BMC Medical Research Methodology

BioMed Central

Research article

On the probability of cost-effectiveness using data from randomized clinical trials

Andrew R Willan[1]

Address: [1]Department of Clinical Epidemiology and Biostatistics, HSC-2C, McMaster University, 1200 Main Street West, Hamilton, ON, L8N 3Z5, Canada and [2]The Centre for Evaluation of Medicines, St. Joseph's Hospital, Hamilton, ON, Canada

E-mail: willana@mcmaster.ca

Published: 6 September 2001

BMC Medical Research Methodology 2001, **1**:8

Received: 23 May 2001
Accepted: 6 September 2001

This article is available from: http://www.biomedcentral.com/1471-2288/1/8

Abstract

Background: Acceptability curves have been proposed for quantifying the probability that a treatment under investigation in a clinical trial is cost-effective. Various definitions and estimation methods have been proposed. Loosely speaking, all the definitions, Bayesian or otherwise, relate to the probability that the treatment under consideration is cost-effective as a function of the value placed on a unit of effectiveness. These definitions are, in fact, expressions of the certainty with which the current evidence would lead us to believe that the treatment under consideration is cost-effective, and are dependent on the amount of evidence (*i.e.* sample size).

Methods: An alternative for quantifying the probability that the treatment under consideration is cost-effective, which is independent of sample size, is proposed.

Results: Non-parametric methods are given for point and interval estimation. In addition, these methods provide a non-parametric estimator and confidence interval for the incremental cost-effectiveness ratio. An example is provided.

Conclusions: The proposed parameter for quantifying the probability that a new therapy is cost-effective is superior to the acceptability curve because it is not sample size dependent and because it can be interpreted as the proportion of patients who would benefit if given the new therapy. Non-parametric methods are used to estimate the parameter and its variance, providing the appropriate confidence intervals and test of hypothesis.

Introduction

In reporting cost-effectiveness analyses alongside clinical trials, authors [1–4] have used various definitions, estimation methods and interpretations for acceptability curves. Acceptability curves provide an excellent means of quantifying the stochastic uncertainty of the estimated incremental cost-effectiveness ratio (ICER) in relation to a particular value ascribed to a unit of effectiveness. It is the certainty, expressed as a probability, that the current evidence would lead us to believe that some new therapy is cost-effective, insofar as the ICER is less than the value ascribed to a unit of effectiveness. In addition, acceptability curves provide an estimator for the ICER and its confidence limits. However, acceptability curves are often interpreted and expressed as the probability that the new therapy is cost-effective. It is argued below that this is subject to misinterpretation, and an alternative definition for a parameter representing the probability that the new therapy is cost-effective is introduced. Data from a clinical trial can be used to make statistical inference

about this parameter. Furthermore, the inference provides a non-parametric estimator for the ICER and its confidence interval.

In a two-arm randomized control trial let e_{ji} and c_{ji} be the respective measures of effectiveness and cost for patient i on therapy j, where j = T (treatment), S (standard); i = 1, 2, . . .n_j ; and n_j is the number of patients randomized to therapy j. Typically, e_{ji} is the patient's survival time (perhaps quality-adjusted) from randomization to death or to the end of the period of interest. Let $\hat{\Delta}_e = \frac{1}{n_T}\sum_{i=1}^{n_T} e_{Ti} - \frac{1}{n_S}\sum_{i=1}^{n_S} e_{Si}$. Define $\hat{\Delta}_c$ similarly. Let $E(\hat{\Delta}_e) = \Delta_e$ and $E(\hat{\Delta}_c) = \Delta_c$, where E is the expectation function. Thus, the incremental cost-effectiveness ratio (ICER) is Δ_c/Δ_e, and is estimated by $\hat{\Delta}_c / \hat{\Delta}_e$. In addition, the incremental net benefit [5–9] (INB) is $\Delta_e\lambda - \Delta_c$, and is estimated by $\hat{\Delta}_e\lambda - \hat{\Delta}_c$, where λ is the value given to a unit of effectiveness. Typically the INB is expressed as a function of λ, allowing readers to provide the value they consider most relevant.

van Hout *et al.* [1] define the acceptability curve as "the probability that the [ICER] is under a certain acceptable limit", say λ. The acceptability curve, then, is a function of λ. If one assumes that the authors are referring to the true ICER ratio, the definition is Bayesian. In the same paper they define the acceptability curve in algebraic terms as $A(\lambda) = \int_{-\infty}^{\infty}\int_{-\infty}^{\lambda e} f(e,c)\,dc\,de$, where f is the joint probability density function for the random vector $(\hat{\Delta}_e, \hat{\Delta}_c)'$. Here the definition is not Bayesian, since it is the probability that, in repeated sampling, the random variable $\hat{\Delta}_e\lambda - \hat{\Delta}_c$ (*i.e.*, the observed net benefit) is greater than 0. In an illustration, the authors substitute the sample estimates for the model parameters in f to yield an empirical density function, call it $\hat{f}$, and refer to the acceptability curve as the probability that the ICER is acceptable. Briggs and Fenn [2] refer to the acceptability curve as "the probability an intervention is cost-effective in relation to different values of" λ. For estimation they propose using the integration of $\hat{f}$, as above, or determining the proportion of bootstrap re-samples in which $\hat{\Delta}_e\lambda - \hat{\Delta}_c$ is greater than 0.

Briggs[3] provides a purely Bayesian approach by defining the acceptability curve as the probability that $\Delta_e\lambda - \Delta_c$ is greater than 0. In an illustration the author interprets the acceptability curve as "probability of cost-effective". Assuming f is the density function for a bivariate normal random vector, and using an uninformative prior, the acceptability curve is given by $A(\lambda) = \int_0^{\infty} g(x)dx$, where g is the probability density function for a normal random variable with mean $\hat{\Delta}_e\lambda - \hat{\Delta}_c$ and variance $\sum_{j=S,T}\frac{1}{n_j}\left(\hat{\sigma}_j^2\lambda^2 + \hat{\omega}_j^2 - 2\hat{\sigma}_j\hat{\omega}_j\hat{\rho}_j\lambda\right)$, where $\hat{\sigma}_j^2$ and $\hat{\omega}_j^2$ are estimates of the variance of e_{ji} and c_{ji}, respectively, and $\hat{\rho}$ is an estimate of the correlation between e_{ji} and c_{ji} . This is exactly the same curve as determined by the integration of $\hat{f}$, given above, and, due to the symmetry, is equal to 1 minus the p-value of the test of the hypothesis $\Delta_e\lambda - \Delta_c < 0$. In reporting the results of a cost-effectiveness analysis, Raikou *et al.* [4] interpret the acceptability curve as the "probability that intervention is cost effective".

By rewriting $Pr(\Delta_e\lambda - \Delta_c > 0)$ as $Pr(\Delta_c/\Delta_e < \lambda)$, assuming $\Delta_e > 0$, the acceptability curve can be interpreted as the posterior distribution for the ICER. Defining $A(\lambda_\gamma) = \gamma$, the estimate for the ICER and its $(1-\alpha/2)100\%$ Baysian limits are given by $\lambda_{0.5}$, $\lambda_{\alpha/2}$ and $\lambda_{1-\alpha/2}$, respectively.

The interpretation that the acceptability curve is the probability that the intervention is cost-effective is not entirely accurate and could easily be misunderstood by policy makers. Consider the situation in which the observed INB for treatment is very small, but due to a very large sample size the acceptability curve at the value of λ of interest is 0.99. Attaching the label "the probability that the intervention is cost-effective" to this quantity could mislead policy makers into thinking that treatment is highly beneficial compared to standard. What, in fact, is high is our confidence that the INB, however small, is not zero. A more accurate interpretation of the acceptability curve is that it is a measure of the certainty with which the current evidence would lead us to believe that treatment is cost-effective, *i.e.*, $\Delta_e\lambda - \Delta_c > 0$. For a Bayesian, this is $Pr(\Delta_e\lambda - \Delta_c) > 0$, and for a frequentist, it is, assuming symmetry, 1 minus the p-value for the test of the hypothesis $\Delta_e\lambda - \Delta_c < 0$. This is not just the traditional confusion between statistical and clinical significance. In significance testing as the sample size increases, the var-

iance of the estimator decreases, but the magnitude of the parameter being estimated stays the same. However, as sample size increase the magnitude of the acceptability curve for a given λ increases.

In the next section we provide a more accurate definition for the probability that treatment is cost-effective. The definition contains no element of certainty, and is the probability of the "next" patient realizing a larger net benefit if he or she is given treatment rather than standard. Using data from a clinical trial, non-parametric methods can be used to estimate this probability, and uncertainty

Methods

The probability that treatment is cost-effective

Let $b_{ji}(\lambda) = e_{ji}\lambda - c_{ji}$. The quantity $b_{ji}(\lambda)$ is the net benefit, expressed in money, realized by patient i on therapy j. An alternative to the acceptability curve for quantifying the probability that treatment is cost-effective is defined as: $\theta(\lambda) \equiv$ the probability, for a given λ, that a patient will receive a larger net benefit with treatment rather than standard. Its definition is not Bayesian because it is a probability statement about random variables, not population parameters. Nonetheless, it relates directly to the notion of the probability of treatment being cost-effective. A policy maker can genuinely interpret θ(λ) as the probability of the "next" patient realizing a larger net benefit if he or she is given treatment rather than standard. Such a direct interpretation is not provided by the acceptability curve. The acceptability curve is the probability that the average net benefit of a group of patients of the same size as in the clinical trial will be greater is they receive treatment rather than standard.

To estimate θ(λ) we borrow methodology from receiver operating characteristic curves [10]. An estimate of θ(λ) is given by:

$$\hat{\theta}(\lambda) = \frac{1}{n_S n_T}\sum_{i=1}^{n_S}\sum_{k=1}^{n_T}\psi\left(b_{Si}(\lambda), b_{Tk}(\lambda)\right), \text{ where } \psi(x,y) = \begin{cases} 1 : y > x \\ \frac{1}{2} : y = x \\ 0 : y < x \end{cases}$$

If there are no ties, $\hat{\theta}(\lambda)$ is the proportion of all the possible comparisons of a patient on treatment with a patient on standard in which the former was observed to have a larger net benefit. The estimate of the variance of $\hat{\theta}(\lambda)$, denoted by $s^2_{\theta}(\lambda)$, is given by:

$$s^2_{\theta}(\lambda) = \frac{1}{n_S(n_S-1)}\sum_{i=1}^{n_S}\left[\frac{1}{n_T}\sum_{k=1}^{n_T}\psi\left(b_{Si}(\lambda), b_{Tk}(\lambda)\right) - \hat{\theta}(\lambda)\right]^2 + \frac{1}{n_T(n_T-1)}\sum_{k=1}^{n_T}\left[\frac{1}{n_S}\sum_{i=1}^{n_S}\psi\left(b_{Si}(\lambda), b_{Tk}(\lambda)\right) - \hat{\theta}(\lambda)\right]^2$$

The 100(1 - α)% confidence interval, defined as $\hat{\theta}(\lambda) \pm s_{\theta}(\lambda) z_{1-\alpha/2}$, can be used to express the uncertainty regarding $\hat{\theta}(\theta)$, where $Z_{1-\alpha/2}$ is the 100(1 - α/2)th percentile of the standard normal distribution. We have made the assumption that in large samples $\hat{\theta}(\lambda)$ will be normally distributed.The value $\lambda_{0.5}$, defined at $\theta(\lambda_{0.5}) = 0.5$, provides a non-parametric estimator of the ICER, since it is that value of λ for which treatment and standard are equally cost-effective in the sense that $\Pr[b_{Tk}(\lambda_{0.5}) > b_{Si}(\lambda_{0.5})] = 0.5$. The quantities λ_L and λ_U, defined as $\hat{\theta}(\lambda_L) + s_{\theta}(\lambda_L)Z_{1-\alpha/2} = 0.5$ and $\hat{\theta}(\lambda_U) - s_{\theta}(\lambda_U)Z_{1-\alpha/2} = 0.5$, respectively, are the corresponding lower and upper non-parametric confidence intervals for the ICER.

The hypothesis H_0: θ (λ) = 0.5 versus H_1: θ (λ) > 0.5 can be tested at the level α by rejecting H_0 if $\frac{\hat{\theta}(\lambda)}{s_{\theta}(\lambda)} \geqslant z_{1-\alpha}$. Rejecting H_0 leads to the conclusion that the data provide evidence that treatment is cost-effective.

Suppose there is an interest in comparing θ (λ) between patient subgroups, say between males and females. To achieve this, let $\hat{\theta}_M(\lambda)$ and $\hat{\theta}_F(\lambda)$ be the estimator of θ (λ) for males and females, respectively, with corresponding estimated variances, $s^2_{M\theta}(\lambda)$ and $s^2_{F\theta}(\lambda)$. Then the hypothesis that θ (λ) is the same for males and females can be rejected, at the α-level, 2-sided, if

$$\frac{\left|\hat{\theta}_M(\lambda) - \hat{\theta}_F(\lambda)\right|}{\sqrt{s^2_{M\theta}(\lambda) + s^2_{F\theta}(\lambda)}} \geqslant z_{1-\frac{\alpha}{2}}.$$

There is an important distinction to be made between A(λ) and $\hat{\theta}(\lambda)$. As sample size increases, A(λ) approaches 1 if $\Delta_e\lambda - \Delta_c > 0$, reflecting the certainty that treatment is cost-effective. If $\Delta_e\lambda - \Delta_c \leq 0$, A(λ) approaches 0, reflecting the certainty that treatment is not cost-effective. The quantity $\hat{\theta}(\lambda)$, being independent of sample size, approaches θ(λ) regardless of $\Delta_e\lambda - \Delta_c$. The certainty with which θ(λ) is estimated is reflected in $s_{\theta}(\lambda)$ which is a decreasing function of the sample size.

Table 1: Sample sizes and parameter estimates for prostate example

	n_j	average effectiveness	average cost	$\hat{\sigma}_j^2/n_j$	$\hat{\omega}_j^2/n_j$	$\hat{\rho}_j\hat{\sigma}_j\hat{\omega}_j/n_j$
Standard	53	28.1	29039	16.4	7,872,681	2,876
Treatment	61	40.9	27322	24.1	6,466,351	2,771

Example

In a trial of symptomatic hormone resistant prostate cancer [11,12], 161 patients were randomized between prednisone alone (S) and prednisone plus mitoxantrone (T). Although there was no statistically significant difference in survival, there was better palliation with T. Cost data, including hospital admissions, outpatient visits, investigations, therapies and palliative care, were collected retrospectively on the 114 patients from the three largest centres. Survival was quality-adjusted using the EORTC quality of life questionnaire QLQ-C30. All patients were followed until death. The sample means and the sample variance-covariance information can be found in Table 1. Cost is given in Canadian dollars (CAD$) and effectiveness in quality-adjusted life-weeks (QALW).

A plot of $\hat{\theta}(\lambda)$, complete with 90% confidence intervals can be found in Figure 1. In this example all three plots have positive slope over the range of λ shown, although this would not be true for all examples. For example, by definition $\theta(0) = Pr(c_{Tk} - c_{Sj} < 0)$ and $\theta(\lambda)$ approaches $Pr(e_{Tk} - e_{Sj} > 0)$ as λ becomes arbitrarily large, for any k and j. Therefore, any example in which $Pr(c_{Tk} - c_{Sj} < 0) > Pr(e_{Tk} - e_{Sj} > 0)$ will have negative slope for some interval of positive λ. In the prostate example, since $\hat{\theta}(\lambda)$ crosses the 0.5 horizontal at λ = -50, for any value greater than -50, the estimate of $\hat{\theta}(\lambda)$ is greater than 0.5. The value -50 is a non-parametric estimate of the ICER, since for that value of λ, treatment and standard are equally cost-effective, in that the probability that a patient on treatment has the same net benefit as a patient on standard is 0.5. This compares to the parametric estimate of -134.

Since the lower bound crosses the 0.5 horizontal at 334, for any value greater than 334 the hypothesis $\theta(\lambda) < 0.5$ can be rejected at the 5% level of significance. Thus, a non-parametric upper bound of the ICER is 334. This compares to 378 using Fieller's theorem [13,14]. For this example a health policy maker can interpret the results as follows: for any positive λ, the estimated probability that treatment is cost-effective is greater than 50%; and for any λ greater than 334 per QALW, the probability that treatment is cost-effective is statistically significantly greater than 50%.

Discussion

As an alternative to the acceptability curve, the quantity $\theta(\lambda)$ is proposed as a definition for the probability that treatment is cost-effective. One advantage is that it is not sample size dependent, *i.e.* it is a population parameter. Another is that it has an appropriate interpretation, namely, it is the proportion of patients that realize a larger net benefit if given treatment rather than standard. The acceptability curve does not provide this, although the language often used regarding it, implies that it does. The use of $\theta(\lambda)$ should be helpful to policy makers, since it does not confuse the magnitude of the benefit with the certainty of its estimate.

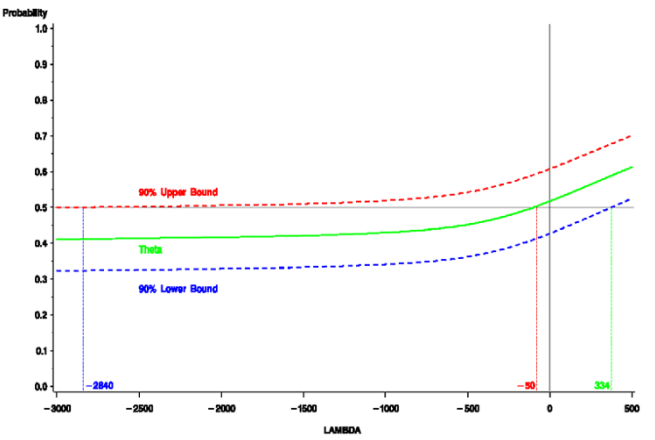

Figure 1
Estimate and confidence limits for θ as a function of λ.

Analysis regarding the quantity $\theta(\lambda)$ is not proposed as an alternative to traditional cost-effectiveness analysis

for allocating health care resources. When allocating a fixed amount of resources to one of two new treatments, the proportion of patients receiving an increase in net benefit would be maximized by choosing the treatment with the larger $\theta(\lambda)$ to Δ_c ratio. However, this would not maximize net benefit, since the ratio may be larger only because the between-patient variability in cost and effectiveness is smaller, resulting in a larger $\theta(\lambda)$.

Non-parametric methods can be used to estimate $\theta(\lambda)$ and its variance. This provides the appropriate confidence intervals and test of hypothesis. In addition, non-parametric estimates of the ICER and its confidence intervals can be determined. This is of particular importance in the presence of highly skewed cost data.

Competing interests
None declared.

Acknowledgements
The author wishes to acknowledge the reviewers whose comments led to a much improved manuscript and to thank Gary Foster for help with the figure.

References
1. van Hout BA, Al MJ, Gordon GS, Rutten EFH: **Cost, effects and C/E-ratios alongside a clinical trial.** *Health Economics* 1994, **3**:309-319
2. Brigg A, Fenn P: **Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane.** *Health Economics* 1998, **7**:723-740
3. Briggs AH: **A Bayesian approach to stochastic cost-effectiveness analysis.** *Health Economics* 1999, **8**:257-261
4. Raikou M, Gray , Briggs , Stevens A, Cull C, McGuire A, Fenn P, Stratton I, Holman R, Turner R: **Cost effectiveness analysis of improved blood pressure control in hypertensive patients with type 2 diabetes: UKPDS 40.** *British Medical Journal* 1998, **317**:720-726
5. Phelps CE, Mushlin AI: **On the (near) equivalence of cost-effectiveness and cost-benefit analysis.** *International Journal of Technology Assessment in Health Care* 1991, **7**:12-21
6. Ament A, Baltussen R: **The interpretation of results of economic evaluation: explicating the value of health.** *Health Economics* 1997, **6**:625-635
7. Stinnett AA, Mallahy J: **Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analsyis.** *Medical Decision Making* 1998, **18 Suppl.**:S68-S80
8. Tambour M, Zethraeus , Johannesson M: **A note on confidence intervals in cost-effectiveness analysis.** *International Journal of Technology Assessment* 1998, **14**:467-471
9. Willan AR, Lin DY: **Incremental net benefit in randomized clinical trials.** *Statistics in Medicine* 2001, **20**:1563-1574
10. DeLong ER, DeLong DM, Clake-Pearson DL: **Comparing the area under two or more correlated receiver operating characteristic curves: a nonparametric approach.** *Biometrics* 1988, **44**:837-845
11. Tannock IF, Osoba D, Stockler MR, Ernst DS, Neville AJ, Moore MJ, Armitage GR, Wilson JJ, Venner PM, Coppin CM, Murphy KC: **Chemotherapy with mitoxantrone plus prednisone or prednisone alone for symptomatic hormone-resistant prostate cancer: a Canadian trial with palliative endpoints.** *Journal of Clinical Oncology* 1996, **14**:1756-1764
12. Bloomfield DJ, Krahn MD, Neogi T, Panzarella T, Warde P, Willan AR, Ernst S, Moore MJ, Neville A, Tannock IF: **Economic evaluation of chemotherapy with mitoxantrone plus prednisone for symptomatic hormone-resistant prostate cancer: based on a Canadian trial with palliative endpoints.** *Journal of Clinical Oncology* 1998, **16**:2272-2279
13. Willan AR, O'Brien BJ: **Confidence intervals for cost-effectiveness ratios: An application of Fieller's theorem.** *Health Economics* 1996, **5**:297-305
14. Chaudhary MA, Stearns SC: **Confidence intervals for cost-effectiveness ratios: an example from a randomized trial.** *Statistics in Medicine* 1996, **15**:1447-1458