DEMISSIE ALEMAYEHU

Columbia University and Pfizer Inc, US, alem@stat.columbia.edu

METHODOLOGICAL CONSIDERATIONS IN COST-EFFECTIVENESS ANALYSIS

Abstract:

In decision making regarding optimal resource allocation to safeguard public health, policymakers and healthcare providers rely on the availability and reliability of data about the relative costs and benefits of competing treatment options. One such an approach is based on cost effectiveness analysis (CEA) which is intended to be used as a combined metric of both the costs and health outcomes of alternative intervention strategies. However, the usual measures used in CEA are not readily analyzable based on standard statistical paradigms for inference. Further, reliable data may not always be available to estimate relevant parameters. Accordingly, it is essential to employ nonstandard procedures to compensate for information gaps and to address inferential difficulties. In this paper, we outline the issues associated with some of the commonly used techniques, with particular emphasis on the so-called network meta-analysis and indirect comparisons. Additional reference is made to the complexities introduced when data are used from observational studies. It is concluded that effective use of CEA in healthcare policy presupposes a careful appreciation of the underlying issues, and implementation of robust remedial measures to mitigate their impacts.

Keywords:

Cost-Effectiveness Analysis, Network Meta-analysis, Indirect Comparisons, Health Economics

JEL Classification: C18

1. Introduction

With the increased focus on the ever-rising cost of healthcare utilization, policymakers need reliable data to make informed decisions about optimizing resource allocation for competing priorities. One such an approach is based on cost effectiveness analysis (CEA) which is intended to be used as a combined metric of both the costs and health outcomes of alternative intervention strategies (Gold, Siegel, Russell and Weinstein, 1996). Typically, CEA presents results in terms of cost per health outcome measure. Common examples include cost per untoward event prevented and cost per life year gained. In general, the goal of CEA is to compare alternative treatment options that have been studied using a common health outcome measure (Saha, Hoerger, Pignone, Teutsch et al., 2001; Weinstein and Stason, 1977).

In contrast to other related metrics, including cost benefit analysis, which assigns monetary values ascribable to a treatment option, and cost utility analysis, which incorporates quality of life measures (e.g., quality-adjusted life years saved), CEA has several desirable features that make it appealing as the metric of choice in healthcare research. Notably, because it is relatively easy to interpret, it is fairly commonly used use in decision making by policy makers and healthcare providers.

While there is a general recognition of the importance of CEA as a fundamental research tool in medical research and health technology assessment, there is also the potential for its misuse on account of inherent methodological issues as well as the interpretation of the results (Doubilet, Weinstein and McNeil, 1986). Accordingly, there have been ongoing efforts to strengthen the evidence generated using CEA through enhancement of the analytical methods, the reliability of the data used as input, and the reporting of the results (see, e.g., Weinstein, Siegel, Gold, et al., 1996; Drummond, Richardson, O'Brien, Levine et al., 1997; O'Brien, Heyland, Richardson, Levine et al. 1997).

One limitation of CEA emanates from the requirement that outcome measures used to assess effectiveness be similar among the different treatment options that are under study. In practice, this is often handled by using broadly general endpoints, such as number of lives saved. However, this may not always be feasible, since it may not be obvious how to transform the endpoints used in the original trials to such general readily interpretable common measures.

A second and more pervasive issue concerns the availability of comparative data on the effectiveness of the treatment options of interest. This is particularly the case when interest lies in performing CEA involving several competing treatment options. When data is available about the relative efficacy of alternative treatment options from randomized controlled trials (RCTs), which are accepted as the gold standard for evidence-based medicine, such data are primarily used in CEA. However, head-to-head comparative data from RCTs may not always be available for treatments that need to be

studied. In comparative effectiveness research the issue is often handled using the so-called indirect comparison procedures (Bucher, Guyatt, Griffith and Walter, 1997; Lu and Ades, 2004). Another potential approach to fill the data gap is to use pertinent efficacy data from non-randomized studies. Despite the appeal of these procedures to serve as viable options in the absence of direct comparative evidence from RCTs, they also have certain inherent shortcomings that require careful scrutiny.

In this paper, we outline the issues associated with some of the commonly used techniques, with particular emphasis on network meta-analysis and use of data from observational studies in relation to CEA. The paper is organized as follows. In Section 2 we expose relevant aspects of CEA and the associated issues with performing inference. In Section 3 we introduce common indirect comparison procedures and underlying assumptions, and provide steps to be taken to assess the validity of underlying assumptions. In Section 4, we underscore the issue of controlling bias when observational data are used to generate effectiveness in CEA. In Section 5 we conclude with a few cautionary remarks.

2. Parameter Estimation and Inference

When there are no competing choices, interest may lie in the relatively simple problem of evaluating the cost effectiveness of a given treatment strategy, say A. Suppose the associated net cost and effect of treatment with A are denoted by C_A and E_A, respectively. The goal then is the estimation of the average cost effectiveness (ACE), defined as:

$$ACE \quad = \frac{C_A}{E_A}$$

As long as reliable input values are available for the numerator and denominator in the above, the computation of ACE is fairly straight forward. However, formal statistical inference may be intricate because of the known complexities that arise when dealing with the ratio of two random variables.

A common, but even more complicated, situation involves the determination of the so-called incremental cost-effectiveness ratio (ICER) which is used when one needs to choose from among competing options. To fix ideas, in the simple case when there are only two treatment options A and B, let the associated costs and effectiveness measures for B be denoted by C_B and E_B. Then the ICER is given by:

$$ICER \quad = \frac{C_A - C_B}{E_A - E_B}$$

Despite the common use of ICER in the pharmacoeconomic literature and in decision making involving healthcare utilization, there are several inferential and other methodological issues associated with the procedure that are nontrivial and require careful and close scrutiny.

From a statistical inference perspective, much has been written about the difficulty in the construction of confidence intervals for the underlying ICER (see, e.g., Laska, Meisner and Siegel, 1997). The traditional approaches, such as the Fieller-based confidence intervals, have known limitations (Wakker and Klassen, 1995). When patient level data are available, bootstrap approaches can be used to provide reliable confidence limits (Briggs, Wonderling and Mooney, 1997). However, since most calculations are performed on the bases of summary statistics for the numerator and denominator, re-sampling techniques are of little or no value. A more detailed discussion of alternative approaches may be found in Polsky, Glick, Willke and Schulman (1997).

In the following, we restrict attention the problem of obtaining reliable estimates for the denominator, especially when there are no data from randomized controlled trials that compared A and B in a head-to-head manner.

3. Managing Pitfalls with Data from Indirect Comparisons

Suppose there is interest in performing CEA pertaining to treatments A and C, but there is inadequate information from RCTs that studied the two treatments in a head-to-head fashion. Assume that A and B were each studied in separate RCTs relative to a common comparator B, with corresponding estimated effects, $d_{AB} = E_A - E_B$, and $d_{CB} = E_C - E_B$. In most applications, d_{AB} and d_{CB} may be functions of the usual efficacy measures, such as odds ratios, risk differences or mean differences. Bucher, Guyatt, Griffith and Walter (1997) proposed estimating the effect of A relative to C indirectly by

$$d_{AC} = d_{AB} - d_{CB}$$

A major advantage of the approach is that it preserves, at least partially, the benefits of randomization, compared to a naïve approach that uses the direct estimates of the effects of the individual treatments. However, the approach not only has the known limitations of standard meta-analytic techniques (Egger, Smith and Altman, 2001), but also presupposes crucial assumptions that are not always readily verifiable. One critical assumption is that the trials contributing evidence are similar with respect to relevant attributes. Further, in situations where it may be desirable to synthesize information from both direct and indirect sources, i.e., when doing mixed treatment comparison (MTC), it is also essential to assume there is consistency in the evidence generated from the direct and indirect sources (Lumley, 2004).

The assumption of similarity or exchangeability is not testable, since its underlying requirement is that the same effect size would have been obtained had each trial compared the treatments studied in the other trial (Alemayehu, 2011). Obviously, this would be the case if the trials were identical in the study methods used, the environment under which they were conducted, and the characteristics of the patient populations they enrolled, among other factors.

While formal measures of assessing the exchangeability assumption are not available, there are qualitative and quantitative approaches that need to be used to add a level of credibility to the results. Some factors that should be inspected for comparability across trials include study level attributes, such as study design features, duration of treatment, efficacy measures and dosing schedules. External factors that could influence study outcome and that should be looked at include health care policy, study location, and when the trials were conducted. Patient characteristics, such as demography, disease state and medical history, need to be compared for any disparity. If the qualitative assessment indicates the presence of imbalance in any factor that has the potential to introduce bias, appropriate analytical measures should be taken to address the issue. In the usual case, when only aggregate data is available, one may perform meta-regression, with the understanding that one needs to guard against the so-called ecological fallacy and the limitations imposed by the usual small number of studies (Berlin, Santanna, Schmid, Szczech et al., 2002). It may also be desirable to examine the consistency of effect estimates for the reference group across trials, as well as any heterogeneity in the original analyses.

For RCTs that may differ only with respect to observed characteristics of enrolled patients, an approach that has been suggested is to identify similar subgroups in each trial mirroring the characteristics of patients in the other, and performing a test of homogeneity (Alemayehu, 2011). Identification of subgroups may be carried using well-established procedures, including the modified minimum volume ellipsoid (MVE) approach. Once similar subgroups are identified, a test of subgroup-by-treatment interaction may be performed, incorporating a term for the subgroup in the analytical model. If the analysis suggests that there is heterogeneity, then the assumption of exchangeability cannot be justified, and appropriate remedial measures should be taken.

When the estimation of comparative effectiveness involves synthesizing data from direct and indirect sources, it is essential to assess the consistency of results from the two sources. This is important because inconsistency is likely to lead to biased results, thereby affecting the value of the ICER. There are methods that have been proposed to assess consistency of evidence. They typically involve performing inference based on a suitably defined measure of discrepancy between the direct and indirect evidence (see, e.g., Bucher, Guyatt, Griffith and Walter (1997); Dias, Welton, Caldwell and Ades, 2010; Dias, Welton, Sutton, Caldwell et al., 2013). When results are inconsistent, it is important

to investigate possible causes of discrepancy, and take appropriate measures before using the estimates in the computation of ICER.

Obviously, neither the qualitative nor the quantitative approaches discussed above are fully satisfactory in terms of giving unequivocal evidence for or against the tenability of the exchangeability or consistency assumptions. In this regard, simulations may play an important role (Caro and Ishak 2010). In a typical simulation exercise, suitable predictive models are constructed based on the characteristics of one trial, and then used to simulate outcomes for patients in the other trials. Depending on the complexity of the data and the trials, either traditional models or nontraditional models may be developed. The common traditional models include logistic regression, linear models, and Cox-proportional hazards models. For more complex situations, especially when traditional tools are not amenable, algorithmic models, such as random forests, support vector machines, or neural network, may be applied (Hastie, Tibshirani and Friedman, 2001).

4. Efficacy Based on Observational Studies

It has long been recognized that although RCTs have internal validity, they lack external validity, in that results do not reflect real-world conditions. This is partly because RCTs have strict protocol criteria that exclude patients for operational or ethical reasons, or that are not consistent with real world use of the treatments. Further, there may be no data available data from RCTs to address a given situation. It may therefore be essential to use data from observational studies to fill the data gaps.

A major issue with observational studies is the potential for bias introduced by both measured and unmeasured confounders. Uncritical use of comparative efficacy data from such studies would inevitably introduce bias in CEA.

For measured confounders, there are commonly used techniques to control for their effects, including the usual analysis of covariance or other techniques, such as propensity score analysis. Although approaches have been proposed to handle unmeasured confounders, e.g., instrumental variables, there is no universally accepted procedure that has been shown to give reliable results. Confounding by indication, which is common in drug safety studies, arises when the indication is also a risk factor for the outcome. In this case, there is always the risk of residual confounding (Salas, Hofman and Strieker, 1999), with the possibility of residual bias even when all attempts are made to control for observed confounders (Alemayehu, Alvir, Jones and Willke, 2011).

In recent studies, observational studies have been shown to be dependent on the choice of database as well as analytical and design strategies. In one study, for example, Madigan, Ryan, Schuemie, Stang et al. (2013) noted that depending on the observational database selected, analytical results may vary from one extreme to another. In a related study, Madigan, Stang, Berlin, Schuemie et al. (2014) also showed

that the choice of study design and analytical techniques could also dramatically influence the study conclusions, all other factors remaining constant.

Thus, while observational data may play a critical role in CEA, as a source of evidence complementary to RCTs, caution should be exercised in the use of such data in the calculation of ICERs and other similar metrics. In addition to the measures summarized above in the context of controlling for bias, best practices should be adopted, particularly those intended to strengthen the values of data from secondary sources (Cox, Marin, Van Staa, Garbe et al., 2009).

5. Conclusion

Cost-effectiveness analysis is important in guiding decision about allocation of scarce resources to improve healthcare. The reliability of CEA is heavily dependent on the degree to which the input variables are accurately determined. Evidently, a key variable in the computation of ICER is the difference in effectiveness of the treatments under study. For the purpose of assessing relative treatment efficacy, the standard sources of data, RCTS, may fall short, especially in circumstances when such data may not be available to do the desired analyses. In other cases, the data from RCTs may not be adequate, since the conditions under which such trials are conducted may not be reflective of the real-world where the drugs are used. Accordingly, it may be essential to resort to indirect ways of obtaining the desired information. In this paper we considered two potential approaches that are commonly used to address such evidentiary gaps, and discussed the need to understand and mitigate the limitations of the procedures. It is acknowledged that some of the issues raised are complex and may not have definitive solutions. However, in light of the potential impacts of the results of CEA on public health, it is stressed that researchers and policy makers should have heightened awareness of the underlying issues.

Recently, there has been a growing interest in comparative effectiveness research (CER) in the United States (US). While the methodological issues discussed in this paper have been thoroughly studied in the context of CER, the literature pertaining to the impacts on CEA is not extensive. This is in part because of the fact that policymakers and most healthcare providers in the US, unlike those in most other industrialized countries, are reluctant to utilize CEA in decision making (Bryan, Sofaer, Siegelberg and Gold, 2009). It is, however, hoped that researchers involved in CEA will continue to leverage the tremendous methodological activities that are ongoing in CER to enhance the evidentiary value of CEA results to promote healthcare delivery and utilization.

BIBLIOGRAPHY

Alemayehu, D. (2011) Assessing exchangeability in indirect and mixed treatment comparisons. *Comparative Effectiveness Research*. 1. p.51 – 55.

Alemayehu, D., Alvir, J., Jones, B. and Willke, R. (2011) Statistical issues with the analysis of non-randomized studies in comparative effectiveness research. *J Manag Care Pharm.* 17(9-a). p.S22-S26.

Berlin, J.A., Santanna, J., Schmid, C.H., Szczech, L.A. et al. (2002 Individual patient-versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat. Med.* 21. p.371–387.

Briggs, A.H., Wonderling, D.E. and Mooney, C.Z. (1997) Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. *Health Economics. 6.* p.*327*–40.

Bryan, S., Sofaer, S., Siegelberg, T. and Gold, M. (2009) Has the time come for cost-effectiveness analysis in U.S. health care? *Health Economics, Policy and Law.* 4. p.425-443.

Bucher, H.C., Guyatt, G.H., Griffith, L.E. and Walter, S.D. (1997) The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol.* 50(6). p.683–691.

Caro, J.J. and Ishak, K.J. (2010) No head-to-head trial? Simulate the missing arms. *Pharmacoeconomics.* 28. p. 957–967.

Cox, E., Marin, B.C., Van Staa, T., Garbe, E. et al. (2009) Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report – part II. *Value Health.* 12. p. 1053–1061.

Dias, S., Welton, N.J., Caldwell, D.M. and Ades, A.E. (2010) Checking consistency in mixed treatment comparison meta-analysis. *Stat. Med.* 29(7-8). p. 932–944.

Dias, S., Welton, N.J,, Sutton, A.J., Caldwell, D.M. et al. (2013) Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med. Decis. Making.* 33. p. 641–656.

Doubilet, P., Weinstein, M.C. and McNeil, B.J. (1986) Use and misuse of the term "cost effective" in medicine. *N Engl J Med.* 314(4). p.253-256.

Drummond, M.F., Richardson, W.S., O'Brien, B.J., Levine, M. et al. (1997) Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA*. 277(19). p. 1552-1557.

Egger, M., Smith, G.D. and Altman, D.G. (eds.) . (2001) *Systematic Reviews in Health Care*: *Meta-analysis in Context*. 2nd Ed. London: BMJ Books.

Gold, M.R., Siegel, J.E., Russell, L.B. and Weinstein, M.C. (1996) *Estimating Costs in Cost-Effectiveness Analysis. Cost-Effectiveness in Health and Medicine.* New York: Oxford University Press.

Hasselblad, V. (1998) Meta-analysis of multitreatment studies. *Med Decis Making*. 18(1). p.37–43.

Hastie, T., Tibshirani, R., and Friedman, J. (2001) *The Elements of Machine Learning: Data Mining, Inference and Prediction*. New York: Springer.

Laska, E.M., Meisner, M. and Siegel, C. (1997) Statistical inference for cost-effectiveness ratios. *Health Econ*. 6(3). p.229-42.

Lu, G. and Ades, A.E. (2004) Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 23(20). p.3105–3124.

Lumley, T. (2002) Network meta-analysis for indirect treatment comparisons. *Stat Med*. 21(16). p.2313–2324.

O'Brien, B.J., Heyland, D., Richardson, W.S., Levine, M. et al. (1997) Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. What are the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. *JAMA*. 277(22). p.1802-1806.

Madigan, D., Ryan, P.B., Schuemie, M., Stang, P.E. et al. (2013) Evaluating the impact of database heterogeneity on observational study results. *Amer. J. Epidemiol*. 178(4). p.645-651.

Madigan, D., Stang, P.E., Berlin, J.A., Schuemie, M. et al. (2014) A systematic statistical approach to evaluating evidence from observational studies. *Annual Review of Statistics and Its Application*. 1(1). p.11-39.

Polsky, D., Glick, H.A., Willke, R. and Schulman, K. (1997) Confidence intervals for cost-effectiveness ratios: a comparison of four methods. *Health Economics*. 6. p. 243–252.

Saha, S., Hoerger, T.J., Pignone, M.P., Teutsch, S.M. et al. (2001) The art and science of incorporating cost-effectiveness into evidence-based recommendations for clinical preventive services. *Am J Preventive Med.* 20(supplement 3). p.36-44.

Salas, M., Hofman, A., and Strieker, B.H.C. (1999) Confounding by indication: An example of variation in the use of epidemiologic terminology. *Am J Epidemiol.* 149(11). p. 981-983.

Wakker, P. and Klassen, M.P. (1995) Confidence intervals for cost/effectiveness ratios. *Health Economics. 4.* p*.73*–81.

Weinstein, M.C. and Stason, W.B. (1977) Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med.* 296913). p.716-721.

Weinstein, M.C., Siegel, J.E., Gold, M.R. et al. (1996) Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA.* 276(15). p.1253-1258.

Wells, G.A., Sultan, S.A., Chen, L., Khan. M. et al. (2009) *Indirect Evidence*: *Indirect Treatment Comparisons in Meta-Analysis*. Ottawa: Canadian Agency for Drugs and Technologies in Health.