainty:

There is moderate certainty in the assessment of the net health benefit of the technology. Moderate certainty implies that the evidence is limited in one or more ways so that it is difficult to estimate the net health benefit with precision. ICER’s approach considers two qualitatively different types of moderate certainty. First, there may be limited certainty in the magnitude of any net health benefit, but there is high certainty that the technology is *at least* as effective as its comparator(s). The second kind of moderate certainty applies to those technologies whose evidence may suggest comparable or inferior net health benefit and for which there is not high certainty that the technology is at least comparable. These two different situations related to “moderate certainty” are reflected in the matrix by the different labels of “Unproven with Potential” and “Insufficient.”

Limitations to evidence should be explicitly categorized and discussed. Often the quality and consistency varies between the evidence available on benefits and that on harms. We follow the GRADE and AHRQ approaches in highlighting key types of limitations to evidence, including:

- a. Internal validity
 - i. Study design
 - ii. Study quality
- b. Generalizability of patients (directness of patients)
- c. Generalizability of intervention (directness of intervention)
- d. Indirect comparisons across trials (directness of comparison)
- e. Surrogate outcomes only (directness of outcomes)
- f. Lack of longer-term outcomes (directness of outcomes)
- g. Conflicting results within body of evidence (consistency)

Low Certainty:

There is low certainty in the assessment of net health benefit and the evidence is insufficient to determine whether the technology provides an inferior, comparable, or better net health benefit.

Net Health Benefit

The horizontal axis of the comparative clinical effectiveness matrix is “net health benefit.” This term is defined as the balance between benefits and harms, and can either be judged on the basis of an empiric weighing of harms and benefits through a common metric (e.g., Quality Adjusted Life-Years, or “QALYs”), or through more qualitative, implicit weightings of harms and benefits identified in the ICER appraisal. Either approach should seek to make the weightings as explicit as possible in order to enhance the transparency of the ultimate judgment of the magnitude of net health benefit.

Whether judged quantitatively or qualitatively, there are two general situations that decision-making groups face in judging the balance of benefits and harms between two alternative interventions. The first situation arises when both interventions have the same types of benefits and harms. For example, two blood pressure medications may both act to control high blood pressure and may have the same profile of toxicities such as dizziness, impotence, or edema. In such cases a comparison of benefits and harms is relatively straightforward. However, a second situation in comparative effectiveness is much more common: two interventions present a set of trade-offs between overlapping but different benefits and harms. An example of this second situation is the comparison of net health benefit between medical treatment and angioplasty for chronic stable angina. Possible benefits on which these interventions may vary include improved mortality, improved functional capacity, and less chest pain; in addition, both acute and late potential harms differ between these interventions. It is possible that one intervention may be superior in certain benefits (e.g. survival) while also presenting greater risks for particular harms (e.g. drug toxicities). Thus the judgment of “net” health benefit of one intervention vs. another often requires the qualitative or quantitative comparison of different types of health outcomes.

Since net health benefit may be sensitive to individual patient clinical characteristics or preferences there is a natural tension between the clinical decision-making for an individual and an assessment of the evidence for comparative clinical effectiveness at a population level. ICER approaches this problem by seeking, through the guidance of its scoping committee, to identify a priori key patient subpopulations that may have distinctly different net health benefits with alternative interventions. In addition, the ICER appraisal will also seek to use decision analytic modeling to identify patient groups of particular clinical characteristics and/or utilities which would lead them to have a distinctly different rating of comparative clinical effectiveness.

The exact boundary between small and moderate-large net benefit is subjective and ICER does not have a quantitative threshold. The rating judgment between these two categories is guided by the deliberation of the Evidence Review Group.

Comparative Value

There are three categories of value: high, reasonable or comparable, and low. The ICER rating for comparative value arises from a judgment that is based on multiple considerations. ICER does not employ a single measure of cost-effectiveness for assignment of comparative value, nor does it rely on a formal threshold for determination of the level of value. Instead, comparative value is informed by multiple measures of potential economic impact, including:

- Impact on service use (e.g., tests, hospitalizations)
- Cost to reduce adverse outcomes (e.g., cost per hospitalization averted)
- Cost to achieve clinical success (e.g., cost per curative outcome)
- Cost per life year gained
- Cost per quality-adjusted life year (QALY) gained
- Budget impact per 1,000 diseased individuals
- System issues (e.g., manpower tradeoffs to invest in new technology)

The advantages for evaluating the full list of economic measures are twofold. First, the importance of these measures varies for individual stakeholders. For example, payers may be most interested in expressions of the clinical value achieved for the additional investment provided (e.g., cost per QALY, cost per event averted), while integrated health systems may ascribe most importance to measures of budgetary or system impact, and patients may be most interested in differential rates of downstream testing or other service use. Second, sole reliance on traditionally-accepted measures of cost-effectiveness such as cost per QALY may mask important considerations in evaluating whether to adopt a new technology. Cost-effectiveness findings may appear to be “reasonable” based on widely-used thresholds (e.g., \$50,000 per QALY gained), when in reality the incremental investment required is for an imperceptible clinical gain.

ICER has developed a method for presenting multiple measures of economic impact together in a format known as the Comparative Value Evidence Table (CVET), which allows for visualization of economic measures important to each healthcare stakeholder. Wherever feasible, the CVET has been designed for interactive modification of certain economic model parameters and visualization of how findings might change. Uncertainty in model results is also explored through “sensitivity analyses” – analyses of the robustness of the economic model to changes in certain probabilities and/or costs. Assignment of comparative value is made based on the performance of the technology in question across all of these measures, in consultation with the ICER Evidence Review Group. An example of the summary table from the CVET can be found below.

Details on the methodology underpinning the design and presentation of cost-effectiveness analyses within ICER appraisals are available on the ICER website at www.icer-review.org.

ICER Comparative Value Evidence Table (CVET)			
Measure	Technology A	Technology B	Difference (B-A)
1. Service Impact			
Tests	27.4	17.9	(9.5)
Visits	31.6	24.8	(6.8)
Hospitalizations	0.0	1.0	1.0
Hospital days	0.0	3.0	3.0
Days of missed work	4.7	5.9	1.2
Pathway Total	63.7	52.6	(11.1)
2. Cost-Consequences			
\$ to Prevent 1 Case of X		\$210,000	
\$ per Cure		\$350,000	
3. Cost per Life-Year Saved		N/A	(equivalent survival)
4. Cost per QALY Gained		\$1,050,000	
% of Cost/QALY <\$100,000		2.63%	
SA 1: Surg Compl. 50% of Basecase		\$547,000	
SA 2: ED 50% of Basecase		\$442,000	
5. Budget Impact (per 1,000, 2 years)		\$1,425,000	
6. Fixed Budget Tradeoffs		19.0	<i>Nurse FTEs @ \$75K each</i>
		11.4	<i>MD FTEs @ \$125K each</i>

Integrated Ratings

The ICER Integrated Evidence Rating™ combines the individual ratings given for comparative clinical effectiveness and comparative value. The overall purpose of the integrated ratings is to highlight the separate considerations that go into each element but to combine them for the purposes of conveying that clinical benefits provided by technologies come at varying relative values based on their cost and their impact on the outcomes of care and the health care system.

ICER Integrated Evidence Rating™: Comparator X vs. Reference Technology Y

Comparative Clinical Effectiveness	Superior: A	Aa	Ab	Ac
	Incremental: B	Ba	Bb	Bc
	Comparable: C	Ca	Cb	Cc
	Inferior: D	Da	Db	Dc
	Unproven/Potential: U/P	Ua	Ub	Uc
	Insufficient: I	I	I	I
		a High	b Reasonable/Comp Comparative Value	c Low