

Evaluating Alternatives to QALYs

KEVIN D. FRICK, PHD; AND JEROMIE M. BALLREICH, PHD, MHS

Health care resource allocation decisions require an objective that includes a metric. Quality-adjusted life-years (QALYs) are one health utility measure that facilitates complex resource allocation decisions. For example, one of us (K.D.F.) was involved in a study comparing surgical treatments for women with dysfunctional uterine bleeding who have symptoms including pain, anxiety, and fatigue.¹ A decision maker may allocate resources that alleviate pain but have only limited impact on fatigue or anxiety. Distilling health effects into a single metric, such as QALYs, facilitates the assessment of trade-offs between the 2 treatments for decision makers.

Researchers who focus on health care resource allocation are likely familiar with arguments regarding the strengths and weaknesses of QALYs.² For accountable care leaders and practitioners, the review by Browne et al in this issue of *The American Journal of Accountable Care*[®] presents a discussion of QALYs being based on assumptions about how society values individual health and the health of groups.³ Browne et al conclude, “Although it is true that science regularly compromises on method for practical reasons, it is important that the nature and scale of this compromise is transparent when the legitimacy of decisions based upon a controversial methodology is debated.” Although we agree on the need for compromise and transparency, we have concerns about whether the authors’ recommendations meet their own criteria.

First, consider the authors’ labels for sections of their paper: “The QALY fails to accurately measure the health of patients,” “The QALY discriminates against some groups,” and “The standard QALY framework does not represent how society actually values health care.” These are all legitimate

concerns to raise, but the solutions that Browne et al provide have their own challenges.

QALYs are intended to measure not health but rather health utility over time. Health utility is a function of health and represents the value that individuals obtain from health, including being healthy, having a longer life expectancy, and having greater capacity for education or work. The authors note that a commonly used health utility instrument does not include a direct measure of vision impairment and blindness. This is true, but it is not a fatal flaw for QALYs, as they measure the utility of life functions related to the impairment or condition rather than trying to measure impairment directly. The authors recommended measuring health impacts with measures more sensitive to changes or by using patient-derived or Delphi-developed assessments. These recommendations have 3 contraindications. Disease-specific measures would not help with allocation decisions across diseases. For measuring health utility, the high end of the scale is intended to be health with no conditions; with disease-specific measures, the high end of the scale is often interpreted as health without the condition in question regardless of other conditions. Lastly, Delphi panel members often have expertise that does not represent the population or their preferences.

Next, consider the authors’ section on QALYs being discriminatory. As they noted, this is not a new revelation. In real-world resource allocation decisions, policy makers rarely use QALYs without qualification; some recognize the implication of resource allocation on groups whose maximum health is less than ideal. The life-extending nature of interventions for patient populations with low maximum health utilities is given special consideration by increasing the weight of QALYs gained

when survival is extended. This is how the United Kingdom's National Institute for Clinical Excellence considers the QALYs from cancer medications.⁴ Other investigators have suggested higher willingness-to-pay thresholds for QALY gains from treatments targeting special populations.⁵ Investigators using QALYs to evaluate interventions that target special populations with low maximum QALYs should contextualize their results. Browne et al used a study of interventions targeting Duchenne muscular dystrophy (DMD)⁶ as an example of shortchanging patients; however, the authors of the DMD study were sensitive to the nature of DMD and reported results in life-years, mitigating the critique of Browne et al. The authors of the DMD study also involved patients and parents throughout the process, suggesting sensitivity to this population and exemplifying the best practices of patient involvement.

Browne et al also suggest that the standard QALY framework does not represent how society actually values health care. We can debate whether society believes in allocating resources based on a measure that allows health states worse than death, that has to be summarized over time with assumptions about how to value changes over time, that values health improvements for those who cannot return to full health appropriately, and that values each person's health improvement appropriately relative to others. Although not everyone may agree with all assumptions, there are pragmatic reasons to use QALYs, most notably being able to compare interventions and allocate where we derive the best "value."

The authors use the phrase "how society actually values health care" as if there is consensus. The US population has demonstrated great variation in values in the past year. Even if a majority agree with a value proposition focusing on the value of health over time and the value of improvements for a population that varies by maximum health achievable and age, leaders must also ascertain the most appropriate health utility summary measure—an average, median, or modal response.

There may not be a single metric that solves all the issues. For QALYs or any measure, we must be explicit about the assumptions that underly a measure's construction and discuss how the assumptions affect conclusions. Decision makers could evaluate the assumptions to choose the best instrument for every decision. The ultimate question for choosing a valuation method may be whether there are assumptions that leave us with less to reevaluate each time. Essentially, we compare the amount saved by not reevaluating each time with the cost of generating those savings.

Many of the arguments that Browne et al have made are made elsewhere in the literature.² We are reminded of Winston Churchill's quote about democracy being a form of government with many imperfections while all others have more

imperfections. A better summary measure would be useful, but it is unclear whether, in a society with free thought, there will ever be sufficiently little variation in valuation to be able to design methods in a way that does not require repeated and costly reevaluation of assumptions.

Author Affiliations: Johns Hopkins Carey Business School (KDF), Baltimore, MD; Johns Hopkins Bloomberg School of Public Health (JMB), Baltimore, MD.

Source of Funding: None.

Author Disclosures: The authors report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of this article.

Authorship Information: Concept and design (KDF, JMB); drafting of the manuscript (KDF, JMB); and critical revision of the manuscript for important intellectual content (KDF, JMB).

Send Correspondence to: Kevin D. Frick, PhD, Johns Hopkins Carey Business School, 100 International Dr, Baltimore, MD 21212. Email: kfrick@jhu.edu.

REFERENCES

1. Frick KD, Clark MA, Steinwachs DM, et al; STOP-DUB Research Group. Financial and quality-of-life burden of dysfunctional uterine bleeding among women agreeing to obtain surgical treatment. *Womens Health Issues*. 2009;19(1):70-78. doi:10.1016/j.whi.2008.07.002
2. Lipscomb J, Drummond M, Fryback D, Gold M, Revicki D. Retaining, and enhancing, the QALY. *Value Health*. 2009;12(suppl 1):S18-S26. doi:10.1111/j.1524-4733.2009.00518.x
3. Browne J, Cryer DR, Stevens W. Is the QALY fit for purpose? *Am J Accountable Care*. 2021;9(2):8-13.
4. PMG9 addendum — final amendments to the NICE technology appraisal methods guide to support the proposed new Cancer Drugs Fund arrangements. National Institute for Health and Care Excellence. 2018. Accessed May 14, 2021. <https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/process-and-methods-guide-addendum.pdf>
5. Modifications to the ICER value assessment framework for treatments for ultra-rare diseases. Institute for Clinical and Economic Review. November 2017. Accessed May 14, 2021. <https://icer.org/wp-content/uploads/2020/10/ICER-Adaptations-of-Value-Framework-for-Rare-Diseases.pdf>
6. Deflazacort, eteplirsen, and golodirsen for Duchenne muscular dystrophy: effectiveness and value. Institute for Clinical and Economic Review. August 15, 2019. Accessed May 14, 2021. https://icer.org/wp-content/uploads/2020/10/ICER_DMD-Final-Report_081519-2.pdf