Using QALYs versus DALYs to measure cost-effectiveness: How much does it matter?

Xue Feng, David D. Kim, Joshua T. Cohen, Peter J. Neumann and Daniel A. Ollendorf

Center for the Evaluation of Value and Risk in Health, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston, MA, USA

Objective. Quality-adjusted life-years (QALYs) and disability-adjusted life-years (DALYs) are commonly used in cost-effectiveness analysis (CEA) to measure health benefits. We sought to quantify and explain differences between QALY- and DALY-based cost-effectiveness ratios, and explore whether using one versus the other would materially affect conclusions about an intervention’s cost-effectiveness.

Methods. We identified CEA s using both QALYs and DALYs from the Tufts Medical Center CEA Registry and Global Health CEA Registry, with a supplemental search to ensure comprehensive literature coverage. We calculated absolute and relative differences between the QALY- and DALY-based ratios, and compared ratios to common benchmarks (e.g., 1× gross domestic product per capita). We converted reported costs into US dollars.

Results. Among eleven published CEA s reporting both QALYs and DALYs, seven focused on pharmaceuticals and infectious disease, and five were conducted in high-income countries. Four studies concluded that the intervention was “dominant” (cost-saving). Among the QALY- and DALY-based ratios reported from the remaining seven studies, absolute differences ranged from approximately $2 to $15,000 per unit of benefit, and relative differences from 6–120 percent, but most differences were modest in comparison with the ratio value itself. The values assigned to utility and disability weights explained most observed differences. In comparison with cost-effectiveness thresholds, conclusions were consistent regardless of the ratio type in ten of eleven cases.

Conclusions. Our results suggest that although QALY- and DALY-based ratios for the same intervention can differ, differences tend to be modest and do not materially affect comparisons to common cost-effectiveness thresholds.

Cost-per-quality-adjusted life-year gained (QALY) and cost-per-disability-adjusted life-year averted (DALY) studies have become commonly used measures in the current practice of cost-effectiveness analysis (CEA) (1;2). In recent decades, childhood and adult mortality rates have continued to decline worldwide (3), whereas morbidity, or time lived with health loss, has become a more serious concern in high-income countries (HICs) and low- and middle-income countries (LMICs) alike. Consequently, QALYs and DALYs are being used increasingly frequently to assess the cost-effectiveness of interventions that affect quality as well as length of life. Because of this, a practical question for researchers and decision makers is which measure should be applied to a given intervention in a specific setting, whether defined by disease, geography, a country’s per capita income, or another characteristic. The QALY-based measure has been recommended by many health technology assessment agencies in HICs, whereas the DALY-based measure is generally preferred in LMICs (4;5). Indeed, the number of published cost-per-DALY studies has substantially increased in LMICs over the past decade (2). One possible reason is that the freely and publicly available disability weights, which required for DALY calculations, can significantly reduce the cost of conducting CEA in resource-constrained settings. In contrast, cost-per-QALY studies, which use utility estimates for specific states and capabilities, may require more time and resources to collect or assemble (6).

QALYs and DALYs: Similarities and Differences

Conceptual Approach

Although the intent in using QALYs or DALYs in CEA s is similar, the theoretical and technical underpinnings of the two metrics differ (1). The concept of the QALY was developed in the 1960s; it represents the products of years lived and the associated utility values, ranging from 0 (dead) to 1 (perfect health). Utility estimates represent the perspective of an individual’s values or preferences, based on the central tenet of “welfarist” economics—that individuals are the best judges of their own welfare, and improved societal welfare as the ultimate goal...
is based on the sum of these individual utilities. However, QALYs also integrate so-called “extra-welfarist” elements to utility assessment, such as the contribution of particular states of health, functioning, and patient preferences to utility estimation (7;8). The primary application of QALYs has been the same ever since their initial use—to compare the benefits and risks of medical interventions (9).

In contrast, the DALY was developed in the 1990s by the Global Burden of Diseases, Injuries, and Risk Factors (GBD) initiative to assess burden of disease at a population level, to understand leading causes of health loss worldwide, and to compare population health across geographic settings (10). DALYs reflect the sum of years of life lost due to premature mortality and years lived with disability. The disability weights used for DALYs are inverse to that of utility weights, with “0” referring to no disability and “1” representing the dead state. DALYs also do not explicitly integrate extra-welfarist concepts; for example, disability weights are defined not based on surveys of individuals but based on expert opinion, as in the view of its developers a single set of weights anchored to specific diseases better facilitated cross-cultural comparisons than did some form of self-assessment (9). In addition, non-health effects are limited to age and sex alone. In recent years, GBD has refined its disability weights to attempt to isolate health loss from welfare loss and social context (11); these weights are intended to be universal and invariant to setting or population but are still undergoing further testing.

**Empirical Comparisons**

A number of other studies have discussed the theoretical differences between QALYs and DALYs (9;12–16), and have generally concluded that both measures have proven serviceable for resource allocation and priority-setting, but do differ in terms of estimation. For example, Sassi et al found that numeric differences between utility and disability weights may lead to further divergence between the QALYs and DALYs (12). Age weighting was also considered a major difference between the two measures (9), although the GBD no longer recommends such weighting. Airoldi et al. found that QALY gained is consistently larger than DALY averted because of the reference age used; differences in cost-per-QALY and cost-per-DALY studies from 1996 to 2017. The search strategies, data collection process, and review methods are similar for the two registries and have been described previously (1;17;18). We used the title and PubMed ID of the article to identify whether there were studies contained by both registries. If so, the identified study was deemed eligible based on our inclusion criteria, as both QALY- and DALY-based ratios were reported in the same article for the same intervention(s).

We also performed a supplemental search to identify articles published since 2018 using databases of PubMed, EMBASE, and Econlit to identify articles reporting results by both measures. We followed the same steps as mentioned above and used keywords of “QALYs,” “quality-adjusted,” “DALYs,” and “disability-adjusted,” to identify candidate papers.

**Methods**

**Inclusion Criteria**

Studies included English-language CEA articles that reported results using both cost-per-QALY and cost-per-DALY measures, published from 1996 through 2018.

**Data Sources**

We utilized two databases maintained by the Center for the Evaluation of Value and Risk in Health at Tufts Medical Center in Boston, Massachusetts: the CEA Registry (http://www.cearegistry.org), with information on 7,287 cost-per-QALY studies collected from 1976 to 2017, and the Global Health (GH) CEA Registry (http://www.ghcearegistry.org), which summarizes 620 cost-per-DALY studies from 1996 to 2017. The search strategies, data collection process, and review methods are similar for the two registries and have been described previously (1;17;18). We used the title and PubMed ID of the article to identify whether there were studies contained by both registries. If so, the identified study was deemed eligible based on our inclusion criteria, as both QALY- and DALY-based ratios were reported in the same article for the same intervention(s).

**Variables and Analysis of Data**

We extracted information from the selected articles, including year of publication, intervention type, study region, disease area, study funder, study perspective, cost discount rate, DALY and QALY discount rate, age-weighting use, sources of disability weights, sources of utilities, cost-per-QALY gained and cost-per-DALY averted results in the base case, the use of a cost-effectiveness “threshold” for decision making as mentioned by the authors, and the conclusions of the study.

We quantified the differences between ratios by QALY and DALY measures based on their absolute and relative difference. Relative difference was defined as the absolute difference divided by the QALY-based ratio. Magnitudes of both types of differences were compared. We also counted the number of cases for which the cost-per-DALY was higher than the cost-per-QALY for each intervention studied. All costs estimated in non-U.S. currency were converted to United States (U.S.) dollars based on the present value year used in each article as we intended to evaluate differences within rather than between studies. In addition, we compared the QALY- and DALY-based ratios to commonly used cost-effectiveness thresholds, including those reported by the articles such as one time gross domestic product (GDP) per capita, as well as any country-specific thresholds mentioned in the articles. Because our sample size was expected to be small, primary analyses were descriptive in nature.

Furthermore, we estimated the net monetary benefit (NMB) based on QALY and DALY measures, respectively (NMB = QALY [DALY] × threshold – ACost) so the results from both measures could be expressed in the same unit of U.S. dollars for further comparison. When calculating the NMB, we applied the threshold reported in each article, whether based on a
commonly used benchmark (e.g., 1× GDP per capita) or a country-specific estimate. Costs were presented in 2018 U.S. dollars for this analysis. The Pearson correlation coefficient was used to examine the relationship between the relative differences as assessed by NMB and the relative differences based on ratios.

**Results**

In total, we obtained eleven articles—ten articles from the two Tufts Medical Center registries and another 2018 article identified through the literature search (Figure 1) (19–29). Among the eleven articles in Table 1, seven (64 percent) focused on infectious diseases (i.e., HIV, TB, hepatitis B, hepatitis C, and rotavirus infections). Most of the articles (82 percent, 9/11) were published from 2015 to 2018. Five (45 percent) of the studies were from high income settings. Pharmaceutical interventions were assessed in seven studies (64 percent, 7/11); other types of interventions included immunization, care delivery, surgery, health education and behavior, legislation, and nutrition. Studies received funding from various sources such as government, foundations, academic institutions, healthcare organizations, the pharmaceutical industry, and other agencies. Studies were conducted using the perspectives of the healthcare sector (36 percent, 4/11) or healthcare payer (36 percent), or with a limited societal perspective (27 percent, 3/11). Most studies applied a discount rate of 3 percent for costs, QALYs, and DALYs. One study reported using age-weighting for the DALY measure. Cost-effectiveness threshold benchmarks of 1× or 3× GDP per capita were mentioned in all studies from LMICs, whereas country-specific thresholds (e.g., Australia: 50,000 AU$; The Netherlands: 20,000 euros) were used in the HICs. Disability weights from GBD sources were cited in eight articles (73 percent, 8/11), and utilities were obtained from a variety of sources, often not specific to the study setting. For example, utilities in a Zambia-based intervention cited a previous study in another African country (19); a Malawi study applied utilities from an Indian setting (29); and a Gambia study used utilities from multiple countries (26).

Most of the included studies (64 percent, 7/11) applied a Markov model; other modeling techniques included decision-tree, stochastic simulation, and metapopulation and compartment modeling. Only two studies stated they used primary data from specific clinical trials (26;28) to inform effectiveness calculations.

Four articles reported that the intervention of interest was “cost-saving” relative to the comparator (i.e., no cost-effectiveness
<table>
<thead>
<tr>
<th>Article</th>
<th>Setting</th>
<th>Disease area</th>
<th>Intervention type</th>
<th>Study funder</th>
<th>Study perspective by reader</th>
<th>Threshold mentioned</th>
<th>Discount rate (%)</th>
<th>Source for disability weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Ryan et al. (19)</td>
<td>Zambia</td>
<td>HIV infection</td>
<td>Pharmaceutical</td>
<td>Government</td>
<td>Healthcare payer</td>
<td>1× GDP per capita</td>
<td>3</td>
<td>GBD</td>
</tr>
<tr>
<td>2 Mangen et al. (20)</td>
<td>The Netherlands</td>
<td>Rotavirus infection</td>
<td>Immunization</td>
<td>Not determined</td>
<td>Limited societal</td>
<td>€20k</td>
<td>3</td>
<td>The Netherlands</td>
</tr>
<tr>
<td>3 Mihalopoulos et al. (21)</td>
<td>Australia</td>
<td>PTSD</td>
<td>Care delivery; health education and behavior; pharmaceutical; legislation</td>
<td>Government, non-profit organization</td>
<td>Healthcare sector</td>
<td>AU$50,000</td>
<td>3</td>
<td>Australia</td>
</tr>
<tr>
<td>4 Long et al. (22)</td>
<td>USA</td>
<td>Obesity</td>
<td>Nutrition; legislation</td>
<td>Government, foundation</td>
<td>Limited societal</td>
<td>NA</td>
<td>3</td>
<td>NA</td>
</tr>
<tr>
<td>5 Diel et al. (23)</td>
<td>UK</td>
<td>TB</td>
<td>Pharmaceutical</td>
<td>Not determined</td>
<td>Limited societal</td>
<td>€10k</td>
<td>3</td>
<td>GBD</td>
</tr>
<tr>
<td>6 Wolfson et al. (24)</td>
<td>Germany</td>
<td>TB</td>
<td>Pharmaceutical</td>
<td>Pharmaceutical industry</td>
<td>Healthcare payer</td>
<td>NA</td>
<td>3.5</td>
<td>GBD</td>
</tr>
<tr>
<td>7 Nguyen et al. (25)</td>
<td>Papua New Guinea, Australia</td>
<td>TB</td>
<td>Care delivery, Pharmaceutical, screening</td>
<td>Academic institutions</td>
<td>Healthcare payer</td>
<td>1× GDP per capita AU$35,000–120,000 AU$40,000–80,000</td>
<td>3</td>
<td>GBD</td>
</tr>
<tr>
<td>8 Nayagam et al. (26)</td>
<td>Gambia</td>
<td>Hepatitis B</td>
<td>Screening, pharmaceutical</td>
<td>Political and economic organization</td>
<td>Healthcare sector</td>
<td>1× GDP per capita $240 per DALY 3× GDP per capita</td>
<td>3</td>
<td>GBD</td>
</tr>
<tr>
<td>9 Aggarwal et al. (27)</td>
<td>India</td>
<td>Hepatitis C</td>
<td>Pharmaceutical</td>
<td>None</td>
<td>Healthcare payer</td>
<td>3× GDP per capita</td>
<td>3</td>
<td>GBD</td>
</tr>
<tr>
<td>10 Löfgren et al. (28)</td>
<td>Uganda</td>
<td>Groin hernia</td>
<td>Surgery</td>
<td>Government, foundation, academic institutions, healthcare organizations, pharmaceutical industry, other</td>
<td>Healthcare sector</td>
<td>1× GDP per capita</td>
<td>NA</td>
<td>GBD</td>
</tr>
<tr>
<td>11 Vetrini et al. (29)</td>
<td>Malawi</td>
<td>Diabetic retinopathy and macular edema</td>
<td>Screening, medical procedure</td>
<td>Foundation</td>
<td>Healthcare sector</td>
<td>1× GDP per capita, 3× GDP per capita</td>
<td>3</td>
<td>GBD</td>
</tr>
</tbody>
</table>

AU, Australian; GBD, global burden of diseases, injuries, and risk factors; GDP, gross domestic product; NA, not applicable; PTSD, post-traumatic stress disorder; TB, tuberculosis; UK, the United Kingdom; USA, the United States.
Among the seven remaining studies with the eleven intervention-specific QALY- and DALY-based ratios, cost-per-DALY results were higher than cost-per-QALY in six cases, whereas the reverse was seen in the other five instances. The magnitude of difference between the two measures also varied across studies. The relative differences between the two measures ranged from 6 to 122 percent, and absolute differences from approximately $2 to $15,000. However, the magnitude of difference was consistently modest, even in cases with seemingly large differences (Figure 2). For example, the study reporting an absolute difference of $15,000 between ratios for rotavirus vaccines in The Netherlands (20), had ratios that were both relatively high; as a result, the relative difference between ratios was only 19 percent. In contrast, the seemingly large relative difference of 122 percent was from a study of low-cost surgical mesh in a LMIC, with an absolute difference of only $9 between ratios (28). We were able to conduct our secondary analysis of NMB on seven interventions. In general, relative differences using these estimates were consistent with those directly employing the QALY- and DALY-based ratios (Pearson correlation coefficient = .67; Supplementary Table).

In Figure 2, we present the ratios of cost-per-QALY and cost-per-DALY compared with a set of threshold benchmarks for decision making for LMICs and HICs separately. Among eleven pairs of QALY- and DALY-based ratios, we identified only one instance of a change in favorability of results when compared to a cost-effectiveness threshold (29). For the remaining ten pairs with consistent conclusions, two pairs from the same study were not considered to be "cost-effective" interventions using country-specific threshold or 1× GDP per capita (20), for both QALY- or DALY-based ratios. Both ratios for another intervention were slightly above the threshold of 1× GDP per capita in the study country (26).

### Table 2. Differences between cost-per-QALY and cost-per-DALY (N = 11)

<table>
<thead>
<tr>
<th>Article</th>
<th>ICER ($/unit of benefit [DALY or QALY])</th>
<th>Absolute difference</th>
<th>Relative difference (%)</th>
<th>QALY-based ratio &gt; DALY-based ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Ryan et al. (19)</td>
<td>94</td>
<td>53</td>
<td>41</td>
<td>43.6</td>
</tr>
<tr>
<td>2 Mangen et al. (20)</td>
<td>Rotarix®</td>
<td>≈69,803</td>
<td>55,842</td>
<td>≈13,961</td>
</tr>
<tr>
<td></td>
<td>RotaTeq®</td>
<td>≈76,356</td>
<td>61,540</td>
<td>≈14,815</td>
</tr>
<tr>
<td>3 Mihalopoulos et al. (21)</td>
<td>Adults</td>
<td>16,967</td>
<td>14,288</td>
<td>2,679</td>
</tr>
<tr>
<td></td>
<td>Children</td>
<td>7,948</td>
<td>7,144</td>
<td>804</td>
</tr>
<tr>
<td>4 Long et al. (22)</td>
<td>Cost-saving</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 Diel et al. (23)</td>
<td>Cost-saving</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 Wolfson et al. (24)</td>
<td>Cost-saving</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7 Nguyen et al. (25)</td>
<td>Papua New Guinea—South Fly</td>
<td>3.5</td>
<td>5.3</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>Australia—Torres Straits Islands</td>
<td>6,077</td>
<td>8,462</td>
<td>2,385</td>
</tr>
<tr>
<td>8 Nayagam et al. (26)</td>
<td>511</td>
<td>540</td>
<td>29</td>
<td>5.7</td>
</tr>
<tr>
<td>9 Aggarwal et al. (27)</td>
<td>Cost-saving</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Löfgren et al. (28)</td>
<td>Low-cost mesh</td>
<td>7.6</td>
<td>16.8</td>
<td>9.2</td>
</tr>
<tr>
<td></td>
<td>Commercial mesh</td>
<td>33.3</td>
<td>58.2</td>
<td>24.9</td>
</tr>
<tr>
<td>11 Vetrini et al. (29)</td>
<td>400</td>
<td>766</td>
<td>366</td>
<td>91.5</td>
</tr>
</tbody>
</table>

DALYs, disability-adjusted life-years; QALYs, quality-adjusted life-years.

ratio was calculated) (Table 2). Among the seven remaining studies with the eleven intervention-specific QALY- and DALY-based ratios, cost-per-DALY results were higher than cost-per-QALY in six cases, whereas the reverse was seen in the other five instances. The magnitude of difference between the two measures also varied across studies. The relative differences between the two measures ranged from 6 to 122 percent, and absolute differences from approximately $2 to $15,000. However, the magnitude of difference was consistently modest, even in cases with seemingly large differences (Figure 2). For example, the study reporting an absolute difference of $15,000 between ratios for rotavirus vaccines in The Netherlands (20), had ratios that were both relatively high; as a result, the relative difference between ratios was only 19 percent. In contrast, the seemingly large relative difference of 122 percent was from a study of low-cost surgical mesh in a LMIC, with an absolute difference of only $9 between ratios (28). We were able to conduct our secondary analysis of NMB on seven interventions. In general, relative differences using these estimates were consistent with those directly employing the QALY- and DALY-based ratios (Pearson correlation coefficient = .67; Supplementary Table).

In Figure 2, we present the ratios of cost-per-QALY and cost-per-DALY compared with a set of threshold benchmarks for decision making for LMICs and HICs separately. Among eleven pairs of QALY- and DALY-based ratios, we identified only one instance of a change in favorability of results when compared to a cost-effectiveness threshold (29). For the remaining ten pairs with consistent conclusions, two pairs from the same study were not considered to be "cost-effective" interventions using the country-specific threshold or 1× GDP per capita (20), for both QALY- or DALY-based ratios. Both ratios for another intervention were slightly above the threshold of 1× GDP per capita in the study country (26).

https://doi.org/10.1017/S0266462320000124 Published online by Cambridge University Press
Discussion

Our study represents an attempt to quantify differences in estimates of cost-effectiveness based on QALY- and DALY measures when both were used in the same evaluation, and to explore possible reasons for these differences. Our findings suggest that differences were modest in relation to each ratio’s magnitude in most cases. Perhaps more importantly, in the vast majority of cases, these differences would not affect CEA conclusions or decisions based on commonly used thresholds for cost-effectiveness.
On the other hand, these modest differences may still have the potential to affect decisions to fund or not fund the health interventions; decision making can be influenced by many factors (e.g., the opportunity cost of other interventions) that vary within the specific contexts of each country. In addition, the motivation to use the two measures for the same intervention was not clearly stated in the included articles. For example, two studies mentioned that the two measures were the most commonly used metrics (21;25), and another posited that the use of the two measures may increase the robustness of the analyses (21). We cannot rule out the possibility of self-selection, however, potentially manifested here by focus on models, treatments, and conditions that would have ensured concordance of results between the two measures.

One of the major issues is that many of the included studies did not provide sufficient details on model specification to explain the factors associated with ratio-based differences. It is likely, however, that source of utilities and disability weights were a major driver. This may be a particular issue in LMICs, because respondents used as the basis of global estimates of disability weights were primarily from high-income settings in GBD studies (30–32), and because utility data often must be obtained from settings other than the location of interest for the study. In addition, comparatively small absolute differences in cost/QALY and cost/DALY ratios were often observed in studies targeted for LMICs, which reflects the relatively low costs of the interventions in these studies. For instance, the total cost of the intervention of low-cost mesh for groin hernia repair was only $49 (28). In such situations, differences in the method used to measure health gain may be less important given that the major driver of results is the low incremental cost itself. Still, this small absolute difference may impose substantial cost for the payers when considering budget planning for the covered population, particularly if the intervention will affect large numbers of individuals. Likewise, depending on population size, absolute differences in CEA estimates using DALYs versus QALYs may have an effect on price negotiations that could have quite considerable implications.

Our findings are consistent with those of previous studies, which concluded that the weight (disability weight vs. utility) used and age-weighting functions are major drivers of differences between QALY and DALY measures (9;13). However, age-weighting was used in only one of the included studies, suggesting that recent studies have adopted the 2010 guidance to remove these weighting functions. The cessation of use of age-weighting was in response to criticisms that the practice was potentially unethical and discriminatory (9;33), given that age-weighting assigns higher values to young- and middle-aged adults because of their higher potential for productivity. Therefore, differences in the utility and disability weights, as well as the sources of those estimates, are likely the major explanatory factors in our sample.

The conclusion about the interventions’ “acceptable” cost-effectiveness was only affected by the type of ratio used in one case. In that case, a study-reported threshold of $679 (per QALY gained or DALY averted) was used; however, if a more common threshold such as one time GDP per capita ($333 in this case) had been used, the intervention would have been found to be cost-ineffective regardless of the type of ratio employed. With regard to policy making, country- and context-specific thresholds have been suggested to decide whether an intervention is considered a priority in healthcare planning (34). These thresholds may be more informative in the process of decision making when one also considers the budget for healthcare spending, and decision makers’ willingness to divert funds from other healthcare interventions and/or consumption outside the healthcare sector. Whether or not decisions would differ for QALY- and DALY-based estimates using specific thresholds needs further exploration.

We acknowledge several limitations in our study. First, given likely differences in model structure, estimation, and programing language, among others, for the same intervention among different CEA, it is likely not feasible to adjust for these differences or directly compare the cost-per-QALY and cost-per-DALY results generated from different studies. Our exclusion of such studies limited the number of articles to those that used both measures in the same evaluation, which in turn limited our sample size and predicted the number of articles to those that used both measures in the same evaluation, which in turn limited our sample size and predicted the use of hypothesis testing. Second, we cannot rule out the possibility that the observed consistency between QALY- and DALY-based measures may be due in part to publication bias, manifested by a predisposition to publish studies with consistent results, whether favorable or unfavorable. Moreover, the calculation of the differences between the two ratios and the use of a single threshold for decision making recommendations is based on an assumption that the QALY and the DALY reflect comparable constructs. As described previously, these measures reflect somewhat different domains of health and may not be readily exchangeable (11;30;35). If this holds true, then different thresholds are likely required to inform decision making. On the other hand, the differences in construction and interpretation of DALYs and QALYs are not likely to affect the interpretation of our findings from the perspective of the current application of CEA to decision making; however, as decision making thresholds for cost-effectiveness have remained relatively constant over time. Although we acknowledge this limitation, we are unaware of any empirical research to quantify differences in QALY- and DALY-based ratios, and so the full implications of our assumptions are not known. We note that findings were similar when QALY- and DALY-based results were presented using common units in our NMB calculations.

Despite these limitations, this is the first study using published CEA to assess the potential relationship between QALYs and DALYs and to compare cost-effectiveness ratios with different thresholds. We find that, although nominal differences in results are observed, conclusions of the CEA are not likely to change based on the use of QALYs versus DALYs to measure health gain, when the commonly used thresholds for CE are applied. Our findings should be of interest to policy makers and researchers in LMICs, particularly those who may be limited to DALY-based analyses because of constraints of resource and data collection costs, as well as those who do have the ability to estimate QALYs, but are concerned about the challenge of doing so in a climate dominated by DALY-based research.

Conclusions
Our results suggest that although QALY- and DALY-based ratios for the same intervention can differ, differences tend to be modest and are unlikely to materially affect resource allocation recommendations. On the other hand, the modest differences may still affect decision making process when considered from a broader perspective, including opportunity cost of other healthcare interventions, budgets for healthcare spending, and price negotiation. Although both QALYs and DALYs can produce cost-effectiveness estimates that assist in healthcare decision making, further studies are warranted to better improve the methodologies and applications of these measures to address local health needs and concerns.
Supplementary Material. The supplementary material for this article can be found at https://doi.org/10.1017/S0266462320000124.

Financial Support. This study was funded by the Bill and Melinda Gates Foundation (grant no. OPP1171680).

Conflict of Interest. Dr. Feng reports no conflict of interest. Dr. Kim reports grants from the Bill and Melinda Gates Foundation, during the conduct of the study. Dr. Cohen reports grants from the Bill and Melinda Gates Foundation, during the conduct of the study; personal fees from Precision Health Economics, personal fees from Sage Therapeutics, personal fees from Merck Corporation, personal fees from Abbvie, personal fees from Indivior, personal fees from Sarepta Pharmaceuticals, outside the submitted work. Dr. Neumann reports advisory boards or consulting from Abbvie, Amgen, Avexis, Bayer, Congressional Budget Office, Vertex, Veritech, Janssen, Merck, Novartis, Novo Nordisk, Precision Health Economics, funding from The CEA Registry Sponsors by various pharmaceutical and medical device companies; grants from Amgen, Lundbeck, Bill and Melinda Gates Foundation, NPC, and Alzheimer’s Association, NIH. Dr. Ollendorf reports grants from the Bill and Melinda Gates Foundation, during the conduct of the study; personal fees and other from Sarepta Therapeutics, LLC, personal fees from DBV Technologies, Inc., personal fees from EMD Serono, other from Gerson Lehman Group, other from The CEA Registry Sponsors, personal fees from Autolus, Inc., outside the submitted work.

References