Research

Bias in published cost effectiveness studies: systematic review

Chaim M Bell, David R Urbach, Joel G Ray, Ahmed Bayoumi, Allison B Rosen, Dan Greenberg, Peter J Neumann

Abstract

Objective To investigate if published studies tend to report favourable cost effectiveness ratios (below $20 000, $50 000, and $100 000 per QALY gained) and evaluate study characteristics associated with this phenomenon.

Design Systematic review.

Studies reviewed 494 English language studies measuring health effects in QALYs published up to December 2001 identified using Medline, HealthSTAR, CancerLit, Current Content, and EconLit databases.

Main outcome measures Incremental cost effectiveness ratios measured in dollars set to the year of publication.

Results Approximately half the reported incremental cost effectiveness ratios (712 of 1433) were below $20 000/QALY. Studies funded by industry were more likely to report cost effectiveness ratios below $20 000/QALY (adjusted odds ratio 2.1, 95% confidence interval 1.3 to 3.3), $50 000/QALY (3.2, 1.8 to 5.7), and $100 000/QALY (3.3, 1.6 to 6.8). Studies of higher methodological quality (adjusted odds ratio 0.58, 0.37 to 0.91) and those conducted in Europe (0.39, 0.33 to 1.1) and the United States (0.44, 0.26 to 0.76) rather than elsewhere were less likely to report ratios below $20 000/QALY.

Conclusion Most published analyses report favourable incremental cost effectiveness ratios. Studies funded by industry were more likely to report ratios below the three thresholds. Studies of higher methodological quality and those conducted in Europe and the US rather than elsewhere were less likely to report ratios below $20 000/QALY.

Introduction

Cost effectiveness analysis can help inform policy makers on better ways to allocate limited resources. Some form of cost effectiveness is now required for health interventions to be covered by many insurers. The quality adjusted life year (QALY) is used to compare the effectiveness of a wide range of interventions. Cost effectiveness analysis produces a numerical ratio—the incremental cost effectiveness ratio—in dollars per QALY. This ratio is used to express the difference in cost effectiveness between new diagnostic tests or treatments and current ones.

Interpreting the results of cost effectiveness analysis can be problematic, making it difficult to decide whether to adopt a diagnostic test or treatment. The threshold for adoption is thought to be somewhere between $20 000 ($11 300, $16 500)/QALY and $100 000/QALY, with thresholds of $50-60 000/QALY frequently proposed.

Regardless of the true value of the willingness of society to pay, studies of healthcare interventions would be expected to report a wide range of incremental cost effectiveness ratios. When published ratios cluster around a proposed threshold, bias may exist, and health policies based on their values may be flawed.

To describe the distribution of reported incremental cost effectiveness ratios and characteristics of studies associated with favourable ratios, we systematically reviewed cost effectiveness studies in health care that used QALYs as an outcome measure. We hypothesised that authors tend to report favourable incremental cost effectiveness ratios, such as those below $50 000 per QALY.

Methods

We conducted a systematic literature search of Medline, HealthSTAR, CancerLit, Current Contents Connect (all editions), and EconLit databases for all original cost effectiveness analyses published in English between 1976 and 2001 that expressed health outcomes in QALYs. Cost effectiveness analyses are reported as dollars per QALY. We used a standard data collection form, and two reviewers independently evaluated each study and abstracted the data. Disagreements were resolved by consensus. Details of the Tufts-NEMC CEA Registry (formerly the Harvard School of Public Health CEA Registry) are available online (http://tufts-nemc.org/cearegistry).

For each article, we documented the name of the journal, the year of publication, the disease category, and the country where the study was carried out. We used the Science Citation Index database to assign an impact factor for the year before publication to each journal. The sources of funding were identified as “industry” (partial or complete funding by a pharmaceutical or medical device company indicated in the manuscript) or “non-industry.” Studies for which a funding source was not listed were identified as “not specified.” We also assigned a quality score to each article, ranging from 1 (low) to 7 (high), based on the overall quality of the study methods, assumptions, and reporting practices.

Because cost effectiveness analyses often compare several programmes and include scenarios specific to patient subgroups or settings, each study may have contributed more than one cost effectiveness ratio. All cost effectiveness ratios were converted to US dollars at the exchange rate prevalent in the year of publication. Because we wanted to test whether the ratios targeted certain thresholds of the willingness of society to pay, such as $50 000/QALY, we did not adjust the ratios to constant dollars.

Statistical analysis

We analysed the distribution of all incremental cost effectiveness ratios and of the smallest and largest ratios from each study. We excluded nine ratios for which both the incremental cost and the incremental QALYs were negative. Although such interventions may be economically efficient, decision makers might not want to adopt interventions associated with reduced health.
Generalised estimating equations were used to evaluate study characteristics associated with incremental cost effectiveness ratios below the threshold values of $20 000, $50 000, and $100 000, as recommended previously.\textsuperscript{c,d,e} We used these equations because they take into account the correlation of cost effectiveness ratios derived from within the same study.\textsuperscript{f} We estimated odds ratios for associations between study characteristics and the presence of a favourable cost effectiveness ratio. Adjusted odds ratios were estimated by fitting a non-parsimonious model that included a priori predictor variables.

We used SAS statistical software version 8.2 for all analyses. Two sided P values less than 0.05 were considered significant.

## Results

We screened more than 3300 study abstracts and identified 533 original cost-utility analyses. Thirty nine studies were excluded because they did not report numerical incremental cost effectiveness ratios. In total, 1433 cost effectiveness ratios were reported in these 494 studies, with a median of 2.0 (interquartile range 1-3) and a range of 1-20 ratios per study. Overall, 130 incremental cost effectiveness ratios (9%) were reported as cost saving (they saved money and improved health simultaneously), 124 (9%) were dominated by their comparators (had worse health outcomes and increased costs), and 1179 (82%) increased costs but improved health outcomes.

Most studies were published in the 1990s (table 1). The citation impact factor in the year before publication was available for 449 studies (91%). Cardiovascular and infectious disease interventions were the most commonly studied. Most studies were from the United States. About 18% were sponsored by industry, almost half were sponsored by non-industry sources, and sponsorship could not be determined in 34% of studies.

Figure 1 shows the frequency distribution of all 1433 incremental cost effectiveness ratios. The median (interquartile range) ratio per QALY was $20 133 ($4520-74 400). Approximately half of the ratios (712; 50%) were below $20 000/QALY, two thirds (974; 68%) were below $50 000/QALY, and more than three quarters (1129; 79%) were below $100 000/QALY. When analysed according to study sponsorship, median (range) ratios per QALY were $13 083 ($3600-35 000) for those sponsored by industry and $27 400 ($4600-96 600) for those with non-industry sponsors. The median (range) cost effectiveness ratio per QALY for studies with unknown sponsorship was $18 900 ($4 960–64 300). Restricting the analysis to the lowest and highest ratios reported by each study yielded median ratios of $8784/QALY and $31 104/QALY (fig 2).

Several study characteristics were associated with reporting incremental cost effectiveness ratios below one or all three thresholds (table 2). The more quoted journals with a lower impact factor were significantly less likely to report ratios below $20 000/QALY (0.48, 0.33 to 0.70) and $50 000/QALY (0.57, 0.39 to 0.83). Within the multivariable model, the association with quality remained significant only for cost effectiveness ratios below $20 000/QALY (adjusted odds ratio 0.58, 0.37 to 0.91; table 2).

### Methodological quality

#### Journal impact factor

- 0-1.0: 214 (43)
- 1.1-2.0: 159 (32)
- 2.1-4.0: 137 (28)
- >4: 121 (25)

#### Region of study

- Europe: 119 (24)
- United States: 306 (62)
- Other: 70 (14)

### Results

#### Disease category

- Cardiovascular: 110 (22)
- Endocrine: 30 (6)
- Infectious: 94 (19)
- Musculoskeletal: 21 (4)
- Neoplastic: 76 (15)
- Neurological or psychiatric: 43 (9)
- Other: 120 (24)

#### Epistemic status

- Not available: 45 (9)

#### Methodological quality

- >1.0: 214 (43)
- 1.1-2.0: 159 (32)
- 2.1-4.0: 137 (28)
- >4: 121 (25)

#### Region of study

- Europe: 119 (24)
- United States: 306 (62)
- Other: 70 (14)

#### Disease category

- Cardiovascular: 110 (22)
- Endocrine: 30 (6)
- Infectious: 94 (19)
- Musculoskeletal: 21 (4)
- Neoplastic: 76 (15)
- Neurological or psychiatric: 43 (9)
- Other: 120 (24)

#### Methodological quality

- >1.0: 214 (43)
- 1.1-2.0: 159 (32)
- 2.1-4.0: 137 (28)
- >4: 121 (25)

### Discussion

About half of all cost effectiveness studies published over a 25 year period reported highly favourable incremental cost effectiveness ratios. The median (interquartile range) ratio per QALY was $20 133 ($4520-74 400). Approximately half of the ratios (712; 50%) were below $20 000/QALY, two thirds (974; 68%) were below $50 000/QALY, and more than three quarters (1129; 79%) were below $100 000/QALY. When analysed according to study sponsorship, median (range) ratios per QALY were $13 083 ($3600-35 000) for those sponsored by industry and $27 400 ($4600-96 600) for those with non-industry sponsors. The median (range) cost effectiveness ratio per QALY for studies with unknown sponsorship was $18 900 ($4 960–64 300). Restricting the analysis to the lowest and highest ratios reported by each study yielded median ratios of $8784/QALY and $31 104/QALY (fig 2).

Several study characteristics were associated with reporting incremental cost effectiveness ratios below one or all three thresholds (table 2). The more quoted journals with a lower impact factor were significantly less likely to report ratios below $20 000/QALY (0.48, 0.33 to 0.70) and $50 000/QALY (0.57, 0.39 to 0.83). Within the multivariable model, the association with quality remained significant only for cost effectiveness ratios below $20 000/QALY (adjusted odds ratio 0.58, 0.37 to 0.91; table 2).

### Discussion

About half of all cost effectiveness studies published over a 25 year period reported highly favourable incremental cost effectiveness ratios. The median (interquartile range) ratio per QALY was $20 133 ($4520-74 400). Approximately half of the ratios (712; 50%) were below $20 000/QALY, two thirds (974; 68%) were below $50 000/QALY, and more than three quarters (1129; 79%) were below $100 000/QALY. When analysed according to study sponsorship, median (range) ratios per QALY were $13 083 ($3600-35 000) for those sponsored by industry and $27 400 ($4600-96 600) for those with non-industry sponsors. The median (range) cost effectiveness ratio per QALY for studies with unknown sponsorship was $18 900 ($4 960–64 300). Restricting the analysis to the lowest and highest ratios reported by each study yielded median ratios of $8784/QALY and $31 104/QALY (fig 2).

Several study characteristics were associated with reporting incremental cost effectiveness ratios below one or all three thresholds (table 2). The more quoted journals with a lower impact factor were significantly less likely to report ratios below $20 000/QALY (0.48, 0.33 to 0.70) and $50 000/QALY (0.57, 0.39 to 0.83). Within the multivariable model, the association with quality remained significant only for cost effectiveness ratios below $20 000/QALY (adjusted odds ratio 0.58, 0.37 to 0.91; table 2).

### Discussion

About half of all cost effectiveness studies published over a 25 year period reported highly favourable incremental cost effectiveness ratios. The median (interquartile range) ratio per QALY was $20 133 ($4520-74 400). Approximately half of the ratios (712; 50%) were below $20 000/QALY, two thirds (974; 68%) were below $50 000/QALY, and more than three quarters (1129; 79%) were below $100 000/QALY. When analysed according to study sponsorship, median (range) ratios per QALY were $13 083 ($3600-35 000) for those sponsored by industry and $27 400 ($4600-96 600) for those with non-industry sponsors. The median (range) cost effectiveness ratio per QALY for studies with unknown sponsorship was $18 900 ($4 960–64 300). Restricting the analysis to the lowest and highest ratios reported by each study yielded median ratios of $8784/QALY and $31 104/QALY (fig 2).

Several study characteristics were associated with reporting incremental cost effectiveness ratios below one or all three thresholds (table 2). The more quoted journals with a lower impact factor were significantly less likely to report ratios below $20 000/QALY (0.48, 0.33 to 0.70) and $50 000/QALY (0.57, 0.39 to 0.83). Within the multivariable model, the association with quality remained significant only for cost effectiveness ratios below $20 000/QALY (adjusted odds ratio 0.58, 0.37 to 0.91; table 2).
effectiveness ratios of less than $20 000/QALY. More than half of the highest ratios reported by each study were below $50 000/QALY. In multivariable analysis, location of the study, methodological quality, and sponsorship were associated with a favourable cost effectiveness ratio. Studies sponsored by industry were more than twice as likely as studies sponsored by non-industry sources to report ratios below $20 000/QALY and were more than twice as likely as studies sponsored by industry to report ratios below $50 000/QALY or $100 000/QALY. Studies sponsored by industry were also more likely to be of lower methodological quality and to be published in journals with lower impact factors.

Few studies have described the distribution of reported incremental cost effectiveness ratios. We reviewed many studies, but we restricted our analysis to studies that measured health outcomes with QALYs. Although the QALY measure remains controversial, it has been endorsed by authoritative bodies, and it potentially allows cost effectiveness analysis to assess both allocative and technical efficiency. It would be interesting to examine whether the use of alternative measures of health outcome results in more or less favourable cost effectiveness ratios. A limitation of our analysis is that some cost effectiveness analyses used in decision making may not have been published. However, analyses of cost effectiveness assessments submitted to the National Institute for Health and Clinical Excellence (NICE) in the United Kingdom found that cost effectiveness ratios submitted by manufacturers were significantly lower than analyses of identical technologies performed by assessors from an academic centre.

**Publication bias**

We found relatively few published incremental cost effectiveness ratios between $50 000/QALY and $100 000/QALY; many were below $20 000/QALY and some were above $100 000/QALY. This indicates that cost effectiveness analyses tend to report “positive” or “negative” results but not intermediate results. There are three possible explanations for these findings. Firstly, they may reflect the true distribution of cost effectiveness ratios. A limitation of our analysis is that some cost effectiveness analyses used in decision making may not have been published. We reviewed many studies, but we restricted our analysis to studies that measured health outcomes with QALYs. Although the QALY measure remains controversial, it has been endorsed by authoritative bodies, and it potentially allows cost effectiveness analysis to assess both allocative and technical efficiency. It would be interesting to examine whether the use of alternative measures of health outcome results in more or less favourable cost effectiveness ratios. A limitation of our analysis is that some cost effectiveness analyses used in decision making may not have been published. However, analyses of cost effectiveness assessments submitted to the National Institute for Health and Clinical Excellence (NICE) in the United Kingdom found that cost effectiveness ratios submitted by manufacturers were significantly lower than analyses of identical technologies performed by assessors from an academic centre.

**Table 2** Characteristics of studies associated with favourable incremental cost effectiveness ratios according to three threshold values. Values are odds ratios (95% confidence intervals)

<table>
<thead>
<tr>
<th>Study characteristic</th>
<th>Code OR (95% CI)</th>
<th>Adjusted OR (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Publication year</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998-91</td>
<td>1.6 (0.98 to 2.7)</td>
<td>1.0</td>
</tr>
<tr>
<td>1992-6</td>
<td>1.3 (0.94 to 1.9)</td>
<td>1.0</td>
</tr>
<tr>
<td>1997-2001</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Journal impact factor†</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2</td>
<td>0.62 (0.42 to 0.91)</td>
<td>1.0</td>
</tr>
<tr>
<td>2-4</td>
<td>0.60 (0.42 to 0.86)</td>
<td>1.0</td>
</tr>
<tr>
<td>&gt;4</td>
<td>0.60 (0.42 to 0.86)</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Disease category</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Endocrine</td>
<td>1.3 (0.88 to 2.2)</td>
<td>1.0</td>
</tr>
<tr>
<td>Infectious</td>
<td>1.1 (0.66 to 1.7)</td>
<td>1.0</td>
</tr>
<tr>
<td>Musculoskeletal</td>
<td>1.4 (0.60 to 3.3)</td>
<td>1.0</td>
</tr>
<tr>
<td>Neoplastic</td>
<td>0.91 (0.56 to 1.5)</td>
<td>1.0</td>
</tr>
<tr>
<td>Neurological/psychiatric</td>
<td>0.76 (0.40 to 1.5)</td>
<td>1.0</td>
</tr>
<tr>
<td>Other</td>
<td>1.2 (0.75 to 1.8)</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Study funding source‡</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-industry</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Industry</td>
<td>2.2 (1.4 to 3.4)</td>
<td>1.0</td>
</tr>
<tr>
<td>United States</td>
<td>0.35 (0.21 to 0.57)</td>
<td>1.0</td>
</tr>
<tr>
<td>Other</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Methodological quality§</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>0.92 (0.64 to 1.3)</td>
<td>1.0</td>
</tr>
<tr>
<td>4.5-5.0</td>
<td>0.48 (0.33 to 0.70)</td>
<td>1.0</td>
</tr>
</tbody>
</table>

GALY: quality adjusted life year.
*Adjusted for all other study characteristics.
†Impact factor in the year before publication.
‡Funding by a pharmaceutical or medical device manufacturer.
§Mean score from two reviewers.
¶Mean score from two reviewers.
Research

ratios for healthcare interventions. Perhaps interventions that manufacturers deem to be economically unattractive are not brought to market.21 Secondly, analysts may not be interested in studying interventions with mid-range cost effectiveness ratios or some journals may not want to publish such studies. Thirdly, some cost effectiveness analyses may be modelled to yield favourable ratios or studies with unfavourable ratios may be suppressed. These findings do not tell us whether authors (and their sponsors) fail to report undesirable cost effectiveness results while reporting those that are positive. It is unclear whether publication bias occurs at a conscious or unconscious level. In any case, our results support concerns about the presence of significant and persistent bias in both the conduct and reporting of cost effectiveness analyses.22 23 It could be argued that all cost effectiveness analyses should be registered before they start, as for randomised clinical trials, but this may be unrealistic given the way they are currently conducted.24

If bias is an important explanation for our findings, our results indicate that the “target” cost effectiveness ratio for a healthcare intervention is between $20 000/QALY and $50 000/QALY. The true willingness of society to pay for an extra QALY is unknown, although some have tried to derive this willingness from economic principles, or to infer it from policy level decisions.25-27 In the US, the threshold most often used is $50 000/QALY, based on Medicare’s coverage of dialysis, although the criteria for selecting this value have been criticised.28 Furthermore, interventions may be adopted despite having an unfavourable ratio if other concerns such as disease burden and health equity are dealt with.29

Recent attempts to standardise the conduct and reporting of economic analyses and modelling studies may help prevent the manipulation of studies.30 31 32-34 Such guidelines offer the producers of cost effectiveness analyses, reviewers, readers, and journal editors a framework against which they can assess analyses, although the complexity of models and publication bias remain important issues.35 Electronic publishing could enhance transparency in modelling by making technical appendices available and improving the presentation of results.36 Furthermore, distribution of the underlying decision analysis models to the public should be considered. Industry sponsorship and the location of the study do not fully explain the tendency for incremental cost effectiveness ratios to fall below desirable thresholds. Fewer than 20% of studies reviewed were funded by industry, and another 33% did not specify their source of funding. Moreover, the median ratio for studies not sponsored by industry was well below the $50 000/QALY threshold. The lower ratios found in US studies cannot be fully accounted for by foreign exchange rates and inflation because the US results were similar to those found in European studies. The methodological quality of studies is a prerequisite for these analyses to compare health management strategies. A heightened awareness of the limitations of cost effectiveness analyses and potentially influential factors may help users to interpret the conclusions of these analyses.

Industry sponsorship and the location of the study do not fully explain the tendency for incremental cost effectiveness ratios to fall below desirable thresholds. Fewer than 20% of studies reviewed were funded by industry, and another 33% did not specify their source of funding. Moreover, the median ratio for studies not sponsored by industry was well below the $50 000/QALY threshold. The lower ratios found in US studies cannot be fully accounted for by foreign exchange rates and inflation because the US results were similar to those found in European studies. The methodological quality of studies is a prerequisite for these analyses to compare health management strategies. A heightened awareness of the limitations of cost effectiveness analyses and potentially influential factors may help users to interpret the conclusions of these analyses.

The paper was presented in abstract form at the Fifth International Congress on Peer Review and Biomedical Publication in Chicago, IL, 18-18 September 2005.

Contributors: All authors contributed to the conception and design of the project, revised the article critically for important intellectual content, and provided final approval of the final version. CMB, DRU, JGR, and AB analysed and interpreted the data and helped to draft the article. CMB had access to all of the data and takes responsibility for the integrity of the data and the accuracy of the data analysis. CMB and PJN are guarantors.

Funding: Agency for Health Care Research and Quality (RO1 HS10919). CMB and JGR are recipients of a phase 2 clinician scientist award and a new investigator award, both from the Canadian Institutes of Health Research. DRU holds a career scientist award from the Ontario Ministry of Health.

Competing interests: None declared.


31 Sharp DW. What can and should be done to reduce publication bias? The perspective of an editor. JAMA 1999;282:1390-1.

32 Chalmers TC, Critchlow CF, Reitman D. Minimizing the three stages of publication bias. JAMA 1990;263:1392-5.