



## RESEARCH ARTICLE

**REVISÉD** **Comparing the cost-per-QALYs gained and cost-per-DALYs averted literatures [version 2; referees: 3 approved]**Peter J. Neumann , Jordan E. Anderson, Ari D. Panzer, Elle F. Pope, Brittany N. D'Cruz, David D. Kim, Joshua T. Cohen

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

**v2** **First published:** 18 Jan 2018, 2:5 (doi: [10.12688/gatesopenres.12786.1](https://doi.org/10.12688/gatesopenres.12786.1))  
**Latest published:** 05 Mar 2018, 2:5 (doi: [10.12688/gatesopenres.12786.2](https://doi.org/10.12688/gatesopenres.12786.2))**Abstract****Background:** We examined the similarities and differences between studies using two common metrics used in cost-effectiveness analyses (CEAs): cost per quality-adjusted life year (QALY) gained and cost per disability-adjusted life year (DALY) averted.**Methods:** We used the Tufts Medical Center CEA Registry, which contains English-language cost-per-QALY gained studies, and the Global Cost-Effectiveness Analysis (GHCEA) Registry, which contains cost-per-DALY averted studies. We examined study characteristics, including intervention type, sponsor, country, and primary disease, and also compared the number of published CEAs to disease burden for major diseases and conditions across geographic regions.**Results:** We identified 6,438 cost-per-QALY and 543 cost-per-DALY studies published through 2016 and observed rapid growth for both literatures. Cost-per-QALY studies most often examined pharmaceuticals and interventions in high-income countries. Cost-per-DALY studies predominantly focused on infectious disease interventions and interventions in low and lower-middle income countries. We found that while diseases imposing a larger burden tend to receive more attention in the cost-effectiveness analysis literature, the number of publications for some diseases and conditions deviates from this pattern, suggesting “under-studied” conditions (e.g., neonatal disorders) and “over-studied” conditions (e.g., HIV and TB).  
**Conclusions:** The CEA literature has grown rapidly, with applications to diverse interventions and diseases. The publication of fewer studies than expected for some diseases given their imposed burden suggests funding opportunities for future cost-effectiveness research.**Keywords**

Quality-adjusted life years, Disability-adjusted life years, Cost-effectiveness

**Open Peer Review****Referee Status:** 

	Invited Referees		
	1	2	3
<b>version 2</b> published 05 Mar 2018			
<b>version 1</b> published 18 Jan 2018	 report	 report	 report

- 1 **Michael Drummond**, University of York, UK
- 2 **Kalipso Chalkidou**, Center for Global Development, UK  
**Alec Morton**, University of Strathclyde, UK
- 3 **Rachel Nugent**, RTI International, USA  
University of Washington, USA

**Discuss this article**

Comments (0)

**Corresponding author:** Peter J. Neumann ([pneumann@tuftsmedicalcenter.org](mailto:pneumann@tuftsmedicalcenter.org))

**Author roles:** **Neumann PJ:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Anderson JE:** Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Panzer AD:** Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Pope EF:** Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **D’Cruz BN:** Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Kim DD:** Investigation, Methodology, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Cohen JT:** Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**How to cite this article:** Neumann PJ, Anderson JE, Panzer AD *et al.* **Comparing the cost-per-QALYs gained and cost-per-DALYs averted literatures [version 2; referees: 3 approved]** Gates Open Research 2018, 2:5 (doi: [10.12688/gatesopenres.12786.2](https://doi.org/10.12688/gatesopenres.12786.2))

**Copyright:** © 2018 Neumann PJ *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Grant information:** Bill and Melinda Gates Foundation [OPP1171680].

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**First published:** 18 Jan 2018, 2:5 (doi: [10.12688/gatesopenres.12786.1](https://doi.org/10.12688/gatesopenres.12786.1))

**REVISED Amendments from Version 1**

Our revised version 2 addresses a number of important comments raised by reviewers.

First, we provide a more complete set of potential explanations for why QALY-based CEAs are more prevalent in high income countries than in low income and lower-middle income countries.

Second, we have revised the figures characterizing the relationship between disease burden and number of CEAs, putting number of studies on the vertical axis. This switch makes it easier to understand the figures because “number of studies” (vertical axis) is best understood as a “response” to diseases burden (horizontal axis). This rearrangement makes it clear that data points above the regression line represent diseases and conditions that are relatively over-studied.

We have also added a [Table 2](#), which compares actual studies conducted to predicted number of studies for each of the seven GBD regions.

Finally, we have made generation of the data used in this paper more easily reproducible by eliminating all manual steps and replacing those steps with computer code that we are making publicly available.

[Dataset 1](#) has been replaced with the following file: [Dataset 1](#). Cleaned QALY Database. [Dataset 2](#) has been replaced with the following file: [Dataset 2](#). Cleaned DALY Database. [Dataset 3](#) has been replaced with the following file: [Dataset 3](#). Regional and Disease Level Stratified Dataset. These datasets are available in our “Version 2” folder at OSF.

We likewise report all code used for analysis.

**See referee reports**

## Introduction

Researchers conducting cost-effectiveness analyses (CEAs) commonly use quality-adjusted life years (QALYs) or disability-adjusted life years (DALYs) as health outcome measures to account for both longevity and quality of life (or life with disability)<sup>1</sup>. These broadly applicable metrics facilitate the comparison of interventions across conditions and diseases.

Analysts have used these measures in different contexts and settings<sup>2–6</sup>. CEAs using the cost-per-QALY metric, which first appeared in the late 1970s, have typically focused on interventions in higher income settings<sup>7,8</sup>. In the 1990s, the World Bank and the World Health Organization (WHO) developed the DALY to quantify disease burden (reflecting both years of life lost (YLL) and years of life with disability (YLD))<sup>9,10</sup>. CEAs using DALYs have tended to focus on lower- and middle-income countries<sup>11</sup>.

QALYs and DALYs, which both quantify health related quality of life by assigning a value ranging from zero to one to each year of life, have somewhat different methodological underpinnings<sup>12</sup>. QALY preference weights range from 0 (corresponding to “dead”) to 1 (corresponding to a hypothetical state of “perfect health”) and reflect a set of health state “attributes,” “dimensions,” or “domains” – e.g., discomfort, mobility, depression, etc. – associated with an individual’s health condition. DALY weights have a similar intuitive interpretation, although for DALYs, 1 corresponds to “dead” and 0 corresponds to “perfect health.” For DALYs,

moreover, each weight corresponds not to a set of health state attributes but to a specific disease<sup>13</sup>.

DALY values have in the past depended on the age of the affected populations. “Age-weighting” reflected the idea that an additional life year accrued during childhood or old age has less value than a year accrued during young and middle adulthood, when productivity contributions to societal well-being are typically greatest<sup>14,15</sup>. Because the unequal treatment of different age groups raised substantial ethical concerns, however, the most recent DALY calculation methods omit age-weighting<sup>16</sup>.

We analyzed the cost-per-QALY gained and cost-per-DALY averted literatures to examine their growth and regional variation, and to investigate the extent to which the focus of each literature corresponds to those diseases and conditions imposing the largest burden on the population.

## Methods

### Data

**The cost-effectiveness analysis literature.** We analyzed two databases maintained by the Center for the Evaluation of Value and Risk in Health at Tufts Medical Center in Boston, Massachusetts: the Cost-Effectiveness Analysis (CEA) Registry ([www.cearegistry.org](http://www.cearegistry.org)), which contains information on cost-per-QALY studies, and the Global Health CEA Registry ([www.ghearegistry.org](http://www.ghearegistry.org)), which houses information on cost-per-DALY studies. Both registries contain information on PubMed-indexed, English-language CEAs published through 2016. Previous publications further detail the search strategies, data collection processes, and review methods, which are similar for the two registries<sup>5,6</sup>. We received an ethics exemption for this study because it did not involve human subjects. Data from these registries used in this analysis appear in [Dataset 1](#) and [Dataset 2](#); [Supplemental File 1](#) and [Supplemental File 2](#) document the variables in these datasets.

**Disease burden.** [Dataset 3](#) contains population disease burden estimates (total DALYs incurred), as reported by the Institute for Health Metrics and Evaluation (IHME), and stratified by Global Burden of Disease (GBD) Super Region<sup>17</sup>. Within each Super Region, we sub-stratified population burden by GBD level two disease category. [Dataset 3](#) also lists the number of articles from the cost-per-QALY literature and from the cost-per-DALY literature for each of these strata and substrata. We counted articles in more than one of the [Table 2](#) strata if, for example, they focused on two countries belonging to two distinct GBD Super Regions.

### Analysis

**Study characteristics.** Using data from [Dataset 1](#) and [Dataset 2](#), and definitions from the World Bank and the GBD initiative, we stratified studies by: GBD Super Region, World Bank income level, intervention type, study funding source category, prevention stage, and GBD category. As detailed in [Table 1](#), some of these categories are mutually exclusive, while others are not. We computed the proportion of studies in each stratum using total article counts for the cost-per-QALY and cost-per-DALY literature from [Dataset 1](#) and [Dataset 2](#), respectively.

**Table 1. Characteristics of published CEAs using cost-per-QALY and cost-per-DALY through 2016.**

	Cost-per-QALY studies	Cost-per-DALY studies	Overall
<b>Number of studies</b>	<b>6438</b>	<b>543</b>	<b>6981</b>
<b>GBD Super Region</b>			
High income	89%	20%	84%
Southeast Asia, East Asia, and Oceania	3%	11%	4%
Sub-Saharan Africa	1%	29%	3%
Multiple Regions#	1%	16%	2%
Latin America and Caribbean	1%	8%	2%
Central Europe, Eastern Europe, and Central Asia	1%	2%	1%
South Asia	0%	8%	1%
North Africa and Middle East	1%	2%	1%
NA	3%	3%	3%
<b>World Bank Income Category</b>			
Low-Income and Lower-Middle-Income	1%	43%	5%
Upper Middle-Income and High-Income	97%	37%	92%
Both	0%	17%	1%
None	2%	3%	2%
<b>Intervention*</b>			
Pharmaceutical	44%	32%	43%
Surgical	13%	8%	13%
Screening	12%	14%	12%
Care delivery	11%	17%	11%
Medical procedure	12%	4%	12%
Health education or behavior	9%	21%	10%
Immunization	6%	27%	8%
Other	19%	38%	20%
<b>Study funder*</b>			
Government	33%	47%	34%
Pharmaceutical or device company	28%	4%	27%
Foundation	10%	27%	11%
Healthcare organization^	4%	9%	5%
None/Not determined	24%	24%	24%
Other	8%	20%	9%
<b>Prevention stage*</b>			
Primary	15%	59%	18%
Secondary	16%	20%	16%
Tertiary	62%	38%	60%
<b>Global Burden of Disease Category</b>			
Neoplasms	18%	3%	17%
Cardiovascular and circulatory diseases	17%	5%	16%
Diabetes, urogenital, blood, and endocrine diseases	12%	5%	11%
Other communicable, maternal, neonatal, and nutritional disorders	9%	7%	9%
Musculoskeletal disorders	10%	1%	9%
Mental and behavioral disorders	6%	8%	6%
HIV/AIDS and tuberculosis	4%	20%	6%
Digestive diseases	4%	1%	4%
Diarrhea, LRI, and other common infectious diseases	2%	20%	3%
Other	18%	31%	19%

Key: # "Multiple regions" refers to studies that reported cost-effectiveness estimates for countries in different regions. ^ Health care organizations include insurance companies, hospitals, HMOs, WHO. \* Not mutually exclusive. GBD: Global burden of disease. GNI: Gross National Income. HMO: Health maintenance organization. LRI: Lower respiratory infection. WHO: World Health Organization.

**Table 2. Standardized residual deviation from projected number of studies for each disease, by GBD region.**

Disease Area	GBD Region							Summary across all GBD Regions	
	Asia and Oceania	Europe and Central Asia	High Income	Latin America and the Caribbean	North Africa and the Middle East	South Asia	Sub-Saharan Africa	Mean	Median (b)
Unintentional injury	-0.80	-0.81	-0.90	-0.85	-0.63	-0.81	-0.29	-0.72	-0.81
Transport injuries	-0.74	-0.94	-0.70	-0.98	-0.80	-0.65	-0.33	-0.73	-0.74
Liver Cirrhosis	-0.60	-0.96	-0.61	-0.89	-0.70	-0.69	-0.11	-0.65	-0.69
Neonatal Disorders	-0.65	-0.56	-0.25	-0.45	-0.73	-1.28	-1.55	-0.78	-0.65
Chronic Respiratory	-0.79	-0.59	-0.49	-0.28	-0.81	-0.91	-0.20	-0.58	-0.59
Nature, War, Legal	-0.49	-0.71	-0.24	-0.81	-0.69	-0.53	-0.04	-0.50	-0.53
Neurological Disorders	-0.53	-0.23	-0.87	-0.74	-0.17	-0.51	-0.03	-0.44	-0.51
Cardiovascular	-0.87	-0.98	1.51	0.67	-0.49	-0.89	0.14	-0.13	-0.49
Musculoskeletal	-0.63	-1.08	-0.31	-0.46	-0.47	-0.91	-0.33	-0.60	-0.47
Nutritional Deficiencies	-0.35	-0.43	-0.39	-0.66	-0.43	0.57	-0.49	-0.31	-0.43
Other, NCD	-0.38	-0.72	-1.13	0.05	-0.84	0.59	-0.34	-0.40	-0.38
Mental or behavior disorders	-0.61	0.84	-1.68	-0.91	-0.38	-0.13	-0.16	-0.43	-0.38
Maternal Disorders	-0.33	-0.52	0.00	-0.33	-0.32	0.51	0.46	-0.08	-0.32
Digestive Diseases	-0.25	-0.09	0.55	-0.74	-0.68	-0.57	0.04	-0.25	-0.25
NTD Malaria	-0.12	0.05	-0.20	0.01	0.83	0.09	0.05	0.10	0.05
Diabetes	0.75	1.14	1.61	0.21	0.50	-0.03	-0.57	0.51	0.50
Neoplasms	2.46	1.40	1.00	1.05	2.21	0.12	0.14	1.20	1.05
HIV and TB	1.23	0.64	0.84	1.53	0.51	2.52	3.83	1.59	1.23
Other Communicable, Maternal, Neonatal, or Nutrition	2.41	1.77	2.52	2.32	1.55	1.08	1.22	1.84	1.77
Diarrhea	1.28	2.39	-0.05	2.12	2.45	2.40	-1.66	1.27	2.12

Note:

(a) Values reported are Studentized residuals.

(b) Table presents diseases and conditions sorted by median deviation. The “unintentional injuries” category appears in the first table row because the median number of published studies was furthest below the corresponding projected number of studies by the greatest amount after standardization (Studentized residual of -0.81). The “diarrhea” category appears in the last table row because the median number of published studies exceeded the corresponding projected number of studies by the greatest amount after standardization (Studentized residual of 2.12).

Abbreviations: NCD (non-communicable disease), NTD (neglected tropical disease), HIV (human immunodeficiency virus), TB (tuberculosis)

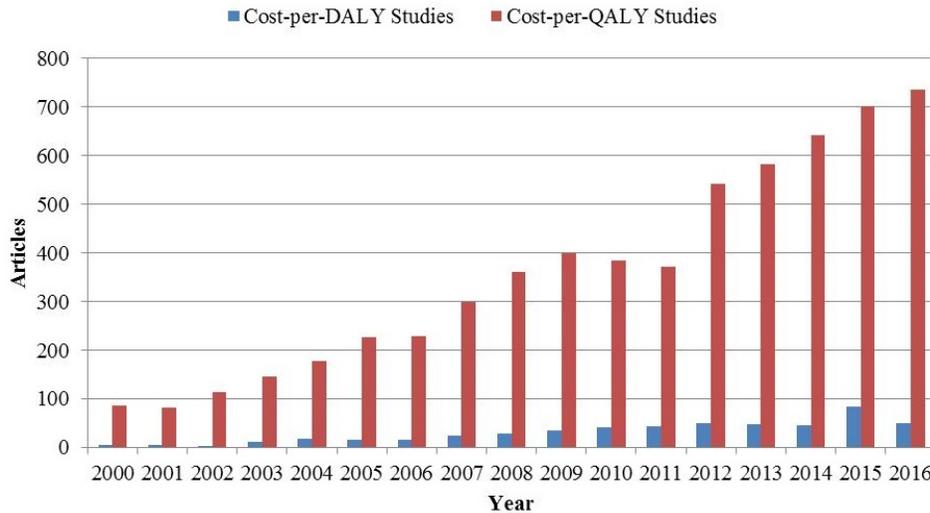
Based on these counts and proportions, we report the proportion of studies in each stratum, number of cost-per-QALY and cost-per-DALY studies published by year, proportion of published CEAs stratified by World Bank country income category and by study type (cost-per-QALY or cost-per-DALY), and number of cost-per-QALY and cost-per-DALY studies focusing on each country.

**Literature coverage vs. disease burden.** We characterized the relationship between the number of CEA studies (cost-per-QALY plus cost-per-DALY) focusing on each disease and the corresponding disease burden by regressing within each of the seven GBD super regions CEA publication count against disease burden using

ordinary least squares linear regression. Graphical plots of the regression results and original data for the three GBD regions with the most publications, and a table of standardized Studentized residuals for all seven regions (SAS Enterprise Guide version 7.1, Cary, NC) characterize which conditions are, in relative terms, over-studied or under-studied in each region, compared to the other conditions.

## Results

We identified 6,438 cost-per-QALY ([Dataset 1](#)) and 543 cost-per-DALY ([Dataset 2](#)) studies published through 2016. The number of published studies in the cost-per-QALY and cost-per-DALY literatures has increased steadily since 2000 ([Figure 1](#)).



**Figure 1. Published cost-per-DALY and cost-per-QALY studies by year.** Journals published 360 cost-per-QALY studies during the years 1976 through 2000. Journals published 13 cost-per-DALY studies during the years 1995 through 2000.

### Study characteristics

Cost-per-QALY studies have tended to focus on upper-middle income and high-income countries (97%); e.g. 2,321 studies focus on the United States, while 1,149 studies focus on the United Kingdom. Cost-per-DALY studies have focused to a much greater extent on low and lower-middle income countries (43%); e.g. 95 studies focus on India, 51 focus on China, and 90 studies focus on Uganda (Table 1, Figure 2, Figure 3A and 3B).

Tertiary prevention (treatment) dominates the cost-per-QALY registry (62%), whereas the cost-per-DALY registry focuses far more on primary prevention (59%). Conditions most frequently addressed by studies in the cost-per-QALY literature include non-communicable diseases, such as cancer (18%) and cardiovascular diseases (17%), whereas most cost-per-DALY registry studies target infectious diseases.

Foundations are the single largest source of non-governmental support for cost-per-DALY studies (27%), while pharmaceutical and device companies are the single largest source of non-governmental support for cost-per-QALY studies (28%).

We classified countries into the following World Bank income categories (quantities expressed in 2016 US dollars): low-income (GNI per capita < \$1,005), lower-middle income (GNI per capita of \$1,006 – \$3,955), upper-middle income (GNI per capita of \$3,956 – \$12,235), and high-income (GNI per capita > \$12,235)<sup>18</sup>. We used GBD Super Region definitions reported in the 2015 GBD study<sup>17</sup>.

In Figure 3A, we excluded one study classified as “international.” We excluded 145 studies because the country of study was unclear.

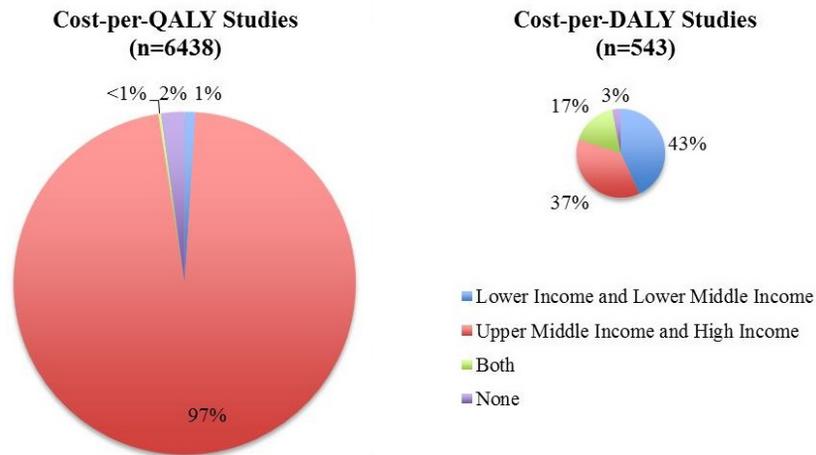
In Figure 3B, we excluded 13 studies classified as “international.” We excluded 17 studies because the country of study was unclear.

### Literature coverage vs. disease burden

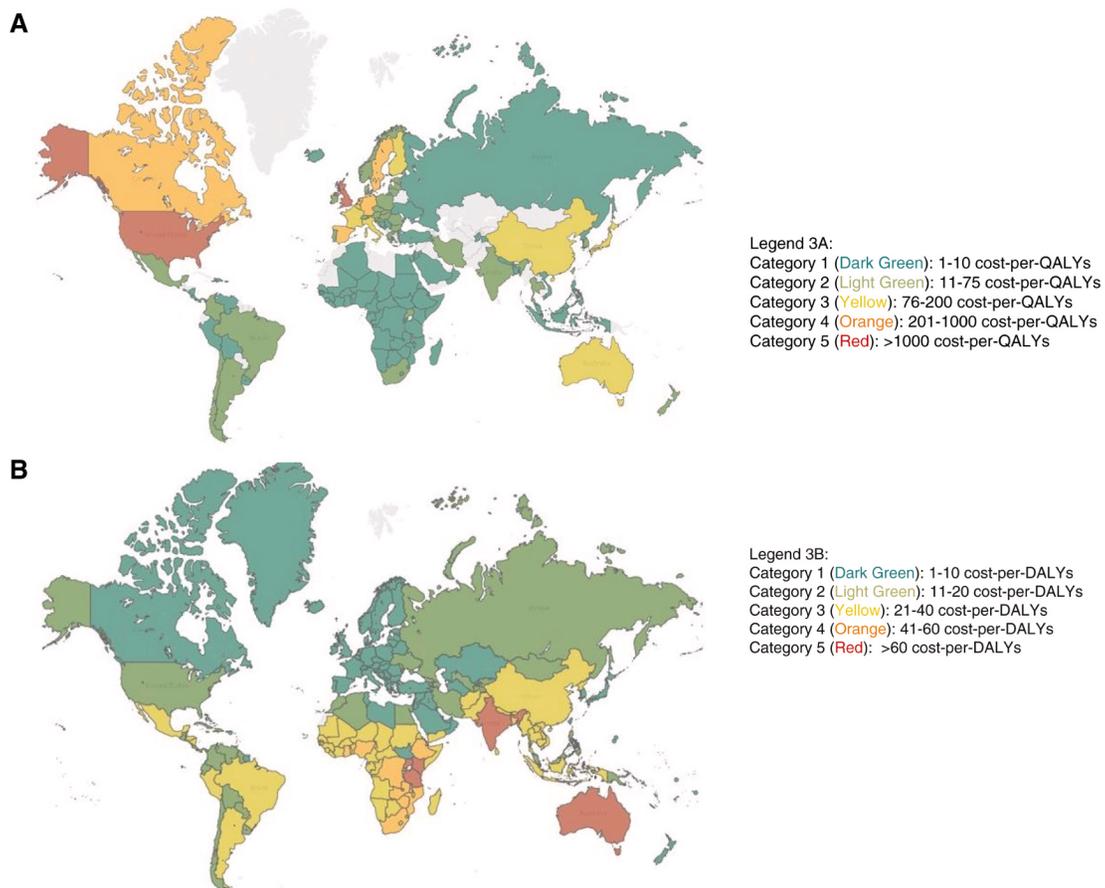
Neoplasms were the most studied diseases in Southeast Asia, East Asia, and Oceania (Figure 4A), while mental and behavioral disorders were less studied relative to their burden. High-income countries had relatively few studies addressing mental and behavioral disorders, and injuries (Figure 4B). Relative to burden, HIV/AIDS and tuberculosis were the most studied diseases in Sub-Saharan Africa, while this region reported fewer studies on nutritional deficiencies (Figure 4C).

Table 2 reports Studentized residuals from the ordinary least square regression for each region, along with the average and median of these residuals for each disease, across all seven GBD regions. Those results suggest that a number of conditions are uniformly “under-studied” because the residuals are negative in all seven regions (e.g., unintentional injuries, transport injuries, liver cirrhosis). Positive residuals across most regions indicate other conditions generally receive more attention than appears warranted by their burden (HIV and TB, neoplasms).

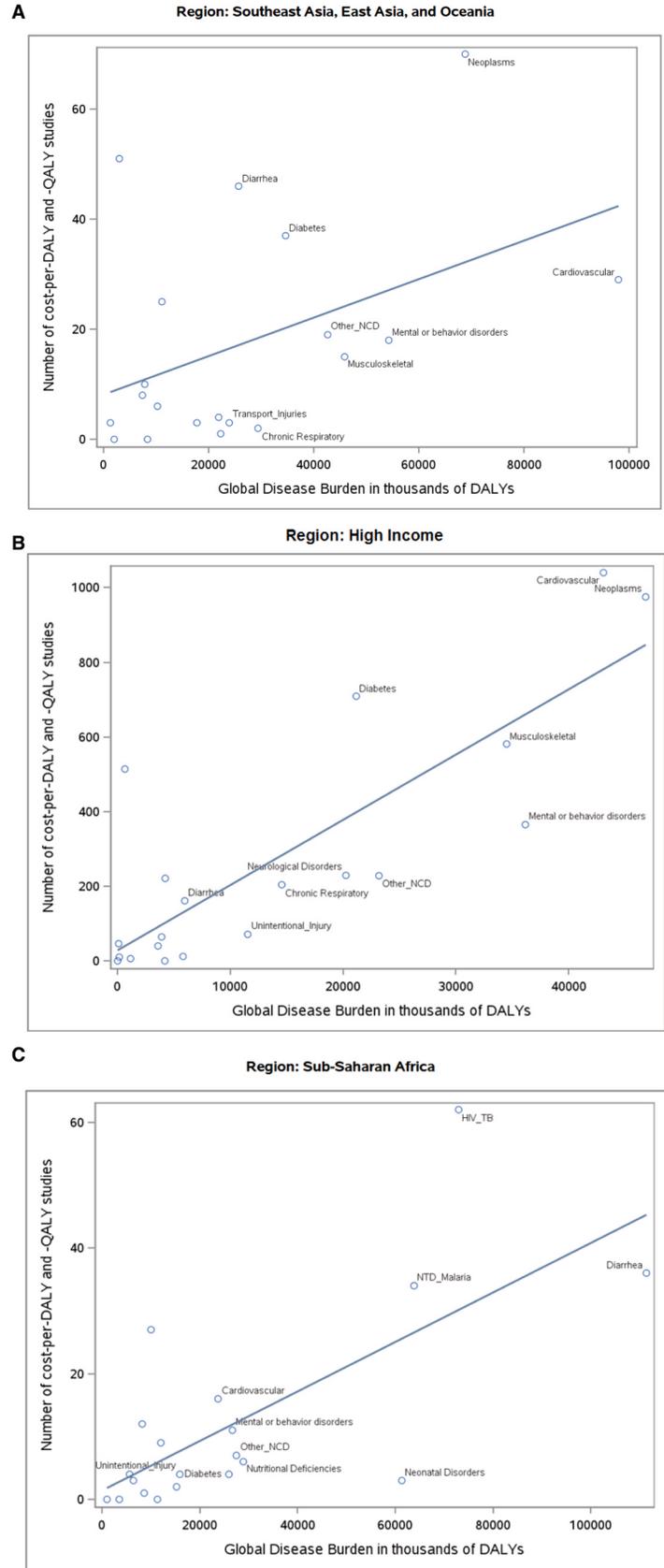
Each Figure 4 panel displays results for the top 10 diseases and includes a diagonal line that represents average studies published as a function of disease burden for each Super Region. The location of a plotted point to the “northwest” of this line indicates a disease that is relatively “over-studied” within that region, because the number of published studies exceeds, on average, the number published studies for other diseases imposing the same burden on the population. The location of a plotted point to the “southeast” indicates a disease that is relatively under-studied.



**Figure 2. Cost-per-QALY vs. cost-per-DALY studies by world bank income level.** The area of each pie chart is proportional to the number of studies catalogued in each registry.



**Figure 3. Geographic distribution of Cost-per-QALY and Cost-per-DALY studies.** The maps present the number of cost-per-QALY studies (Figure 3A) and cost-per-DALY studies (Figure 3B) for each country. Gray indicates countries with no associated studies. If a study reported a cost-effectiveness estimate for two or more countries, we counted a CEA for each country (e.g. if a study reported an intervention's cost-effectiveness ratio for both Canada and the United States, we incremented the study count in both countries). If a study reported a "global" cost-effectiveness ratio, we excluded it from all country counts. We also excluded from these counts studies that did not clearly specify an applicable country or region.



**Figure 4.** Number of CEAs vs. normalized disease burden for selected diseases and GBD Super Regions. **(A)** Southeast Asia, East Asia, and Oceania. **(B)** High Income Countries. **(C)** Sub-Saharan Africa.

## Discussion

Our review reveals a notable increase in the publication of cost-per-QALY and cost-per-DALY studies since 2000, thus making ever more cost-effectiveness information available to aid decision makers in their efforts to prioritize resources. The literature spans a wide range of interventions, diseases, and geographic regions.

The data demonstrate key differences between the cost-per-QALY and cost-per-DALY literatures (Table 1). For example, the cost-per-QALY literature tends to focus on high-income countries, while cost-per-DALY studies focus more on lower- and middle-income income nations. Differences extend to the types of interventions and diseases represented: cost-per-QALY studies tend to address diseases prevalent in wealthier countries (e.g., cardiovascular disease and cancer), while cost-per-DALY studies address diseases more prevalent in low-income countries (e.g., infectious diseases, such as tuberculosis and HIV). The two literatures also differ in terms of the interventions on which they focus. More cost-per-QALY studies evaluate pharmaceuticals, while cost-per-DALY studies focus more often on immunizations.

Several factors may explain why cost-per-QALY studies predominate in high-income countries, while cost-per-DALY studies are more popular in lower and middle-income countries. The differences could, for example, reflect the availability of health utility weights, needed to estimate QALYs, in high-income countries and the lack of such information in lower-income settings. Researchers conducting CEAs in countries with limited data capacity may find it easier and less expensive to use the cost-per-DALY metric.

The differences could also reflect the preferences and traditions of organizations that fund CEA studies. Foundations funding global health research may prefer the DALY metric, given the historic use of DALYs to measure global disease burden. In contrast, health authorities in high-income countries (e.g., the National Institute for Health and Care Excellence (NICE) in the United Kingdom) have tended to recommend the use of QALYs in CEAs. The geographic differences between the cost-per-QALY and cost-per-DALY literature deserve further investigation, as our effort did not gather information on why authors used these measures.

Our data also indicate inconsistencies between literature coverage and disease burden. Some diseases and conditions (e.g., cardiovascular disease and mental health in Southeast Asia, South Asia and Oceania) are relatively “under-studied,” while other diseases and conditions (e.g., HIV and TB in all regions) are relatively “over-studied”.

There is no clear explanation for these inconsistencies. As we have noted elsewhere, decisions to fund or conduct economic evaluations reflect not just the disease burden imposed by the targeted condition, but also the number of promising interventions or programs<sup>19,20</sup>. Because specialty drugs for diseases such as cancer represent important new interventions in high-income countries, and because pharmaceutical companies have the resources and incentive to characterize value for those interventions, much of the cost-per-QALY literature has recently focused on specialty drug therapies. These financial incentives are less pronounced in the lower- and middle-income countries that are much more the focus of the cost-per-DALY literature. In addition to disease burden,

priorities in the cost-per-DALY literature may reflect the visibility and emotional salience of diseases, the influence of advocacy groups, the vagaries of reimbursement decisions<sup>19</sup>, and institutional priorities of the organizations sponsoring the research.

In any case, the incongruities we observed between literature coverage and disease burden raise important questions about opportunities for the re-direction of future CEA research funding so that resources for such research can generate the highest return on investment.

Our work has the following limitations. First, the databases we used are restricted to English-language articles indexed in PubMed. This restriction may have depressed the number of cost-per-DALY studies we identified to a greater extent proportionally than it may have depressed the number of cost-per-QALY studies we identified because a smaller proportion of the cost-per-DALY literature focuses on English-speaking countries. Second, categorizing studies (e.g., whether an intervention targets primary or secondary prevention) depends on judgment, and other researchers may have classified articles differently.

In the future it will be important to further explore trends in the CEA literature in terms of diseases and geographic regions covered, funding patterns among donor organizations, the country of origin or study authors, the prevalence and patterns of CEAs published in languages other than English, the variation in methods used in analyses, and whether published studies address society’s most pressing needs<sup>21</sup>. It will also be useful to continue to investigate the methodological underpinnings of QALYs and DALYs and how much the choice of metric influences CEA results and the decisions based on them<sup>22,23</sup>.

## Data availability

We have made the data used in this analysis available through the Open Science Foundation (OSF): <http://doi.org/10.17605/OSF.IO/3BEK5><sup>24</sup>.

License: CC0 1.0 Universal.

### Dataset 1. Cleaned QALY Database.

Includes the cost-per-QALY data used in this paper.

### Dataset 2. Cleaned DALY Database.

Includes the cost-per-QALY data used in this paper.

### Dataset 3. Regional and disease level stratification dataset.

Includes disease burden and literature coverage data used in this paper.

## Competing interests

No competing interests were disclosed.

## Grant information

Bill and Melinda Gates Foundation [OPP1171680].

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

## Supplementary material

**Supplementary File 1. Cost-per-QALY manual.** Documents the variables collected in the cost-per-QALY database.

[Click here to access the data.](#)

**Supplementary File 2. Cost-per-DALY manual.** Documents the variables collected in the cost-per-DALY database.

[Click here to access the data.](#)

## References

- Neumann PJ, Sanders GD, Russell LB, *et al.*: **Cost-Effectiveness in Health and Medicine.** 2nd Edition, New York, NY: Oxford University Press; 2016.  
[Reference Source](#)
- Airoidi M, Morton A: **Adjusting life for quality or disability: stylistic difference or substantial dispute?** *Health Econ.* 2009; **18**(11): 1237–1247.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Teerawattananon Y, Tantivess S, Yamabhai I, *et al.*: **The influence of cost-per-DALY information in health prioritisation and desirable features for a registry: a survey of health policy experts in Vietnam, India and Bangladesh.** *Health Res Policy Syst.* 2016; **14**(1): 86.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hutubessy R, Chisholm D, Edejer TT: **Generalized cost-effectiveness analysis for national-level priority-setting in the health sector.** *Cost Eff Resour Alloc.* 2003; **1**(1): 8.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Neumann PJ, Thorat T, Zhong Y, *et al.*: **A Systematic Review of Cost-Effectiveness Studies Reporting Cost-per-DALY Averted.** *PLoS One.* 2016; **11**(12): e0168512.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Neumann PJ, Thorat T, Shi J, *et al.*: **The changing face of the cost-utility literature, 1990-2012.** *Value Health.* 2015; **18**(2): 271–277.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gold MR, Siegel JE, Russell LB, *et al.*: **Cost-Effectiveness in Health and Medicine.** New York: Oxford University Press; 1996.  
[Reference Source](#)
- Drummond MF, Sculpher MJ, Torrance GW, *et al.*: **Methods for the Economic Evaluation of Health Care Programmes.** 3rd ed. Oxford, UK: Oxford University Press; 2014.  
[Reference Source](#)
- World Bank: **World Development Report 1993; Investing in Health.** New York: Oxford University Press, 1993.  
[Reference Source](#)
- Murray CJ, Salomon JA, Mathers CD, *et al.*: **Summary measures of population health: concepts, ethics, measurement and applications.** Geneva, 2002.  
[Reference Source](#)
- Deveesschauwer B, Havelaar AH, Maertens de Noordhout C, *et al.*: **DALY calculation in practice: a stepwise approach.** *Int J Public Health.* 2014; **59**(3): 571–574.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Sassi F: **Calculating QALYs, comparing QALY and DALY calculations.** *Health Policy Plan.* 2006; **21**(5): 402–408.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Gold MR, Stevenson D, Fryback DG: **HALYS and QALYS and DALYS, Oh My: similarities and differences in summary measures of population Health.** *Annu Rev Public Health.* 2002; **23**: 115–134.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Arnesen T, Nord E: **The value of DALY life: problems with ethics and validity of disability adjusted life years.** *BMJ.* 1999; **319**(7222): 1423–1425.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Robberstad B: **QALYs vs DALYs vs LYs gained: what are the differences, and what difference do they make for health care priority setting?** *Norsk Epidemiologi.* 2005; **15**(2): 183–191.  
[Publisher Full Text](#)
- Murray CJ, Ezzati M, Flaxman AD, *et al.*: **GBD 2010: design, definitions, and metrics.** *Lancet.* 2012; **380**(9859): 2063–2066.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Institute for Health Metrics and Evaluation. 2017.  
[Reference Source](#)
- World Bank Country and Lending Groups. 2017.  
[Reference Source](#)
- Neumann PJ, Rosen AB, Greenberg D, *et al.*: **Can we better prioritize resources for cost-utility research?** *Med Decis Making.* 2005; **25**(4): 429–36.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Drummond M: **Referee Report For: Comparing the cost-per-QALYs gained and cost-per-DALYs averted literatures [version 1; referees: 3 approved].** *Gates Open Res.* 2018; **2**: 5.  
[Publisher Full Text](#)
- Santatiwongchai B, Chantarastapornchit V, Wilkinson T, *et al.*: **Methodological variation in economic evaluations conducted in low- and middle-income countries: information for reference case development.** *PLoS One.* 2015; **10**(5): e0123853.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Airoidi M, Morton A: **Adjusting life for quality or disability: stylistic difference or substantial dispute?** *Health Econ.* 2009; **18**(11): 1237–47.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Robberstad B: **QALYs vs DALYs vs LYs gained: What are the differences, and what difference do they make for health care priority setting?** *Norsk Epidemiologi.* 2009; **15**(2).  
[Publisher Full Text](#)
- Neumann P: **A comparison of cost-effectiveness analyses reporting cost-per-QALYs gained and cost-per-DALYs averted.** 2018.  
[Data Source](#)

## Open Peer Review

Current Referee Status:   

---

### Version 1

Referee Report 05 February 2018

doi:[10.21956/gatesopenres.13846.r26225](https://doi.org/10.21956/gatesopenres.13846.r26225)



**Rachel Nugent** <sup>1,2</sup>

<sup>1</sup> RTI International, Seattle, WA, USA

<sup>2</sup> Department of Global Health, University of Washington, Seattle, WA, USA

The authors have provided a useful summary of the up-to-date contents of the Tufts Medical Center CEA Registry and Global Health CEA Registry, which they manage. In particular, they contrast the contents of the two databases in regard to number of studies, geographic, disease burden, and disease-specific content. This provides a useful - if somewhat simplistic - overview of the availability and contents of current CEA studies. A few comments regarding the results as presented are provided below, along with a few suggestions about additional ways to interrogate the databases.

1. I am interested to see the time series of cost per QALY and cost per DALY studies presented in Figure 1. There is nothing especially surprising here for anyone working in the field, and it is heartening to see the steady increase in economic evaluations for health. I would appreciate the authors highlighting a few aspects of the databases that help one interpret the data. The methods clearly state that the databases draw from English-language articles indexed in PubMed, but it would be worth underscoring that selection creates a downward bias on the true number of cost per DALY studies. It would be very interesting to know if any literature has assessed the change over time of economic evaluations in local-language journals, which would provide an additional signal of the state of economic capacity in LMIC regions.
2. The authors have a rich longitudinal database that could be further analyzed to assess such questions as how changes in the funding sources or disease patterns over time affect the number of cost per QALY or cost per DALY studies. Looking specifically at the cost per DALY numbers over time, how does one understand the growth in study numbers? Is it strongly correlated with growth in global health funding? (I would guess so), and are numbers of disease-specific studies correlated with change in disease burden? (I would guess not). The authors could show rates of growth year by year which would make comparisons across years and types of studies easier.
3. The metric of "under-" and "over-" studied as determined by the DALY burden is also interesting and mostly unsurprising. Perhaps more could be said about the countries and sub-regions that show up green on both maps - such as North Africa, Middle East, and parts of Latin America. Those are the regions truly deficient in economic evaluations. Another point about the literature coverage relative to disease burden is to consider the demographics of the respective Super Regions. Since sub-Saharan Africa and South Asia have younger populations, they also merit more analysis of childhood conditions. If an age-specific disease burden measure were used as

the scalar, would the conclusions about "over-" and "under-" studied be the same?

4. Like other reviewers, I find some issue with the statement about "historic proclivities" driving the choice between cost per QALY and cost per DALY, but for a different reason. The methodological underpinnings of the two measures require different types of data, some of which is culturally or contextually determined. Measuring disease prevalence is more straightforward - albeit not simple - than measuring attributes and states of health, and therefore more readily available in countries with limited data capacity; thus creating the means to produce more cost per DALY studies.

**Is the work clearly and accurately presented and does it cite the current literature?**

Yes

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**Referee Expertise:** Non-communicable disease economic evaluation

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 27 Feb 2018

**Peter Neumann,**

**Comments from Rachel Nugent, RTI International, Seattle, WA, USA; Department of Global Health, University of Washington, Seattle, WA, USA**

***Comment #1. The authors have provided a useful summary of the up-to-date contents of the Tufts Medical Center CEA Registry and Global Health CEA Registry, which they manage. In particular, they contrast the contents of the two databases in regard to number of studies, geographic, disease burden, and disease-specific content. This provides a useful - if somewhat simplistic - overview of the availability and contents of current CEA studies. A few comments regarding the results as presented are provided below, along with a few suggestions about additional ways to interrogate the databases.***

***I am interested to see the time series of cost per QALY and cost per DALY studies***

***presented in Figure 1. There is nothing especially surprising here for anyone working in the field, and it is heartening to see the steady increase in economic evaluations for health. I would appreciate the authors highlighting a few aspects of the databases that help one interpret the data. The methods clearly state that the databases draw from English-language articles indexed in PubMed, but it would be worth underscoring that selection creates a downward bias on the true number of cost per DALY studies. It would be very interesting to know if any literature has assessed the change over time of economic evaluations in local-language journals, which would provide an additional signal of the state of economic capacity in LMIC regions.***

Response:

We have added the following text to a new limitations section of the Discussion:

*... the databases we used are restricted to English-language articles indexed in PubMed. This restriction may have depressed the number of cost-per-DALY studies we identified to a greater extent proportionally than it may have depressed the number of cost-per-QALY studies we identified because a smaller proportion of the cost-per-DALY literature focuses on English-speaking countries.*

***Comment #2. The authors have a rich longitudinal database that could be further analyzed to assess such questions as how changes in the funding sources or disease patterns over time affect the number of cost per QALY or cost per DALY studies. Looking specifically at the cost per DALY numbers over time, how does one understand the growth in study numbers? Is it strongly correlated with growth in global health funding? (I would guess so), and are numbers of disease-specific studies correlated with change in disease burden? (I would guess not). The authors could show rates of growth year by year which would make comparisons across years and types of studies easier.***

Response:

These are interesting questions, although we believe they go beyond the scope of what we set out to address. We have added text to the Discussion section of the paper to note areas for future research, including trends in the CEA literature in terms of diseases and geographic regions covered, funding patterns among donor organizations, and whether published studies correspond to society's most pressing needs.

***Comment #3. The metric of "under-" and "over-" studied as determined by the DALY burden is also interesting and mostly unsurprising. Perhaps more could be said about the countries and sub-regions that show up green on both maps - such as North Africa, Middle East, and parts of Latin America. Those are the regions truly deficient in economic evaluations. Another point about the literature coverage relative to disease burden is to consider the demographics of the respective Super Regions. Since sub-Saharan Africa and South Asia have younger populations, they also merit more analysis of childhood conditions. If an age-specific disease burden measure were used as the scalar, would the conclusions about "over-" and "under-" studied be the same?***

Response:

These points are likewise interesting. We defer to future researchers to organize the data as needed and conduct these analyses.

***Comment 4. Like other reviewers, I find some issue with the statement about "historic***

*proclivities" driving the choice between cost per QALY and cost per DALY, but for a different reason. The methodological underpinnings of the two measures require different types of data, some of which is culturally or contextually determined. readily available in countries with limited data capacity; thus creating the means to produce more cost per DALY studies.*

Response:

We have added text offering further explanation for the discrepancy between use of QALYs and DALYs by country wealth level (see response to Comment #1 from Michael Drummond).

**Competing Interests:** No competing interests were disclosed.

Referee Report 01 February 2018

doi:[10.21956/gatesopenres.13846.r26223](https://doi.org/10.21956/gatesopenres.13846.r26223)



**Kalipso Chalkidou<sup>1</sup>, Alec Morton<sup>2</sup>**

<sup>1</sup> Center for Global Development, London, UK

<sup>2</sup> Department of Management Science, University of Strathclyde, Glasgow, UK

Thank you for the chance to review this paper.

- This is a useful and timely study presenting informative analyses of the cost-per-QALY gained and cost-per-DALY averted literatures, relating them to geographical regions and to diseases and conditions prominent in these regions, as well as to wealth levels, types of intervention assessed and funding source. The authors draw conclusions as to the respective literatures' evolution over the years and to things such as how well these link to disease burdens in their respective geographies.
- We would challenge the classification of pharmaceuticals as "tertiary prevention/treatment". According to WHO, pharmaceuticals make up the bulk of OOP spending in most LICs (~77% based on the 2011 World Medicines Situation) and given fees, access and availability of facilities, self-medication is a major component of healthcare systems in LICs and LMICs.
- We were surprised almost half of the DALY studies have received government funding. Is it possible to tell whether this is national governments of LICs and LMICs or donor governments? Given DALYs are mostly used in LICs and LMICs, are the governments of these countries commissioning this work? According to another study which we believe is worth citing<sup>1</sup>, BMGF seems to be the single most commonly cited funder of DALY studies in LMICs. Our analysis as part of this paper (unpublished data) found that LIC government funded studies in malaria, TB and HIV studies (using DALYs (mostly) as an outcome measure), made up only 13%, 5% and 7% of the total in each disease area, respectively. Perhaps a more nuanced (eg broken down by decision maker global and domestic) analysis of funding source may reveal important messages assuming the data are available? Such a study would supplement nicely the PLoS paper cited earlier.
- Though probably not for this paper, perhaps a discussion as to why the discrepancy between QALYs and DALYs by wealth level and what the message may be for transitioning countries, is warranted. So the database could be expanded perhaps in the future to include data on the country of origin of authors, which would in turn allow capturing a likely (but unproven) trend from

poorer countries where publications come from mostly western authors funded by foreign donor foundations or governments, focus on MDGs and using DALYs to MICs/HICs where local authors funded by local money dominate, and with the focus shifting to NCDS and the use of QALYs. Such an analysis, if it confirms our hypothesis (and there are plenty of anecdotes from countries like the Philippines, China, Thailand, Brazil, Mexico and South Africa where national reference cases use QALYs as the outcome measure of preference), could then help reflect on what this might mean for countries in transition and the type of data and capacity they need to support their transition.

- Figure 1 is useful, but it would be helpful to show the time trends not just in the counts of the papers but in the composition of papers by GBD category and super-region, as well as intervention. A more developed Figure 1 would set up nicely a discussion of what the future might hold and which we touch on earlier in our discussion regarding transition.
- One of us (AM) has downloaded the cost/QALY data set to take a look and found that for 5895 out of 6438 records the field for publication year is blank. This info is needed to generate Figure 1 and so it should be there. It considerably lowers confidence in the integrity of the analysis when one discovers these things within a few seconds of downloading the database.
- Figure 4 is also very useful but why not show similar analysis for the other GBD regions? Why so few regions? In an increasingly multipolar world it seems highly appropriate to conclude that research priorities in each region should be different. If one could generate graphs for all the GBD regions that would strengthen that case and give a sense of the extent of global diversity.
- The authors write: “The contrast seems to reflect the historical proclivities rather than any inherent advantages for one metric’s use for a particular category of countries” – what might such “inherent advantages” be or indeed the historical proclivities (might the funding source have a role to play given the preference by BMGF a major funder of this work, for DALYs - <https://beta.nice.org.uk/Media/Default/About/what-we-do/NICE-International/projects/MEEP-report>.)? Why not standardise on the QALY as both the methodological (see Airoidi and Morton<sup>2</sup>) and empirical foundations of the method are more well-established and there is evidence that when domestic payers make investment decisions in HICs and UMICs, QALY is their preferred outcome?
- The findings relating to under- and over-studied conditions seem to us to be very interesting and relevant (perhaps more so than the QALY/DALY debate). Could the paper be retitled and/or the abstract rewritten to give these findings more prominence?

## References

1. Santatiwongchai B, Chantarastapornchit V, Wilkinson T, Thi boonboon K, Rattanavipapong W, Walker DG, Chalkidou K, Teerawattananon Y: Methodological variation in economic evaluations conducted in low- and middle-income countries: information for reference case development. *PLoS One*. 2015; **10** (5): e0123853 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Airoidi M, Morton A: Adjusting life for quality or disability: stylistic difference or substantial dispute?. *Health Econ*. 2009; **18** (11): 1237-47 [PubMed Abstract](#) | [Publisher Full Text](#)

## Is the work clearly and accurately presented and does it cite the current literature?

Partly

**Is the study design appropriate and is the work technically sound?**

Yes

**Are sufficient details of methods and analysis provided to allow replication by others?**

Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**

Yes

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Partly

**Competing Interests:** No competing interests were disclosed.

**We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 27 Feb 2018

**Peter Neumann,**

**Comments from Kalipso Chalkidou, Center for Global Development, London, UK**

**Alec Morton, Department of Management Science, University of Strathclyde, Glasgow, UK**

**Approved**

***Comment #1. This is a useful and timely study presenting informative analyses of the cost-per-QALY gained and cost-per-DALY averted literatures, relating them to geographical regions and to diseases and conditions prominent in these regions, as well as to wealth levels, types of intervention assessed and funding source. The authors draw conclusions as to the respective literatures' evolution over the years and to things such as how well these link to disease burdens in their respective geographies.***

Response:

No response needed.

***Comment #2. We would challenge the classification of pharmaceuticals as "tertiary prevention/treatment". According to WHO, pharmaceuticals make up the bulk of OOP spending in most LICs (~77% based on the 2011 World Medicines Situation) and given fees, access and availability of facilities, self-medication is a major component of healthcare systems in LICs and LMICs.***

Response:

Note that we make no assumptions about an intervention's prevention stage based on its type. For example, we do not assume that pharmaceuticals are tertiary treatments. Instead, we assign the prevention level based on how the article describes the disease and the treatment.

While we had intended the original text to provide examples of typical primary and tertiary treatments, we see that the presentation of the results may have been confusing. We have therefore eliminated those examples and just report the overall proportion of articles in two categories. The text now reads:

*Tertiary prevention (treatment) dominates the cost-per-QALY registry (62%), whereas the cost-per-DALY registry focuses far more on primary prevention (59%).*

***Comment #3a. We were surprised almost half of the DALY studies have received government funding. Is it possible to tell whether this is national governments of LICs and LMICs or donor governments? Given DALYs are mostly used in LICs and LMICs, are the governments of these countries commissioning this work? According to another study which we believe is worth citing<sup>1</sup>, BMGF seems to be the single most commonly cited funder of DALY studies in LMICs. Our analysis as part of this paper (unpublished data) found that LIC government funded studies in malaria, TB and HIV studies (using DALYs (mostly) as an outcome measure), made up only 13%, 5% and 7% of the total in each disease area, respectively. Perhaps a more nuanced (eg broken down by decision maker global and domestic) analysis of funding source may reveal important messages assuming the data are available? Such a study would supplement nicely the PLoS paper cited earlier.***

Response:

We do not have the information needed to assess whether the governments of these countries are commissioning the work. The final paragraph of the Discussion now cites both papers identified by the reviewer and notes the need for further research on this and on other issues.

***Comment #3b. Though probably not for this paper, perhaps a discussion as to why the discrepancy between QALYs and DALYs by wealth level and what the message may be for transitioning countries, is warranted. So the database could be expanded perhaps in the future to include data on the country of origin of authors, which would in turn allow capturing a likely (but unproven) trend from poorer countries where publications come from mostly western authors funded by foreign donor foundations or governments, focus on MDGs and using DALYs to MICs/HICs where local authors funded by local money dominate, and with the focus shifting to NCDS and the use of QALYs. Such an analysis, if it confirms our hypothesis (and there are plenty of anecdotes from countries like the Philippines, China, Thailand, Brazil, Mexico and South Africa where national reference cases use QALYs as the outcome measure of preference), could then help reflect on what this might mean for countries in transition and the type of data and capacity they need to support their transition.***

Response:

We have added text offering further explanation for the discrepancy between use of QALYs and DALYs by country wealth level (see response to Comment #1 from Michael Drummond) and on the need for further research in this area.

***Comment #4. Figure 1 is useful, but it would be helpful to show the time trends not just in the counts of the papers but in the composition of papers by GBD category and super-region, as well as intervention. A more developed Figure 1 would set up nicely a discussion of what the future might hold and which we touch on earlier in our discussion***

**regarding transition.**

Response:

We appreciate that providing time trends for other study characteristics, including GBD and super-region would be useful and could provide insight regarding the direction of the literature. In the revised paper, we have noted that as an area for future research and believe that as the cost-per-DALY literature in particular increases in size, the inferences that can be drawn will increase.

**Comment #5. One of us (AM) has downloaded the cost/QALY data set to take a look and found that for 5895 out of 6438 records the field for publication year is blank. This info is needed to generate Figure 1 and so it should be there. It considerably lowers confidence in the integrity of the analysis when one discovers these things within a few seconds of downloading the database.**

Response:

We very much appreciate the reviewers pointing out errors in our data extract. In response, we have regenerated the data extract, this time doing so by implementing all steps in a computer program to reduce the risk of introducing errors through manual manipulation of the original data. We are posting the computer program (written in STATA) and the extracted data. We have checked the distributions of the extracted data to make sure they appear to be reasonable.

Note that because we used the original dataset for our statistical analysis in version #1 of this paper, the errors in the extract did not affect the results.

**Comment #6. Figure 4 is also very useful but why not show similar analysis for the other GBD regions? Why so few regions? In an increasingly multipolar world it seems highly appropriate to conclude that research priorities in each region should be different. If one could generate graphs for all the GBD regions that would strengthen that case and give a sense of the extent of global diversity.**

Response:

We have chosen to include figures for only the three regions with the largest number of studies. But to address the reviewer's comment, we have also added a table that reports the standardized residual for each disease in each super region, relative to the regression line. We also report the mean and median residual for each disease (across all seven super regions) to characterize which diseases tend to be over- and under-studied in general.

**Comment #7. The authors write: "The contrast seems to reflect the historical proclivities rather than any inherent advantages for one metric's use for a particular category of countries" – what might such "inherent advantages" be or indeed the historical proclivities (might the funding source have a role to play given the preference by BMGF a major funder of this work, for DALYs - <https://beta.nice.org.uk/Media/Default/About/what-we-do/NICE-International/projects/MEEP-re> Why not standardise on the QALY as both the methodological (see Airoidi and Morton2) and empirical foundations of the method are more well-established and there is evidence that when domestic payers make investment decisions in HICs and UMICs, QALY is their preferred outcome?**

Response:

Making recommendations as to what measure the field should use is beyond the scope of this paper. We do, however, provide expanded text in an effort to explain why these measures are each used, and why the QALY measure is used more in high-income countries, and the DALY measure more in lower- and middle-income countries. See response to Comment #1 from Michael Drummond.

***Comment #8. The findings relating to under-and over-studied conditions seem to us to be very interesting and relevant (perhaps more so than the QALY/DALY debate). Could the paper be retitled and/or the abstract rewritten to give these findings more prominence?***

Response:

As this paper is the first comprehensive effort to describe the cost-per-DALY literature and compare it to the cost-per-QALY literature, we prefer to stick with emphasizing this aspect of the work in the title.

***Competing Interests:*** No competing interests were disclosed.

Referee Report 29 January 2018

doi:[10.21956/gatesopenres.13846.r26221](https://doi.org/10.21956/gatesopenres.13846.r26221)



**Michael Drummond**

Centre for Health Economics, University of York, York, UK

Neumann et al. examine the literature on economic evaluation through 2016, focusing specifically on studies measuring the health benefits in quality-adjusted life-years (QALYs) and disability-adjusted life-years (DALYs). A number of the findings of their research are as expected. First, cost per QALY studies have tended to focus on upper-middle and high-income countries, whereas cost per DALY studies have tended to focus on low and lower-middle income countries. This is likely to reflect the greater availability of preference values for health states in higher income countries and the preference of international donors, such as WHO and the World Bank, for studies estimating DALYs in lower income countries. Secondly, while the literature in both cost per QALY and cost per DALY studies is growing over time, there are more than 10 times the number of studies using QALYs than those using DALYs. This is likely to reflect the higher number of economist researchers and greater availability of funding for studies in high-income countries.

However, another finding of the research is not so easily explained. While the focus on topics for research, tertiary prevention (treatment) for studies using QALYs and primary prevention for studies using DALYs, it is surprising that the literature coverage is not closely aligned to disease burden in either high income or low income countries. Neumann et al. suggest that 'the most commonly studied diseases, regions and interventions may reflect the financial interests of the CEA funders'. One can see why this might be the case in higher income countries, where many studies are funded by pharmaceutical countries, but it's not clear why international donors might be favouring some diseases over others in lower income countries.

The analysis by Neumann et al. cannot directly answer that question, but one important factor driving economic evaluation in all countries is the number of promising interventions or programmes to evaluate.

In this sense, the literature on economic evaluation mostly follows the priorities for research of technology manufacturers or public health specialists. For example, in recent years the research priorities of pharmaceutical companies in higher income countries have focused on specialty drugs for diseases such as cancer. This could be driven by discoveries in basic research or the pursuit of profits, or both. However, in all countries one might expect priorities for research to be driven not by the absolute level of disease burden, but the potential for *modifying* that burden through the development and implementation of health care treatments and programmes.

One final issue touched on in the paper by Neumann et al. concerns the analytic choice between QALYs and DALYs in conducting economic evaluations. In commenting on the contrast in approach between higher and lower income countries, the authors state that 'this contrast seems to reflect the historic proclivities of health economist researchers, rather than any inherent advantages for one metric's use for a particular category of countries'. In my view this issue deserves much deeper investigation.

In many lower income countries, health economist researchers may not have a realistic choice of approach, as QALYs may not exist for the country concerned. But which approach should the analyst use in a country for which both QALYs and DALYs are available? Comparisons between QALYs and DALYs and the implications for health policy decisions have been discussed in the papers by Airoidi and Morton (2009)<sup>1</sup> and Robberstad (2005)<sup>2</sup>, with the conclusion that different decisions might be reached.

Although there are some minor differences in the theoretical constructs of QALYs and DALYs, two practical issues may be critical to the choice of approach. On the one hand QALYs are likely to be more 'bespoke' to the country where the study is being conducted and are more likely to reflect the health state preferences in the country concerned. However, on the other hand there is considerable variability in the methods used to elicit the preferences for health states in QALYs, which may threaten any standardized approach to decision-making. This issue has been recognized by the National Institute for Health and Care Excellence (NICE) in the United Kingdom, which, while recommending the use of QALYs, specifies the characteristics of the instrument that should be used to estimate them (NICE, 2013). By an extension of the same argument, an international donor requiring some standardization of approach to evaluation across several countries is likely to recommend the use of DALYs.

I answered 'Partly' to the question "Are sufficient details of methods and analysis provided to allow replication by others?" as access to the databases would be required for full replication.

### References

1. Airoidi M, Morton A: Adjusting life for quality or disability: stylistic difference or substantial dispute?. *Health Econ.* 2009; **18** (11): 1237-47 [PubMed Abstract](#) | [Publisher Full Text](#)
2. Robberstad B: QALYs vs DALYs vs LYs gained: What are the differences, and what difference do they make for health care priority setting?. *Norsk Epidemiologi.* 2009; **15** (2). [Publisher Full Text](#)

### Is the work clearly and accurately presented and does it cite the current literature?

Yes

### Is the study design appropriate and is the work technically sound?

Yes

### Are sufficient details of methods and analysis provided to allow replication by others?

Partly

**If applicable, is the statistical analysis and its interpretation appropriate?**

I cannot comment. A qualified statistician is required.

**Are all the source data underlying the results available to ensure full reproducibility?**

Yes

**Are the conclusions drawn adequately supported by the results?**

Yes

**Competing Interests:** No competing interests were disclosed.

**I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Author Response 27 Feb 2018

**Peter Neumann,**

**Comments from Michael Drummond, Centre for Health Economics, University of York, York, UK**

**Approved**

***Comment #1. Neumann et al. examine the literature on economic evaluation through 2016, focusing specifically on studies measuring the health benefits in quality-adjusted life-years (QALYs) and disability-adjusted life-years (DALYs). A number of the findings of their research are as expected. First, cost per QALY studies have tended to focus on upper-middle and high-income countries, whereas cost per DALY studies have tended to focus on low and lower-middle income countries. This is likely to reflect the greater availability of preference values for health states in higher income countries and the preference of international donors, such as WHO and the World Bank, for studies estimating DALYs in lower income countries. Secondly, while the literature in both cost per QALY and cost per DALY studies is growing over time, there are more than 10 times the number of studies using QALYs than those using DALYs. This is likely to reflect the higher number of economist researchers and greater availability of funding for studies in high-income countries.***

Response:

We agree with the reviewer comments and have revised the Discussion to incorporate these points. We have added the following text to the Discussion:

*Several factors may explain why cost-per-QALY studies predominate in high-income countries, while cost-per-DALY studies are more popular in lower and middle-income countries. The differences could, for example, reflect the availability of health utility weights in high-income countries and the lack of such information in lower-income settings. Researchers conducting CEAs in countries with limited data capacity may find it easier and less expensive to use the cost-per-DALY metric.*

*The differences could also reflect the preferences and traditions of organizations that fund CEA studies. Foundations funding global health research may prefer the DALY metric, given the historic use of DALYs to measure global disease burden. In contrast, health authorities in high-income countries (e.g., the National Institute for Health and Care Excellence (NICE) in the United*

Kingdom) have tended to recommend the use of QALYs in CEAs. The geographic differences between the cost-per-QALY and cost-per-DALY literature deserve further investigation, as our effort did not gather information on why authors used these measures.

**Comment #2.** However, another finding of the research is not so easily explained. While the focus on topics for research, tertiary prevention (treatment) for studies using QALYs and primary prevention for studies using DALYs, it is surprising that the literature coverage is not closely aligned to disease burden in either high income or low income countries. Neumann et al. suggest that ‘the most commonly studied diseases, regions and interventions may reflect the financial interests of the CEA funders’. One can see why this might be the case in higher income countries, where many studies are funded by pharmaceutical countries, but it’s not clear why international donors might be favouring some diseases over others in lower income countries.

Response:

We agree with the reviewer and have added the following paragraph to the Discussion: *There is no clear explanation for these inconsistencies. As we have noted elsewhere, decisions to fund or conduct economic evaluations reflect not just the disease burden imposed by the targeted condition, but also the number of promising interventions or programs* 19, 20. Because specialty drugs for diseases such as cancer represent important new interventions in high-income countries, and because pharmaceutical companies have the resources and incentive to characterize value for those interventions, much of the cost-per-QALY literature has recently focused on specialty drug therapies. These financial incentives are less pronounced in the lower- and middle-income countries that are much more the focus of the cost-per-DALY literature. In addition to disease burden, priorities in the cost-per-DALY literature may reflect the visibility and emotional salience of diseases, the influence of advocacy groups, the vagaries of reimbursement decisions<sup>19</sup>, and institutional priorities of the organizations sponsoring the research.

**Comment #3.** The analysis by Neumann et al. cannot directly answer that question, but one important factor driving economic evaluation in all countries is the number of promising interventions or programmes to evaluate. In this sense, the literature on economic evaluation mostly follows the priorities for research of technology manufacturers or public health specialists. For example, in recent years the research priorities of pharmaceutical companies in higher income countries have focused on specialty drugs for diseases such as cancer. This could be driven by discoveries in basic research or the pursuit of profits, or both. However, in all countries one might expect priorities for research to be driven not by the absolute level of disease burden, but the potential for modifying that burden through the development and implementation of health care treatments and programmes.

Response:

See response to Comment #3.

**Comment #4.** One final issue touched on in the paper by Neumann et al. concerns the analytic choice between QALYs and DALYs in conducting economic evaluations. In commenting on the contrast in approach between higher and lower income countries, the authors state that ‘this contrast seems to reflect the historic proclivities of health economist researchers, rather than any inherent advantages for one metric’s use for a particular category of countries’. In my view this issue deserves much deeper

**investigation.**

***In many lower income countries, health economist researchers may not have a realistic choice of approach, as QALYs may not exist for the country concerned. But which approach should the analyst use in a country for which both QALYs and DALYs are available? Comparisons between QALYs and DALYs and the implications for health policy decisions have been discussed in the papers by Airoidi and Morton (2009)<sup>1</sup> and Robberstad (2005)<sup>2</sup>, with the conclusion that different decisions might be reached.***

***Although there are some minor differences in the theoretical constructs of QALYs and DALYs, two practical issues may be critical to the choice of approach. On the one hand QALYs are likely to be more 'bespoke' to the country where the study is being conducted and are more likely to reflect the health state preferences in the country concerned. However, on the other hand there is considerable variability in the methods used to elicit the preferences for health states in QALYs, which may threaten any standardized approach to decision-making. This issue has been recognized by the National Institute for Health and Care Excellence (NICE) in the United Kingdom, which, while recommending the use of QALYs, specifies the characteristics of the instrument that should be used to estimate them (NICE, 2013). By an extension of the same argument, an international donor requiring some standardization of approach to evaluation across several countries is likely to recommend the use of DALYs.***

Response:

See response to Comment #1.

***Comment #5. I answered 'Partly' to the question "Are sufficient details of methods and analysis provided to allow replication by others?" as access to the databases would be required for full replication.***

Response:

We have made all data used in the analysis available, along with the computer code used for analyses and to create tables and figures.

***Competing Interests:*** No competing interests were disclosed.