HTA and economics in the United States: a systematic review of ICER reports to evaluate trends, identify factors associated with recommendations, and understand implications

Denise T Kruzikas, PhD, MPH; Daniel C Malone, RPh, PhD, FAMCP; Sissi Pham, PharmD; Tyler K Reinsch, PharmD; and Ron Akehurst, DSc, Hon MFPH

What is already known about this subject

- Health technology assessments (HTAs) for the Institute for Clinical and Economic Review (ICER) are created by an independent organization and could potentially frame the costs and benefits of emerging technologies if methods follow established research standards.
- ICER methodologies have evolved over time with a substantial update to their value assessment framework in 2017 and a more recent update in 2020.

What this study adds

- This study can help raise further awareness of the quality and attributes of ICER reports, which influence results and corresponding value to decision makers.
- This assessment identifies areas for ICER to more closely align cost-effectiveness methodological approaches with widely accepted standards.

Author affiliations

Denise T Kruzikas, PhD, MPH; Sissi Pham, PharmD; and Tyler K Reinsch, PharmD, AESARA, Durham, NC. Daniel C Malone, RPh, PhD, FAMCP, University of Utah, Salt Lake City, and Ron Akehurst, DSc, Hon MFPH, Bresmed, Sheffield, UK.

AUTHOR CORRESPONDENCE:
Tyler Reinsch, 573.619.7028,
tyler@aesara.com

J Manag Care Spec Pharm.
2020;26(12):1548-57

Copyright©2020, Academy of Managed
Care Pharmacy. All rights reserved.

ABSTRACT

BACKGROUND: The Institute for Clinical and Economic Review (ICER) is a prominent health technology assessment (HTA) entity in the United States that considers costs and applies economic analyses to derive price-based recommendations. ICER continues to adjust its value framework, yet discussion persists regarding whether ICER methodologies align with established research standards. This work evaluates ICER assessments relative to those standards, providing a benchmark with the release of ICER's most recent value framework update.

OBJECTIVES: To evaluate ICER economic assessments for trends, factors related to

recommendations, and quality for use in U.S. decision making.

METHODS: We evaluated all ICER final evidence reports published between 2006 and August 31, 2019, with regard to base-case result trends over time, pricing sources, comparator selection, analytic perspectives, model uncertainty, how modeling results aligned with ICER's determinations of value for money, and comparison of ICER methodological approaches with established modeling standards. Analyses were stratified by time period, where appropriate, to account for changes in ICER's framework over time.

RESULTS: Of 58 ICER final evidence reports, 47 used the most commonly reported outcome (cost per quality-adjusted life-year

[QALY]); ICER-developed models evaluated 131 interventions and comparators with 238 base-case results. Pricing sources for ICER reports became more standardized in 2017, although sources were not associated with the likelihood of falling below ICER's cost-effectiveness thresholds. In 30% of base-case analyses (n = 72), ICER did not use a clinical comparator, although reasonable treatments were available. In modified societal perspectives scenarios applied in later assessments, 75% of analyses (n = 76) included productivity but did not specify how it was quantified. Reports did not explain how sensitivity and scenario analyses were selected or implications of results. ICER value for money determinations generally aligned with cost-effectiveness results, although

2 of 33 (6%) interventions ranked as low value and 3 of 5 (60%) interventions ranked as low-moderate value, met a \$150,000 per QALY threshold, and 14 of 37 (38%) moderate-value interventions exceeded this threshold; the most common rationale was related to national budget impact.

CONCLUSIONS: While some progress has been made, further improvement is needed to ensure that ICER assessments address the most relevant questions for target audiences, adhere to established research standards, and are reported in a manner that can be readily interpreted and applied to policymaking.

Although multiple entities consider health technology assessment (HTA) in the United States, only the Institute for Clinical and Economic Review (ICER) specifically considers costs and applies economic analyses to derive price-based recommendations. Established in 2006, ICER evaluates the clinical effectiveness, safety, and cost-effectiveness of select interventions and comparators using its own value framework to “translate evidence into policy decisions that lead to a more effective, efficient, and just health care system.”¹ ICER notes that its reports include full analyses of an intervention’s effectiveness, economic value, and other important elements to patients and families.¹

ICER has released numerous assessments evaluating pharmaceutical, diagnostic, device, and programmatic interventions, with the first final report issued in November 2007.² Its approach to value assessment has evolved; the first formal effort to establish an ICER value framework was undertaken in 2014–2015, and in 2017, ICER implemented a substantial update based on stakeholder input and public comment.³ This update introduced a conceptual framework that included greater specification of long-term value for money and cost-effectiveness assessments to inform decisions aimed at achieving sustainable access to high-value care for all patients.

Methodological changes applied to ICER’s approach to incremental cost-effectiveness analysis included specifying the health system as the base-case perspective and including a “modified societal perspective” (MSP) as a scenario analysis accounting for work productivity, both counter to the Second Panel on Cost-Effectiveness; extending cost-effectiveness thresholds to include \$50,000, \$100,000, and \$150,000 per quality-adjusted life-year (QALY); and establishing the cost per QALY as the primary measure for its cost-effectiveness analyses, as recommended by the Second Panel.^{4–7}

Despite the 2017 updates, the approaches used by ICER have consistently undergone scrutiny and remain in debate in recent publications.^{8–10}

The purpose of this study was to evaluate the economic analyses within ICER assessments for their quality, trends, influential factors, and usefulness for payer and access decision making.

Methods

We reviewed ICER final evidence reports following standards established by the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) as they could be applied to the specific purpose of evaluating ICER intervention assessments.¹¹ We included all ICER reports on health conditions and/or health interventions published between 2006 and August 31, 2019. Reports on methodological approaches or position statements were excluded. We reviewed all ICER final reports and, where available, report-at-a-glance and supplemental materials. When ICER reviewed a therapy area previously considered, these reviews were evaluated as individual assessments because they included new interventions and/or comparators, underwent separate ICER evaluations, and were reported in their own individual ICER final reports. We summarized the number of assessments across interventions, comparators, and/or specific populations for eligible reports.

Our analysis focused on assessments using cost per QALY, the most commonly reported economic outcome. Using this measure allowed us to evaluate trends, recommendations, and quality throughout the majority of ICER reports. Further, this outcome aligns with the recommendation of the First and Second Panels on Cost-Effectiveness in Health and Medicine for cost-effectiveness analyses to account for health effects in terms of QALYs.^{4,7}

For final reports released between 2007 and 2019, we examined cost per QALY results for patterns related to the time of assessment. To examine the updated ICER value framework, we did a subevaluation of reports released between 2017 and 2019, including the full year of 2017, which allowed us to have a baseline before the implementation of the value framework update in July 2017. In this subanalysis, we separated outliers greater than \$1,000,000 per QALY to evaluate results on a more consistent basis, following ICER’s 2017 update to its value framework, and to examine outlier patterns related to rare diseases, gene therapies, or other more costly considerations.

We were supported by a Bresmed analysis, which evaluated the number of interventions that underwent cost-effectiveness assessment by the National Institute for Health and Care Excellence (NICE) and met the NICE

thresholds of £30,000–£50,000 per QALY during 2017–2018 (unpublished data, Bresmed Health Solutions, analysis, 2019). This allowed us to observe potential trends in ICER cost per QALY results relative to an established HTA over a similar time period. While trends are not directly comparable, given the different regions and health care systems, this comparison offers insights into the relative influence of each body on intervention pricing and likelihood to meet cost-effectiveness thresholds.

To analyze factors specific to the ICER modeling framework, we removed assessments that used only previously published models. Focusing on those assessments based on models developed by ICER, we examined (a) rationale for the base-case price input; (b) type of comparators and rationale for the comparators selected; (c) analytic perspective; and (d) assessment of uncertainty. For all ICER reports, we explored how cost per QALY results of interventions related to those of comparators, when reported, and how cost-effectiveness results corresponded to ICER's selected final determination of value for money. We additionally considered ICER's methodological approaches relative to establishing modeling of good research practices.^{4,7,12}

The source or method for determining an intervention's base-case price was used to evaluate the rationales before and after the update to ICER's value framework that addressed this attribute. We also considered whether specific rationales appeared to be associated with cost-effectiveness ratios less than \$150,000 per QALY.

For the comparator analyses, we undertook a standardized, 2-step categorization approach. The first step identified whether ICER selected comparators based on treatments currently available to address the condition of interest, reflecting the perspective of evaluating the potential incremental value of the new interventions in the market. We classified comparators based on labeled indication as available in the product labeling and approved by the U.S. Food and Drug Administration, as well as practice guidelines and/or therapeutic compendia such as MicroMedex¹³ or UptoDate¹⁴ to 1 of 3 broad categories: (1) *active comparator*, selected comparator was approved and is used in routine clinical practice; (2) *no active comparator, despite approved clinical treatments used in routine clinical practice*; or (3) *no active comparator and no clinical interventions approved and/or used in routine clinical practice*.

In a second step, we identified whether ICER noted the rationale for comparator selection, as recommended by good research practices for modeling conceptualization

and transparent reporting. We classified ICER's reported rationales for selecting comparators into 1 of 6 categories:

1. *Clinically appropriate treatment alternative*: comparator was considered a viable treatment in real-world clinical practice for condition of interest.
2. *Clinical trial comparator*: comparator was based on use in clinical trials.
3. *Data constraints*: selection was driven by data availability.
4. *Usual care*: treatment consisted of usual, supportive, or other nonclinical care delivered at home.
5. *Unclear*: comparator was discussed, but a clear rationale did not emerge.
6. *Not specified*: a comparator was listed, but the rationale was not reported.

With the health care system as ICER's base-case perspective, we evaluated reports that included an MSP as a scenario analysis accounting for work productivity. We analyzed incorporation of an MSP with respect to trends over time and the inputs considered, as well as percent differences between health care system and MSP inclusive results and how differences may have affected ICER recommendations.

We examined common methods for testing model uncertainty, such as 1-way sensitivity analyses, probabilistic sensitivity analyses (PSA), and scenario analyses and the resulting effect on conclusions.

We evaluated ICER determinations of "value for money," which were categorized as low, low-moderate, moderate, moderate-high, high, insufficient evidence, or no vote taken. This categorization was based on the majority vote or ICER's overriding determination as reported in the final report. Where value for money was reported as "reasonable," "comparable," or "intermediate," these results were categorized as moderate. Of note, ICER does not have a formal construct for determining value for money across the multiple factors that will inform this decision. Review of final reports indicated that ICER's value for money conclusions are based on multiple factors of the ICER assessment, including cost-effectiveness, budget impact, panel voting, expert and patient input, contextual considerations, and/or other relevant items or discussions. Therefore, our analyses examined where value for money determinations aligned with cost-effectiveness results, and where cost-effectiveness results did not align with value for money conclusions (e.g., interventions with cost-effectiveness results below \$150,000 per QALY were considered to be low value), we identified the other factors that influenced the final ICER determination.

FIGURE 1 ICER Base-Case Results 2007-2019

ICER=Institute for Clinical and Economic Review; QALY=quality-adjusted life-year.

Results

Between November 23, 2007, and August 31, 2019, ICER released 58 assessments eligible for inclusion ([Supplementary Table 1](#), available in online article). Within the same assessment, multiple interventions can be evaluated with multiple comparators. Therefore, within 58 reports, 193 interventions with 284 base-case results were evaluated.

ICER has generally released an increasing number of reports each year; the highest number of final evidence reports released was 12 in 2018, and the most interventions and modeling analyses were conducted in 2017 ($n=48$ and $n=80$, respectively; Figure 1).

Cost per QALY was assessed in 47 (82%) final reports (137 interventions, 71%; 252 base-case modeling results, 89%). Of these, 42 (89%) final reports were based on models developed by ICER (131 interventions, 96%; 238 base-case modeling results, 94%).

Figure 1 presents the results for all primary results reported from January 2007 to August 2019. The number of analyses increased over time. In the subanalysis focusing on reports from 2017 to 2019 and separated outliers ($> \$1,000,000/\text{QALY}$, $n=11$), no discernible patterns in the cost per QALY results were observed over this time frame (Figure 2).

The Bresmed analyses found that the majority of interventions recommended by NICE ($n=65$, 72%) met the UK established thresholds of £30,000–£50,000 per QALY ([Supplementary Figure 1](#), available in online article). The majority of the interventions not recommended by NICE ($n=6$, 55%) exceeded £50,000 per QALY (unpublished data, Bresmed Health Solutions, analysis, 2019).

When examining the rationale used for the base-case price input, marked differences were observed over the 2 time periods (Figure 3). Before July 2017, the most common source of price pharmaceuticals was wholesale acquisition

FIGURE 2 ICER Base-Case Results <\$1,000,000 per QALY, 2017-2019^a

^aOutliers that exceeded \$1,000,000 per QALY included the following interventions: nusinersen for symptomatic patients with later-onset (type II/III) spinal muscular atrophy (\$8,156,000/QALY); C1 esterase inhibitor as long-term prophylaxis for hereditary angioedema (\$5,950,000/QALY); siponimod for treatment of overall secondary progressive multiple sclerosis (\$3,760,000/QALY); obeticholic acid for treatment of nonalcoholic steatohepatitis (\$2,748,300/QALY); niraparib for treatment of ovarian cancer (\$1,910,000/QALY); inotersen for the treatment of hereditary transthyretin amyloidosis (hATTR; \$1,730,000/QALY); siponimod for the treatment of active secondary progressive multiple sclerosis (\$1,565,024/QALY); evolocumab for the treatment of high cholesterol (\$1,300,000/QALY); nusinersen for the treatment of patients with infantile-onset (type I) spinal muscular atrophy (\$1,120,000/QALY); lanadelumab for the treatment of hereditary angioedema (\$1,110,000/QALY); and deutetrabenazine for the treatment of tardive dyskinesia (\$1,100,000/QALY).

ICER=Institute for Clinical and Economic Review; QALY=quality-adjusted life-year.

cost (WAC) minus an estimated discount (38%) followed by WAC alone (15%). After July 1, 2017, the most common approach to pricing was net (37%), followed by WAC minus an estimated discount (25%) and WAC alone (14%). However, no trends were observed in the source for base-case pricing inputs and modeling results relative to cost-effectiveness thresholds ([Supplementary Table 2](#), available in online article).

An active treatment was used as a comparator for 64% (n=153) of analyses; most were based on an established standard of care, the most common treatment, or a clinically valid alternative (69%, n=105). Treatment guidelines were specifically noted as a basis for determining comparators in

20% (n=21) of analyses. An additional 13% (n=20) were based on the intervention used in a clinical trial. Justification for comparator selection was not specified for 17% (n=26). Among those without an active comparator (n=85), 85% (n=72) evaluated conditions with reasonable treatment alternatives available. In these reports, ICER did not specify a reason in 57% (n=41) of the analyses and noted data constraints in 21% (n=15) and usual care for 17% (n=12); an ICER-specified decision not to use an active comparator accounted for the remaining 6% (n=4). For assessments in which an active comparator was neither used nor available, usual or best supportive care served as the comparator.

FIGURE 3 Intervention Base-Case Price Input Source Before and After July 2017

^aOther=AHFS Drug Monographs; assumption; canadadrugs.com; Medicare payment schedule; previous nonmodeling publication, rationale not specified; previously published economic model, rationale not specified; survey of Washington state surgeons; Washington State Health Care Authority.

^bOther=Comparator price; patient out-of-pocket price (not otherwise covered); analyst estimate.

AHFS=American Hospital Formulary Service; ASP=average selling price; AWP=average wholesale price; WAC=wholesale acquisition cost.

Analyses including an MSP scenario increased over time. Of the 238 base-case analyses, 102 (43%) included an MSP scenario; within the subanalysis (2017–2019), 91% included MSP scenarios. The majority ($n=76$, 75%) of MSP scenarios accounted for productivity; of these, 25 (33%) specifically noted patient productivity, and 8 (11%) specified that productivity was accounted for both patients and caregivers.

We further evaluated the difference in cost per QALY between MSP and health care system perspectives, expressed as a percent difference. The situations where the MSP findings had lower cost per QALY values than those results based on a health care system perspectives were from the following assessments: (a) CAR-T therapy analyses relative to 2 different comparators, both of which accounted for patients' and caregivers' productivity and time for treatment; (b) Duchenne muscular dystrophy, which considered caregiver costs in the dual-base case; and (c) hereditary angioedema, which did not fully describe elements considered in the MSP. For 96 of 102 analyses, incorporating an MSP did not change the cost-effectiveness conclusions relative to a threshold of \$150,000 per QALY.

For 6 analyses (3 for rheumatoid arthritis, 2 for episodic migraine, and 1 for voretigene in patients with biallelic RPE65-mediated inherited retinal disease and a mean age of 3 years), the MSP results were <\$150,000 per QALY, whereas results based on the health care system exceeded this threshold.

Among the 42 final reports based on an ICER-developed model with cost-effectiveness results reported as cost per QALY, ICER included 1-way sensitivity analyses in nearly all assessments ($n=41$, 98%). Beginning in June 2017, 1-way sensitivity analyses were conducted across all model parameters to identify key drivers; previously, ICER analyzed a subset of parameters but did not explain how these parameters were selected. Final reports did not typically provide interpretations of model stability or discuss the uncertainty of data inputs and/or implications of these analyses.

PSAs were considered in 25 of the 42 (60%) final assessments informed by ICER-developed models with cost per QALY results. At the intervention level, PSAs were conducted for 155 (65%) modeling base-case assessments and

FIGURE 4 ICER Results for Interventions in Same Populations with Same Comparators, <\$500,000 per QALY

Note: Results from multiple intervention-comparator base cases reported in a single assessment are presented along the same y-axis line; this representation allows interventions to be evaluated relative to other available treatments. Cost-effectiveness results for daclizumab are noted in orange relative to other interventions considered within the same ICER report, as an illustrative example of how a new intervention may compare with other interventions already on the market. In assessment 16, other interventions included fingolimod, dimethyl fumarate, interferon B-1b 250 mcg (Extavia, Betaseron), peginterferon B-1a, and natalizumab—the comparator was generic glatiramer acetate 20 mg (Glatopa), as the most commonly used DMT and lowest priced version. In assessment 17, other interventions included teriflunomide 7 mg, teriflunomide 14 mg, interferon B-1a 22 mcg, interferon B-1a 30 mcg, interferon B-1a 44 mcg, glatiramer acetate 20 mg (Copaxone, Glatopa), fingolimod, dimethyl fumarate, interferon B-1b 250 mcg (Extavia, Betaseron), peginterferon B-1a, natalizumab, alemtuzumab—the comparator was best supportive care. Of the 22 assessments, the majority (n=20) exceeded cost-effectiveness thresholds of \$50,000–\$150,000 per QALY with 2 exceptions (interferon B-1b 250 Extavia compared with generic glatiramer acetate 20 mg [\$144,873/QALY], alemtuzumab compared with best supportive care [\$38,277]). Daclizumab cost-effectiveness results similarly exceeded thresholds but were within range of, and had lower cost per QALY results than, other available interventions. Daclizumab was deemed by ICER as “low value.”

ICER=Institute for Clinical and Economic Review; QALY=quality-adjusted life-year.

were more common in reports released from 2017 to 2019 (n=125, 81%). The mean probability that PSAs would meet the thresholds considered is shown in [Supplementary Table 3](#) (available in online article). The most common threshold considered in PSAs was \$150,000; the mean likelihood of meeting this benchmark was 31% (SD=41). Only 4 PSAs reported the likelihood to be cost saving. Higher thresholds were used for some ultra-rare conditions, the most common of which was \$200,000 per QALY with a mean probability of meeting this threshold of 44% (SD=43).

At least 1 scenario analysis accompanied 152 base-case analyses (64%), with a total of 811 scenario analyses across

all final reports. Of the base-case analyses that did not report any scenario analyses, 79 (92%) were from final evidence assessments released before July 2017. Reports did not formally or consistently provide information on the rationale for scenario analyses conducted.

Base-case results were reported for multiple interventions based on the same population and comparator in 20 final evidence reports. These assessments considered 94 interventions and comparators and generated 59 base-case results. Figure 4 presents base-case results <\$500,000 per QALY (n=146 from 16 final evidence reports; multiple comparators in the same report produced 28 intervention-comparator assessments). The highlighted

TABLE 1 Base-Case Cost-Effectiveness Results and ICER Value-for-Money Determinations

	Low	Low-Moderate	Moderate	Moderate-High	High	Insufficient Evidence	No Vote
Votes taken							
n	39	5	37	2	11	10	151
Base-case result, \$							
Min	31,112	-44,084	7,800	79,200	3,929	36,759	1,803
Max	8,156,000	328,000	643,813	89,610	124,541	96,846	2,748,300
Base-case result <\$150,000/QALY or dominant							
n	2	3	23	2	11	4	78
%	6	60	62	100	100	40	52

ICER=Institute for Clinical and Economic Review; QALY=quality-adjusted life-year.

example, from the 2017 multiple sclerosis final evidence report, illustrates how the base-case results for the new-to-market intervention of interest, daclizumab, compared with the other available interventions. This report evaluated interventions against 2 comparators: generic glatiramer acetate (assessment 16 in Figure 4) and best supportive care (assessment 17). ICER deemed daclizumab “low value for money” in both assessments.

Finally, we explored the associations between intervention base-case results and ICER’s determinations of value for money (Table 1). Where interventions were deemed to be of moderate-high or high value, all base-case results fell below cost-effectiveness thresholds. In the low, low-moderate, and moderate voting categories, some base-case results did not align with the final ICER value for money determination. In 2 (6%) analyses in which the interventions were deemed to be of low value, the base-case results met a \$150,000 per QALY threshold; among interventions voted as low-moderate value, 3 (60%) met this threshold. Among interventions deemed to be of moderate value, 14 (38%) exceeded the \$150,000 per QALY threshold.

Explanations of the rationales for the value for money determinations varied widely. With 1 exception, deciding factors were unrelated to other benefits and/or contextual considerations. The most common rationale pertained to potential health care costs and the ICER-calculated national budget impact threshold, although this related to just 4 (21%) analyses. For 10 analyses in which modeling was conducted and reported, ICER found insufficient evidence to make a value for money determination.

Discussion

The number of assessed ICER final reports and interventions has increased over time, with a noticeable upward trend in the number of reports from 2017 to 2019. Since 2017, more results were above \$1,000,000 per QALY, although these represent a minority (8%) of all base-case results reported over these 3 years. Among reports issued from 2017 to 2019, after removing analyses with cost per QALY above \$1,000,000 per QALY, we observed a random dispersion across the results from this subset of reports, which suggest 2 items for consideration.

First, it suggests that topics typically selected for ICER review may not be focused solely on the most expensive or least cost-effective interventions. Second, these results may indicate that ICER does not have a substantial effect on pricing. This may be because new interventions are either entering markets with established treatments and pricing, or they may be first-in-kind treatments without available alternatives. This is in contrast to the experience in the United Kingdom, where intervention cost-effectiveness tends to correspond with NICE thresholds of £30,000-£50,000 per QALY. One might expect NICE to have greater influence, given the single-payer system as observed. It is also possible that in the United States effect may occur behind closed doors as payers and manufacturers negotiate net pricing.

ICER’s reported source for base-case cost inputs has been more consistent since the update of its value framework in July 2017; net and WAC pricing both appear to be the most commonly used base-case inputs.

Substantial variation was observed with respect to comparator selection and the rationales provided. While ICER’s stated objective is to help improve the effectiveness and efficiency of our current health care system, nearly one third of analyses did not include an active clinical comparator, although an alternative was available in clinical practice. The reasons for not including such comparators were not consistently provided, and reports released as recently as 2018 did not provide a justification for comparator selection. When rationales were provided, data constraints commonly drove the comparator selection. Established good research standards specify that models should be based on the decision

problem or research question of interest and not other factors, such as data availability.¹²

Evaluations of specific reports reveal further insights into how comparator selection may be problematic. For example, the cholesterol assessments conducted between 2017 and 2019 noted the use of a placebo comparator based on clinical trials, while the actual trial populations used statins. Using a placebo comparator is inappropriate for clinical conditions where there is widespread acceptance of treatment vis-à-vis treatment guidelines. In a more recent example of spinal muscular atrophy, zolgensma and nusinersen were compared with best supportive care, and the treatments were not evaluated directly. ICER's rationale was that "evidence for zolgensma in this setting is based on 12 patients, while the evidence for Spinraza comes from a randomized trial with over 100 patients. As in previous reports, we feel it is inappropriate for a therapy to appear cost-effective simply by offsetting costs of a recently introduced very expensive alternative."¹⁵ These examples underscore the need for a more standardized approach to comparative selection.

While ICER has increasingly included an MSP in its more recent assessments, there often are few details on how they are quantified, and approaches are inconsistent. As an example, the 2019 Duchenne muscular dystrophy dual base-case MSP included relatively comprehensive indirect costs, although caregiver utilities were excluded (\$390,000/QALY); utilities were considered only in scenario analyses (\$202,000/QALY and \$136,000/QALY for 1 and 2 caregivers, respectively). A rationale for an MSP base case that does not fully account for societal value was not provided, and such lack of transparency could lead to misinterpretation. While the MSP is not only inconsistent across reports, it also does not adhere to recommended modeling standards. The Second Panel recommends not only that a societal perspective should be included, but that such a reference case fully include direct and indirect medical costs.⁴

One-way sensitivity analyses were often conducted on all parameters, and while key drivers were appropriately noted, the assessments did not fully discuss the implications of the results. PSAs were performed more frequently after 2017. Such analyses would be more useful if the reporting reflected the range of results and therefore better characterized the degree of variability and uncertainty inherent in the model.

Scenario analyses became more common in recent assessments, were clearly identified, and the results were generally reported for each scenario. However, the ICER reports seldom describe the rationale for selecting scenarios. Moreover, the majority of results were not interpreted within a context or with a discussion of the implications.

To better inform decision making, greater detail and clarity should be provided so that readers can comprehend the selection for parameters and scenarios, as well as the degree of uncertainty associated with parameter estimates and modeling results.

ICER's value determinations generally align with the cost-effectiveness modeling results. However, when there are discrepancies, the results of ICER's national budget impact calculations have provided the best explanation for the ICER value for money determinations given. The ICER budget impact calculations follow UK NICE practice, which is based on a single-payer system. In the fragmented U.S. health care system, no single payer bears the national budget of health care interventions, and budget impacts can fluctuate widely across payers. Further, as highlighted in the example of daclizumab, value determinations based on cost-effectiveness thresholds may not capture the context of the U.S. market, in which competition can lead to new entrants where previous treatments have set pricing precedent.

Delivering an overarching recommendation for individual payers, based on an approach that may not be relevant to some or any individual U.S. payers, has limitations. Payers likely have varying willingness to pay thresholds, different populations, coverage criteria, and budgets.

LIMITATIONS

This review has its own limitations. First, an evaluation of comparators was particularly challenging and required a degree of subjectivity to create categories for quantified analyses. The authors discussed the various treatment alternatives at length, but other categories, definitions, and assignments may have been possible.

Second, ICER final evidence reports contain large amounts of text and data; given ICER's short timelines and number of individuals involved, ICER reporting is not always consistent. Some information conflicted across different sections of the same final evidence report or between the report and the corresponding report-at-a-glance. Where information conflicted, we based inputs for our analyses on the evidence that in our judgement was aligned with the totality of information provided in the final report.

This evidence-based analysis illustrates the progress made by ICER in its economic assessments and the areas where further advancement is required. Remaining improvements include (a) ensuring that modeling is determined by the most relevant research question rather than data availability; (b) better provision of rationales and explanations for choices made during the assessment, including comparators, sensitivity analyses, and scenario analyses; and (c) reporting results with the full context and totality of evidence. This

particular improvement is pivotal, since cost-effectiveness is not well applied when recommendations are based on singular point estimates. Assessments will benefit from standardized reporting, and similar information should be provided in the same sections across all assessments. This would enhance the quality of the reporting and likely the quality of the analyses.

Conclusions

Following the July 2017 update, ICER's approach to intervention base-case pricing inputs is more consistent across assessments. In addition, an MSP is more often included in scenario analyses for common conditions and as a dual base case for ultra-rare diseases since 2017.

We recommend improvements related to interpretability and application of ICER's assessments for decision making. While the comparators are clearly noted, adequate justification of comparators may be lacking; such justification is particularly important when possible clinical alternatives exist but are not chosen.

Also, ICER reports lack explanation for why specific sensitivity and scenario analyses were conducted, descriptions of model drivers that may affect the stability and confidence of the results, and overall interpretations of the degree of uncertainty that may exist. We observed significant variation across reports in the level of detail. ICER recommendations are currently based on the base case; they do not consider the totality of all the evidence generated through the analyses. It is the understanding of the magnitude and direction of uncertainty, relative to a base case, that drives utility of economic models.

While some advances have been made in ICER's economic assessments, more needs to be done to ensure that (a) analyses address relevant research

questions, (b) methodologies adhere to established research standards, and (c) results are generated and reported in a manner that can be readily understood and appropriately applied to policymaking related to patient access to interventions. Otherwise, policies based on ICER assessments as currently conducted may not be fully informed and could have unintended consequences.

DISCLOSURES

No outside funding supported this study. The authors have nothing to disclose.

REFERENCES

1. Institute for Clinical and Economic Review. About. Accessed November 15, 2020. <https://icer-review.org/about/>
2. Institute for Clinical and Economic Review. Intensity modulated radiation therapy (IMRT) for localized prostate cancer. Final appraisal document. November 23, 2007. Accessed November 15, 2020. https://icer-review.org/wp-content/uploads/2016/02/IMRT_Final.pdf
3. Institute for Clinical and Economic Review. Overview of the ICER value assessment framework and update for 2017-2019. Accessed November 15, 2020. <http://icer-review.org/wp-content/uploads/2017/06/ICER-value-assessment-framework-Updated-050818.pdf>
4. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for the conduct, methodological practices and reporting of cost-effectiveness analyses: Second Panel on Cost-Effectiveness in Health and Medicine. *JAMA*. 2016;316(10):1093-103.
5. Anderson JL, Heidenreich PA, Barnett PG, et al. ACC/AHA statement on cost/value methodology in clinical practice guidelines and performance measures: a report of the American College of Cardiology/American Heart Association Task Force on Performance Measures and Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2014;63(21):2304-22.
6. Neumann PJ, Cohen JT, Weinstein MC. Updating cost-effectiveness – the curious resilience of the \$50,000-per-QALY threshold. *N Engl J Med*. 2014;371(9):796-97.
7. Weinstein MC, Siegel JE, Gold MR, et al. Recommendations of the Panel on Cost-Effectiveness in Health and Medicine. *JAMA*. 1996;276(15):1253-58.
8. Neumann PJ, Cohen JT. America's "NICE"? Health Affairs blog, March 12, 2018. Accessed November 15, 2020. <https://www.healthaffairs.org/doi/10.1377/hblog20180306.837580/full/>
9. Dubois RW, Westrich K. As value assessment frameworks evolve, are they finally ready for prime time? *Value Health*. 2019;22(9):977-80.
10. Neumann PJ. Yes, improve the U.S. value frameworks, but recognize they are already in prime time. *Value Health*. 2019;22(9):975-76.
11. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151(4):264-69.
12. Caro JJ, Briggs AH, Siebert U, Kuntz KM. Modeling good research practices – overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-1. *Value Health*. 2012;15(6):796-803.
13. IBM Watson Health. IBM Micromedex solutions. Database. Retrieved March 1, 2019-August 1, 2020. Accessed November 15, 2020. <https://www.ibm.com/watson-health/about/micromedex>
14. Wolters Kluwers. UpToDate. Product. Retrieved March 1, 2019-August 1, 2020. Accessed November 15, 2020. <https://www.uptodate.com/home/product?&redirect=true>
15. Institute for Clinical and Economic Review. Spinraza and zolgensma for spinal muscular atrophy: effectiveness and value. February 22, 2019. Accessed November 15, 2020. <https://icer-review.org/material/sma-evidence-report/>