Available online at www.sciencedirect.com
ScienceDirect
Journal of Health Economics 27 (2008) 476–495
JOURNAL OF HEALTH ECONOMICS
www.elsevier.com/locate/econbase

Measuring and illustrating statistical evidence in a cost-effectiveness analysis

Jeffrey S. Hoch [a,b,*], Jeffrey D. Blume [c]

[a] *Centre for Research on Inner City Health, The Keenan Research Centre in the Li Ka Shing Knowledge Institute of St. Michael's Hospital, Toronto, Ontario, Canada*
[b] *Department of Health Policy, Management and Evaluation, University of Toronto, Toronto, Ontario, Canada*
[c] *Center for Statistical Sciences, Brown University, Providence, RI 02912, USA*

Received 11 March 2005; received in revised form 2 July 2007; accepted 18 July 2007
Available online 7 January 2008

Abstract

Recently, there has been much interest in using the cost-effectiveness acceptability curve (CEAC) to measure the statistical evidence of cost-effectiveness. The CEAC has two well established but fundamentally different interpretations: one frequentist and one Bayesian. As an alternative, we suggest characterizing the statistical evidence about cost-effectiveness using the likelihood function (the key element of both approaches). Its interpretation is neither dependent on the sample space nor on the prior distribution. Moreover, the probability of observing misleading evidence is low and controllable, so this approach is justifiable in the traditional sense of frequentist long-run behaviour. We propose a new graphic for displaying the evidence about cost-effectiveness and explore the strengths of likelihood methods using data from an economic evaluation of a Program in Assertive Community Treatment (PACT).

JEL classification: I18

Keywords: Cost-effectiveness analysis; Net benefit regression framework; Likelihood methods

1. Introduction

There is growing interest in improving the quality of trial-based economic evaluations (see, for example Chapter 6 in Drummond and McGuire (2001); Chapter 8 in Drummond et al. (2005) and Ramsey et al. (2005)). The International Society for Pharmacoeconomics and Outcomes Research's Task Force on Good Research Practices (RCT-CEA Task Force) recently concluded that "conducting high quality economic analyses alongside clinical studies is desirable because they provide timely information with high internal validity" (Ramsey et al., 2005). This conclusion was framed in the context of an "increasing demand [for] evidence of economic value for health-care interventions" (Ramsey et al., 2005). An accompanying editorial stressed the need to assess evidence in the data, using the word "evidence" more than 10 times in the space of less than two pages (Backhouse, 2005). So it would

* Corresponding author at: Centre for Research on Inner City Health, 30 Bond Street, Toronto, Ontario M5B 1W8, Canada. Tel.: +1 416 864 6060x2194; fax: +1 416 864 5485.
E-mail address: jeffrey.hoch@utoronto.ca (J.S. Hoch).

0167-6296/$ – see front matter
doi:10.1016/j.jhealeco.2007.07.002

seem desirable to provide consumers of trial-based economic evaluations with an assessment of the strength of the evidence provided by the data that a new treatment or intervention is cost-effective. Ideally, this assessment would describe only "what the data say" about the cost-effectiveness of the intervention. This paper describes how Likelihood methodology can meet this need; we explain its advantages and limitations, and we contrast it with more common methods.

Depending on the methodology chosen, statistical analysis of an intervention's cost-effectiveness can answer one of the following three questions: "Now that we have seen the data,

(1) Should we *act* as if the intervention is cost-effective?;
(2) Do we *believe* that the intervention is cost-effective?; or
(3) What is the *evidence* that the intervention is cost-effective?"

The answer to question (3) is derived entirely from the data, whereas the answers to the first two questions depend on more that just the data themselves (e.g., the impending action's probable gains and losses, or what was initially believed to be true) (Berkson, 1942, 2003; Blume, 2002; Cox, 1958; Edwards, 1970; Fisher, 1959; Neyman, 1950; Royall, 1997). As a result, the three questions and their distinct answers must be considered separately. The tools and principles of frequentist Decision theory (i.e., Neyman–Pearson–Wald hypothesis testing) and Bayesian methods can be used to answer the first two questions[1]; however, many have argued that problems arise when these frequentist or Bayesian methods are applied to question (3) (Blume and Peipert, 2003; Goodman, 1993; Goodman and Royall, 1988; Royall, 1997).[2]

Depending on its interpretation, the cost-effectiveness acceptability curve (CEAC) (Briggs and Fenn, 1998; Fenwick et al., 2001, 2004; Lothgren and Zethraeus, 2000; van Hout et al., 1994) is often used to address either question (1) or question (2), but it is insufficient for answering question (3). A separate theory or paradigm dedicated to measuring statistical evidence is needed. Many have argued that such a paradigm must focus on the likelihood function alone (Barnard, 1949; Birnbaum, 1962; Edwards, 1970; Fisher, 1922; Royall, 1997). While the likelihood function figures prominently in both frequentist and Bayesian methodologies, it is neither the focus nor the endpoint of either methodology. In contrast, the likelihood paradigm, with the likelihood function as its centrepiece, embodies a theory of statistical evidence (Blume, 2002; Edwards, 1972, 1992; Royall, 2000a,b). Under the likelihood paradigm, the statistical evidence is characterized by the likelihood function and the strength of the evidence for one hypothesis over another is measured by their likelihood ratio. Thus, question (3) can be answered by using the likelihood function to characterize the statistical evidence about cost-effectiveness.[3] Empirical examples of the likelihood paradigm can be found in the scientific literature (Blume, 2002; Mellen and Royall, 1997).

This paper considers how characterizing the strength of evidence in the data through the use of likelihood methods could complement current techniques for analyzing trial-based cost-effectiveness data. As a framework for our likelihood analysis, this paper begins by briefly tracing some of the major advances in statistical cost-effectiveness analysis. Next, we examine the different interpretations of the CEAC and discuss how they address questions (1) and (2). Subsequently, we introduce the likelihood paradigm, describe why it represents a compromise between the frequentist and Bayesian approaches, discuss its frequentist behaviour (e.g., how often strong misleading evidence is observed), and show how to measure evidence of cost-effectiveness in the net benefit regression framework (Hoch et al., 2002). In addition, we propose a new graphic that displays the statistical evidence about cost-effectiveness and apply it to data from an economic evaluation of a program of assertive community treatment (ACT) for homeless people with

[1] Significance testing, as Fisher proposed it, was intended to address question (3). However, as both Goodman (1993) and Royall (1997) point out, today's version of significance testing is actually a combination of hypothesis testing concepts and Fisher's original significance testing ideas, and the resulting 'methodology' and application is much closer to hypothesis testing. So, for the purposes of this paper, we classify such methods along with those of hypothesis testing (although we are careful to make the necessary distinctions through the paper). Note also that neither incarnation of significance testing is adequate for answering the third question (Royall, 1997), and that we do not in this context delineate between Bayesian decision theory which addresses question (1) and Bayesian belief theory which addresses question (2).

[2] Notable examples are the controversies surrounding adjustment for multiple comparisons and multiple looks at data, as well as the proper use and interpretation of *p*-values (Royall, 1986; Goodman and Royall, 1988; Goodman, 1998; Blume and Peipert, 2003).

[3] We do not exclude the possibility that competing theories of statistical evidence will be (or have been!) proposed and developed. Rather, our contention is that neither hypothesis testing nor Bayesian methods represent competing theories of statistical evidence, largely because their derivations and intended uses were for different purposes.