CURRENT MEDICAL RESEARCH AND OPINION
2022, VOL. 38, NO. 2, 165–170
https://doi.org/10.1080/03007995.2021.2006535
Article ST-0327.R2/2006535

Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

Real-world impact of disease on functioning and activity: what is missed when using general instruments to estimate quality-adjusted life years

Tingjian Yan[a], Jesse D. Ortendahl[a], Eunice Chang[a], Zac Wessler[b], Amanda L. Harmon[a] and Michael S. Broder[a]

[a]Partnership for Health Analytic Research LLC, Beverly Hills, CA, USA; [b]Amgen, Thousand Oaks, CA, USA

ABSTRACT

Objective: Economic evaluations conducted to inform healthcare resource allocation often rely on quality-adjusted life years (QALYs) to measure therapeutic benefit. However, QALYs, with underlying health utilities estimated using the EQ-5D or SF-36, may fail to capture the impact of disease for all patients. How well-being and heath utility differ across several common conditions was explored.

Methods: This study examined eight diseases: arthritis, asthma, cancer, depression, diabetes, heart disease, lung disease and stroke. Health utilities for each disease were obtained from published literature. Other measures of disease burden, including physical functioning, cognitive functioning and physical activity, were estimated from the National Health and Nutrition Examination Survey (NHANES). Group rankings by these measures were compared to rankings by health utility.

Results: Health utilities were lowest for patients with depression (0.44), and highest for those with cancer (0.81). Physical functioning was most limited (higher score) among those with stroke (28.2) and had the least impact for cancer (24.4). Physical activity was most impacted by heart disease (27.3) and least impacted by depression (40.7). Cognitive functioning was lowest in stroke (41.6) and highest in asthma (52.0).

Conclusion: Differences in rankings of disease severity by metric indicate that the results of cost–utility analyses might be biased against treatments for certain diseases. As patient preferences for clinical outcomes vary, the full burden of disease should be considered in evaluations. Restricting access to treatments based on an incomplete estimate of burden could lead to misallocation of resources and a withholding of therapies that patients find valuable.

ARTICLE HISTORY
Received 11 May 2021
Revised 20 October 2021
Accepted 8 November 2021

KEYWORDS
Cost-effectiveness; value; resource allocation; health technology assessment

Introduction

When an avid runner falls ill, their primary concern may be to return to previous levels of physical activity. For an elderly individual with dementia, retaining cognitive functioning may be a priority. For someone accustomed to living independently, maintaining the ability to perform basic activities of daily living such as dressing and meal preparation may be of utmost importance. Each of these individuals might measure the impact of disease differently and might assign different values to the same treatment. However, typical value assessment assumes there is an average patient and determines the value of interventions that broadly maximize clinical benefits, ignoring patient heterogeneity.

The use of health technology assessments (HTAs) has been increasing over the past decade as a method for efficiently allocating scarce healthcare resources. HTAs of pharmaceutical innovations are typically conducted as cost-effectiveness analyses (CEAs) and estimate value by calculating the incremental cost-effectiveness ratio (ICER) of the novel therapy compared to existing alternatives for the average patient. These analyses commonly define the benefit of a therapy in life years or quality-adjusted life years (QALYs) gained[1]. QALYs are calculated by assigning a health utility, representing the severity of disease, to a given stage of disease, and calculating both the quantity and quality of life for patients with a condition. Defining benefit in terms of QALYs has the advantage of allowing for the comparison of different diseases when making decisions about the optimal use of health resources. Although they do make comparisons easier, metrics like QALYs and life years can fail to capture the full impact of disease on an individual or on society.

Specifically, the QALY metric has been criticized for a number of reasons, including inconsistency across heterogeneous patients with a specific condition, insensitivity to small but clinically meaningful changes, and equity in how the elderly and those with existing conditions are considered[2]. QALYs can underestimate the burden of disease, thereby miscalculating benefits and making true assessment of a treatment's value difficult. Another concern with using QALYs in value assessments is that the collections of health utilities used to inform their calculation is commonly done using broad, disease-agnostic instruments such as the EuroQol-5 Dimension (EQ-5D) and 36-Item Short Form Survey (SF-36). These capture broad measures of health (e.g.

CONTACT Jesse D. Ortendahl Jesse@pharllc.com 280 South Beverly Dr, Suite 404, Beverly Hills, 90212, CA, USA

ability to walk), but fail to account for benefits of more granular increases in physical functioning, cognitive functioning, ability to care for one's self and other impacts of disease[2–4].

Another method of assessing the impact of disease on functioning and activity is to use data from the National Health and Nutrition Examination Survey (NHANES), a nationally representative, cross-sectional study conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), conducted every two years to monitor the health and nutritional status of the civilian, non-institutionalized US population of all ages. While many questions in NHANES overlap with questions in the EQ-5D and SF-36, it also contains questions about functioning and activity not included in other instruments. For example, EQ-5D has a single question about mobility, while NHANES has three (i.e. any difficulty walking, difficulty walking 10 steps, difficulty walking a quarter mile). SF-36 has questions about the impact of disease on functioning such as moderate and intense activity, whereas NHANES better captures issues with fine motor coordination (e.g. ability to use a fork).

Since patient access to life-improving and life-extending therapies can be driven by the findings of economic evaluations, it is imperative that these evaluations capture the full nature of disease. The comparative impact of many common diseases on well-being has frequently been made using QALYs but less frequently with measures that incorporate additional elements of value. By comparing disease impact measured using a broader set of health measures with impact measured by QALYs, we hope to better understand whether QALYs fully capture the important contributors to quality of life (and hence to value).

Methods

Methods overview

We began by defining a general framework for measuring disease impairment and the value of interventions that can reduce the burden. While there are other components of value that are likely important to patients and society, the domains we assess in this study, noted in Figure 1, represent the elements that can be easily measured. To compare the burden of disease across domains, we assessed eight common diseases: arthritis, asthma, cancer, depression, diabetes, heart disease, lung disease and stroke. These diseases were chosen based on data availability and reflect some of the most common diseases in the US[5]. We estimated the severity of each of the eight diseases using multiple measures of health and functioning, and then compared the results to the severity of the same diseases as measured by health utilities and mortality.

Study participants and instruments

We identified respondents who had any of the eight common diseases as well as those with none (healthy controls) using data from NHANES. Specifically, we used 2015–2016 dataset to assess physical functioning and physical activity among participants ≥20 years, and the 2013–2014 dataset to assess cognitive functioning among participants ≥60 years as the latter was not conducted in the most recent survey and is limited to older individuals.

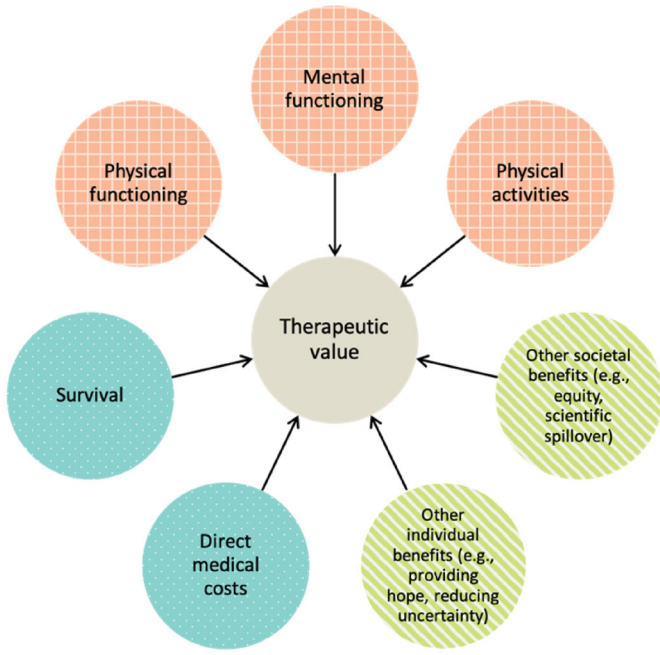

Figure 1. Conceptualization of therapeutic value. Blue bubbles indicate aspects of therapeutic value that are typically included in value assessments. Orange bubbles indicate those that could influence therapeutic value but may not be captured fully QALYs and are further assessed in this study. Green bubbles indicate areas that could influence value and are not addressed in this analysis.

Physical functioning

The NHANES physical functioning questionnaire collects self-reported data on functional limitations caused by long-term physical, mental and emotional problems or illness[6]. Participants ≥20 years old were asked to report difficulties in performing 20 tasks that assess ability in the following functional domains: (1) activities of daily living (ADLs), (2) instrumental activities of daily living (IADLs), (3) lower extremity mobility, (4) general mobility, and (5) social and leisure activities. Each task was scored on a 4 point Likert scale (1 = no difficulty to 4 = unable to perform). The sum of these 20 scores was taken to create an overall measure of physical functioning, ranging from 20 to 80.

Physical activity

Physical activity is defined as any bodily movement produced by skeletal muscles that results in energy expenditure[7]. Each NHANES participant ≥20 years old completed a physical activity questionnaire based on the Global Physical Activity Questionnaire (GPAQ) regarding the number of hours per week an individual was active[6]. This questionnaire assesses vigorous- and moderate-intensity physical activity in three domains: (1) at work, (2) traveling to and from places,

and (3) during recreational activities. Work-related physical activity refers to paid or unpaid work, studying or training, household chores and yard work. Recreational activity refers to sports and fitness activities. Vigorous-intensity activities are activities that require hard physical effort and cause large increases in breathing or heart rate, and moderate-intensity activities are activities that require moderate physical effort and cause small increases in breathing or heart rate. The suggested metabolic equivalent (MET) scores for vigorous work-related physical activity, moderate work-related physical activity, walking or bicycling for transportation, vigorous leisure-time physical activity and moderate leisure-time physical activity were 8.0, 4.0, 4.0, 8.0 and 4.0, respectively. For subjects who responded with no such activity, the hours per week for that activity were coded as zero. We multiplied the average number of hours per week spent in each activity by the suggested MET scores to get an estimate of MET-hours per week.

Cognitive functioning

The Digit Symbol Substitution Test (DSST) was used to measure cognitive performance of the US non-institutionalized population of adults aged 60 and over. The DSST, a subtest of the Wechsler Adult Intelligence Scale, Third Edition (WAIS–III)[8], is a cognitive test conducted using a paper form that has a key at the top containing nine numbers paired with symbols. Participants have 2 min to copy the corresponding symbols in the 133 boxes that adjoin the numbers. The score is the number of correct symbols drawn within the time limit. The maximum score is 133. DSST offers a practical and effective method to monitor cognitive functions over time in clinical practice. Performance on the DSST correlates with real-world functional outcomes and recovery from functional disability[9].

Health utilities and mortality

Utilities were based on a previous publication that estimated health utilities for patients using the EQ-5D. The impact of disease on health utility was defined as the difference between those with the condition and an age and sex matched cohort without the condition[10]. Mortality measures were based on the same publication and defined as the difference in rate per 100 person years between those with and without the condition, or excess mortality.

Data analysis

Descriptive statistics were generated for the eight disease groups and the healthy population. Means, standard deviations (SD) and medians were reported for continuous measures, and frequencies and percentages for categorical measures. To assess the incremental impact of disease, Cohen's *d* effect size was calculated to measure the differences in the three domains between the healthy population and each disease group. Cohen's *d* is a useful method for comparing the magnitude of findings across two or more different instruments, even when the scales of the instruments differ[11]. It was determined by calculating the mean difference divided by the pooled standard deviations for the two populations. A small effect size is generally considered to be at least 0.2, a medium effect size at least 0.5, and a large effect size 0.8 and above[12]. The larger the effect size, the more severe the disease burden on that specific domain of health. Sampling weights provided by NHANES were used to estimate nationally representative frequencies and means. Analyses of NHANES were conducted to rank disease severity on these three domains of health from largest to smallest. This ranking process was repeated within the same disease considering health utility and mortality.

Results

The number of participants included in the 2015–2016 analytic sample was weighted to be nationally representative. The weighted *N* ranged from 6,428,682 for stroke to 83,628,148 for the healthy population nationwide. The average age of the participants ranged from 41.3 years old for the healthy population to 66.2 years old for heart disease. In the 2013–2014 analytic sample, the number of participants ranged from 4,700,157 for stroke to 32,984,989 for arthritis. The average age of the participants ranged from 69.3 years for depression to 77.2 years for stroke and heart disease.

Real-world impact of disease on physical functioning, physical activity and cognitive functioning

Among the eight disease groups, physical functioning was most impacted (higher score) by stroke (28.2), lung disease (27.3) and depression (27.1), and was less impacted in patients with cancer (24.4) and diabetes (24.6). For those without disease, this value was 21.9 (Figure 2).

Heart disease and lung disease had the greatest impact on physical activity, with average MET-hours per week of 27.3 h and 29.0 h, respectively. Patients with depression were least impacted, with average MET-hours per week of 40.7 h. The corresponding value among those without disease was 50.2.

Cognitive functioning was most impacted in stroke patients (41.6) and had the lowest impact in asthma patients (52.0). Among those without disease, cognitive functioning was 54.5.

When assessing Cohen's *d* effect size for each disease and considering each domain, disease generally had the largest impact on physical functioning, followed by cognitive functioning (Figure 3). As an example, heart disease has an effect size of 0.99 on physical functioning, 0.65 on cognitive functioning and 0.35 on physical activity.

Health utilities and mortality

Health utilities were lowest for depression (0.44) and stroke (0.76), and highest for those with cancer (0.81) and diabetes (0.79) (Figure 2).

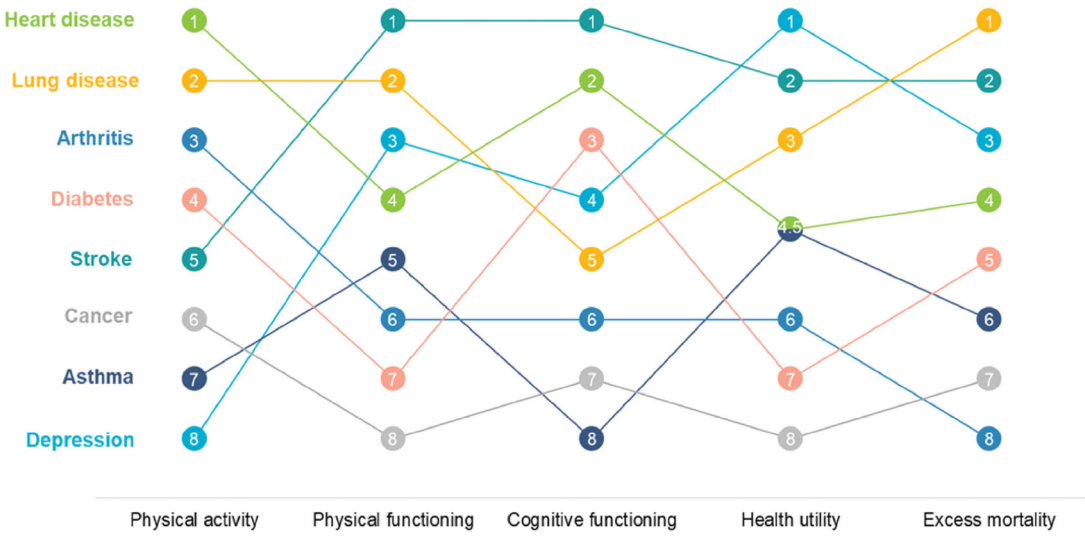

Figure 2. Rank ordering of diseases based on impact on health utility, excess mortality, functioning and activity. Diseases are ranked from highest impact (1) to lowest impact (8). Among the eight disease groups, physical functioning was most impacted (higher score) by stroke (28.2), lung disease (27.3) and depression (27.1), and was less impacted in patients with cancer (24.4) and diabetes (24.6). For those without disease, this value was 21.9. Health utilities were lowest for depression (0.44) and stroke (0.76), and highest for those with cancer (0.81) and diabetes (0.79).

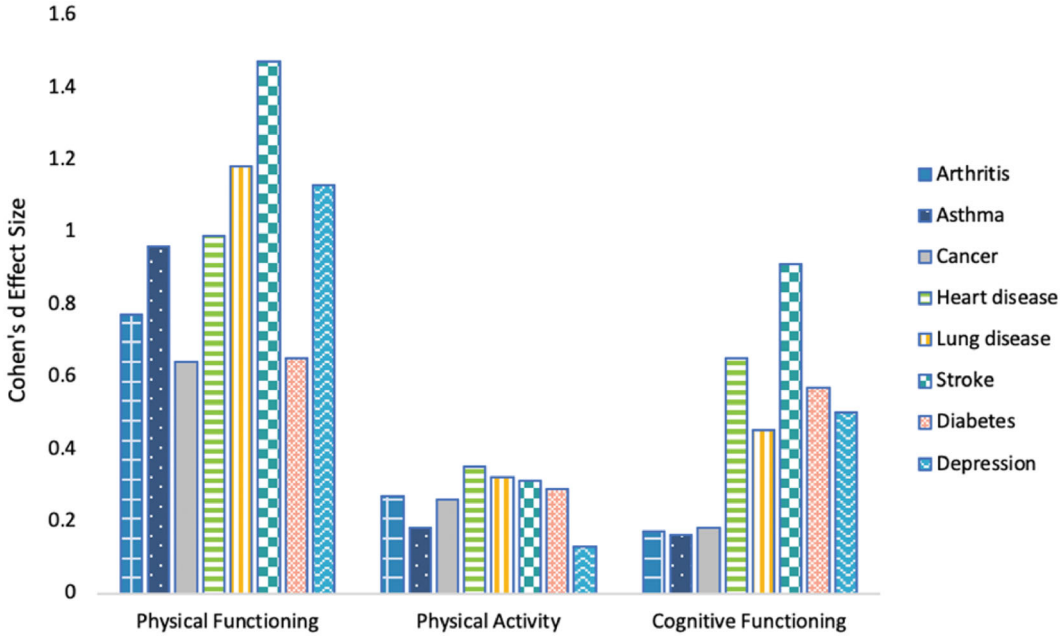

Figure 3. Impact of each disease on functioning and activity in comparison to a healthy population: Cohen's *d* effect size. When assessing Cohen's *d* effect size for each disease and considering each domain, disease generally had the largest impact on physical functioning, followed by cognitive functioning and physical activity.

The excess mortality associated with having one of the eight diseases was highest for lung disease (7.1 deaths per 100 person-years) and stroke (4.5 deaths per 100 person-years), and lowest for arthritis (0.2 deaths per 100 person-years).

Discussion

We found that the measured impact of eight common diseases varies substantially depending on how that impact is measured. The outcomes most commonly used in CEAs are QALYs based on health utilities, but rankings of disease burden with QALYs differ substantially from the rankings based on other, possibly more relevant measures. The choice of measurement instrument therefore directly affects resource allocation decisions, and prioritization of therapies by payers and policy makers could change depending on the metric used to measure value. For example, depression had the highest impact on health utility, but the lowest impact on physical activity. If there were two interventions for depression that had the same cost, but one had more impact on overall health utility while the other had a greater impact on activity level, the therapy that had the greater impact on utility would be prioritized under current approaches despite being sub-optimal to some patients. In the context of the US, where formalizing value assessment of new technologies is being considered, it is worth improving upon QALYs to establish a superior outcome measure that can adequately

capture heterogeneity in patient preferences. While it is the case that decision makers internationally do not solely rely on CEAs in value assessment, they are the primary quantitative tool that is considered and therefore society should understand how they are conducted and be confident that following such an approach would maximize societal values such as length and quality of life. Development of novel methods to estimate value to an individual could better serve patients and provide more benefit to the population as a whole. This study also supports the need for continued expansion of patient-centered decision making and allowing patients to choose the treatments and prioritize the outcomes that they find most important, as opposed to having a single set of recommendations that apply to all patients.

The criticisms that QALYs rely on health utilities have been published extensively, covering multiple issues, including measurement difficulties and concerns around equity[3,13–15]. However, one aspect that is particularly relevant to our findings is the assumption that the patient population is homogeneous, and all patients have the same preferences for clinical outcomes. While this has been discussed in the theoretical[16] and shown to be problematic within a single disease area[17], our analysis is the first to use nationally representative real world evidence and compare across diseases to see the impact of different diseases. In reality, some individuals might be willing to sacrifice physical functioning to maintain cognitive functioning, whereas others might be willing to forgo physical activity in order to maximize physical functioning. Within each of these hypothetical populations, the perceived burden of disease, and therefore the value of treatments that relieve this burden, would differ. While standard cost-effectiveness analyses struggle to incorporate this heterogeneity, multiple-criteria decision analysis (MCDA) allows for weighting of different outcomes in a manner that could be better suited for individual-level decision making[18]. By using MCDA, researchers could assess treatments for a group of patients with differing preferences, and better determine whether a given therapy might be beneficial to some even if it did not provide greater utility on average.

Additionally, there are concerns as to whether QALYs fully capture all the elements one might consider valuable when measuring the benefits of treatment. The EQ-5D and SF-36 are limited in their ability to assess some common changes in health status due to disease. If they do not capture the harms of having a condition, they cannot fully capture the benefits of treatment. There are benefits to the simplicity of the EQ-5D in reducing responder burden and allowing for rapid calculation of utility values; however, there are elements of each domain measured in NHANES that allow for more granularity. To better understand the true value of interventions, it is worth exploring instruments beyond the EQ-5D and SF-36 to measure quality of life. Other approaches to measuring the impact of disease, for example the Social Return on Investment method, may better capture the full impact of disease by estimating all potential impacts of disease[19]. Currently, analyses such as CEAs are being used to inform insurance coverage decisions, such as in the UK where NICE explicitly considers cost-effectiveness, and in the US where some insurers will consider evaluations conducted by the Institute for Clinical and Economic Review when making coverage decisions[20]. Therefore, failure of CEAs to correctly identify the full burden of disease can lead to patients being denied medications that best address their needs. In improving the measurement of health utilities used to inform QALY in analyses, gathering health utility information from instruments that are more detailed than the EQ-5D or SF-36 could further highlight therapeutic value, as would a creation of methods to map results from disease-specific instruments to formal health utilities.

Conclusions

Results of this study should be considered in light of its limitations. Responses from NHANES are self-reported and subject to recall bias. Differences in patients assessed in NHANES across diseases could impact the findings; however, the sample collected was large and all results were weighted to be representative of the US population. Estimates of health utilities were based on a single source, and utility values can vary substantially between studies. But we would expect that relying on a single source for all diseases would increase comparability, as the utility values may differ depending on collection methods, but the relative burden for each disease would be consistent. While we compared across conditions and most CEAs compare interventions within a single disease area, we feel the same issues would apply in both situations. In the case where two interventions for a single condition can improve different aspects of health (e.g. cognitive functioning and physical functioning), there would still likely be a differential value based on each treatment depending on patient preferences. In this study, we did not set out to improve upon the QALY, rather point out limitations and support the need to improve the measurement of value. We recognize that it is difficult to combine different outcomes into a composite endpoint, or to capture heterogeneity when considering allocation decisions, but believe it is imperative to do so in future studies. Such studies could consider subgroups of populations with different preferences, and for each subgroup could speak to the value of interventions. Plausible ranges or other sorts of statements around the confidence and generalizability in the findings of a CEA could also better account for patient heterogeneity.

While the QALY is commonly used and has advantages over other approaches that fail to capture the quality of life, sole reliance on health utilities as currently captured, and failure to consider patient heterogeneity for treatment outcomes, is suboptimal for patients. If insurers use current CEA results to make coverage decisions and restrict access to treatments based on an incomplete estimate of burden of disease, this will lead to misallocation of resources and a withholding of therapies that patients find valuable.

Transparency

Declaration of funding

The work described in this manuscript was funded by Amgen.

Declaration of financial/other relationships

At the time of the analysis, T.Y. was an employee of Partnership for Health Analytics and Research LLC, a health services research company paid by Amgen to conduct this research. J.D.O., E.C., A.L.H. and M.S.B. have disclosed that they are employees of Partnership for Health Analytics and Research LLC, a health services research company paid by Amgen to conduct this research. Z.W. has disclosed that he is currently employed by Amgen Inc. and owns stock in Amgen Inc.

Author contributions

All authors (1) made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data; or the creation of new software used in the work; (2) drafted the work or revised it critically for important intellectual content; (3) approved the version to be published; and (4) agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All authors read and approved the final manuscript.

Conceptualization: J.D.O., Z.W., M.S.B.; methodology: T.Y., J.D.O., E.C., Z.W., M.S.B.; formal analysis and investigation: T.Y., J.D.O., E.C.; writing – original draft preparation: T.Y., J.D.O., E.C., A.L.H., M.S.B.; writing – review and editing: T.Y., J.D.O., E.C., Z.W., A.L.H., M.S.B.; funding acquisition: J.D.O., Z.W., M.S.B.; resources: T.Y., E.C.; supervision: J.D.O., Z.W., M.S.B.

Data availability statement

The datasets generated during and/or analyzed during the current study are available in the NHANES repository, https://www.cdc.gov/nchs/nhanes/index.htm.

Code availability statement

All data and materials as well as software application or custom code support published claims and comply with field standards.

ORCID

Michael S. Broder http://orcid.org/0000-0002-2049-5536

References

[1] Health Equality Europe. Understanding Health Technology Assessment (HTA) [Internet]. 2008 [cited 2020 Feb 4]. Available from: https://htai.org/wp-content/uploads/2018/02/PCISG-Resource-HEE_ENGLISH_PatientGuidetoHTA_Jun14.pdf

[2] Knapp M, Mangalore R. The trouble with QALYs … Epidemiol Psichiatr Soc. 2007;16(4):289–293.

[3] Duru G, Auray JP, Béresniak A, et al. Limitations of the methods used for calculating quality-adjusted life-year values. Pharmacoeconomics. 2002;20(7):463–473.

[4] Normand C. Measuring outcomes in palliative care: limitations of QALYs and the road to PalYs. J Pain Symptom Manage. 2009;38(1):27–31.

[5] IHME. United States of America [Internet]. The Institute for Health Metrics and Evaluation. 2020 [cited 2021 Aug 11]. Available from: http://www.healthdata.org/united-states

[6] Centers for Disease Control and Prevention. About the National Health and Nutrition Examination Survey [Internet]. The National Health and Nutrition Examination Survey. 2017 [cited 2019 Nov 1]. Available from: https://www.cdc.gov/nchs/nhanes/about_nhanes.htm

[7] Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep Wash DC 1974. 1985;100:126–131.

[8] Weschler D. Wechsler adult intelligence scale, 3rd ed (WAIS–III). San Antonio (TX): Psychological Corporation; 1987.

[9] Jaeger J. Digit symbol substitution test: the case for sensitivity over specificity in neuropsychological testing. J Clin Psychopharmacol. 2018;38(5):513–519.

[10] Jia H, Lubetkin EI. Impact of nine chronic conditions for US adults aged 65 years and older: an application of a hybrid estimator of quality-adjusted life years throughout remainder of lifetime. Qual Life Res. 2016;25(8):1921–1929.

[11] Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res. 2005;14(6):1523–1532.

[12] Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale (NJ): L. Erlbaum Associates; 1988.

[13] Pettitt D, Raza S, Smith J. The limitations of QALY: a literature review. J Stem Cell Res Ther [Internet]. 2016 [cited 2020 Feb 4]. Available from: https://www.omicsonline.org/open-access/the-limitations-of-qaly-a-literature-review-2157-7633-1000334.php?aid=70859

[14] Nord E, Enge AU, Gundersen V. QALYs: is the value of treatment proportional to the size of the health gain? Health Econ. 2010;19(5):596–607.

[15] Dolan P, Shaw R, Tsuchiya A, et al. QALY maximisation and people's preferences: a methodological review of the literature. Health Econ. 2005;14(2):197–208.

[16] Browne J, Cryer DR, Stevens W. Is the QALY fit for purpose? Am J Accountable Care. 2021;9(2):8–13.

[17] Scharff RL, Jessup A. Evaluating chronic disease for heterogeneous populations: the case of arthritis. Med Care. 2007;45(9):860–868.

[18] Hansen P, Devlin N. Multi-criteria decision analysis (MCDA) in healthcare decision-making. In: Oxford research encyclopedia of economics and finance [Internet]. Oxford University Press; 2019 [cited 2019 Dec 2]. Available from: https://doi.org/10.1093/acrefore/9780190625979.013.98

[19] Laing CM, Moules NJ. Social return on investment: a new approach to understanding and advocating for value in healthcare. J Nurs Adm. 2017;47(12):623–628.

[20] Alliance for Aging Research. What is ICER and how does it promote discriminatory drug pricing? [Internet] [cited 2021 Aug 12]. Available from: https://www.agingresearch.org/icer-facts/