Abstract Title Page

Not included in page count.

Title: Understanding the Uncertainty of an Effectiveness-cost Ratio in Educational Resource Allocation: A Bayesian Approach

Authors and Affiliations:
Yilin Pan
Teachers College, Columbia University

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

Despite a wide-ranging support of the message that both effectiveness and cost information should be taken into account for program selection, the methodological standards of conducting cost-effectiveness analysis in education are still under discussion (Levin and Belfield, 2014). One methodological issue in debate is whether it is reasonable and sufficient to compare the alternatives only based on a single, scalar efficiency measure, i.e., the cost-effectiveness ratio estimate derived from the observed sample for each program of interest. The ratio estimate conveys information about what happened once in the specific evaluation settings. However, if the program is replicated (either in the original evaluation settings or at a different site), it is almost impossible to obtain the same cost-effectiveness ratio due to measurement error, time-to-time variability, site-to-site variability, or other factors that contribute to the uncertainty. Therefore, compared to a single cost-effectiveness ratio estimate that tells what happened, more useful information for practitioners would be 1) the best guess for what to anticipate in terms of the trade-off between effectiveness and cost, and 2) the comparatively worst-case and best-case scenarios. The underlying methodological challenge is to identify a probability distribution of an efficiency measure, based on which the expectation, the 2.5th-quantile and the 97.5th-quantile can be calculated to answer these two questions that practitioners are interested in.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

Given the necessity to bridge the gap between what happened and what is likely to happen, this paper aims to explore how to apply Bayesian inference to cost-effectiveness analysis so as to capture the uncertainty of a ratio-type efficiency measure. The first part of the paper summarizes the characteristics of the evaluation data that are commonly available in educational research, discusses the ratio property and proposes different estimators of interest. The second section synthesizes two perceptions of uncertainty in the literature, and reviews the conventional quantitative methods that address the uncertainty of a ratio under each perception. The third part proposes two Bayesian models that differ in the assumption of site-level variability, and demonstrates the estimation, presentation and interpretation of the results using the comparison of two high school dropout prevention programs: New Chance and JOBSTART. The last section summarizes the strengths and limitations of the Bayesian method, and lists some directions for future exploration.

Significance / Novelty of study:

Description of what is missing in previous work and the contribution the study makes.

In the literature of statistics, there are generally two categories of perception towards uncertainty: sampling variation, and incomplete information. The way to perceive uncertainty as sampling variation is derived from the inference framework that the observed dataset is one random sample of the population, and the inference from the sample to the population is based on an imaginary situation in which the sampling process is repeated infinite times. Measures of the sampling error such as standard error and confidence interval, are used to model the uncertainty of the point estimate in terms of estimation precision. Since a ratio estimator does not have a mathematical tractable formula to calculate the variance (Briggs et al., 2002), researchers usually use Delta method or Fieller's theorem to approximate the confidence interval of a ratio along with reporting a single point estimate from the sample, or rely on bootstrapping or Monte Carlo method to generate the sampling distribution of the estimator of interest. Note that sampling variation is not the main source of uncertainty that matters to educational practitioners given the rarity of replicating a program many times in practice. In addition, the applicability of Delta Method, Fieller' Theorem, bootstrapping and Monte Carlo method is highly restricted by the limited sample size, a property commonly observed among the available datasets in educational cost-effectiveness analysis since the unit of analysis is site rather than individual.

Compared to sampling variation, the way to perceive uncertainty as it arises from incomplete information is probably more intuitive: one is uncertain about what happened in the past or what will happen in the future because not all information is obtainable, reliable or certain. Therefore, even though the true efficiency level of a program is a fixed value, what we know about it entails some randomness because of the limited availability of information; and the more information one has, the less uncertainty there is. Under this perception, there are two categories of methods to quantify the uncertainty of the estimation in cost-effectiveness analysis: conventional sensitivity analysis and Bayesian approach. In conventional sensitivity analysis, researchers arbitrarily determine which assumption to manipulate and what possible values to impose on the key assumption, which may generate intentional or unintentional selection bias. The deficiency of sensitivity analysis to fit in a statistical inference framework calls for an approach that entails advantages of both intuitive interpretations and standard statistical inference. Bayesian inference happens to be the one.

Statistical, Measurement, or Econometric Model:

Description of the proposed new methods or novel applications of existing methods.

<u>Estimators</u>

For a program, let ATEj represent the average treatment effect at Site j; ACj be the average cost at Site j; nj be the scale of the program at Site j. The site-level EC ratio (ECRj) of a program is

$$ECR_j = \frac{ATE_j}{AC_j} \tag{2}$$

The weighted EC ratio (W ECR) of a program is

$$WECR = \frac{\sum w_j ATE_j}{\sum w_j AC_j},\ where\, w_j = \frac{n_j}{\sum n_j} \tag{3}$$

The weighted EC ratio is the weighted average treatment effect across all sites divided by the weighted average cost, with the weights proportional to the scale of the sites. This paper will investigate methods to estimate the expectation and the confidence interval for both estimators.

<u>Models</u>

1) Complete pooling model

I first assume that the true values of average treatment effect and average cost at all sites are the same. Let Ej and Cj represent (the linear transformations of) the estimated average treatment effect and average cost for site j. The Bayesian model can be expressed as follows.

$$\begin{bmatrix} E_j \\ C_j \end{bmatrix} \sim MVNormal \left[\begin{pmatrix} u_E \\ u_C \end{pmatrix}, \quad \Delta\Lambda\Delta \right] \qquad \text{[Likelihood]}$$

$$u_E \sim N(0, 10) \qquad \text{[Normal prior for } u_E\text{]}$$

$$u_C \sim N(0, 10) \qquad \text{[Normal prior for } u_C\text{]}$$

$$\Delta \sim \tfrac{1}{\delta} \qquad \text{[Jeffrey's prior for each element of } \Delta\text{]}$$

$$\Lambda \sim LKGCorr(\eta = 1) \qquad \text{[LKJ prior for correlation matrix]}$$

2) Hierarchical model

The assumption that all of the sites come from the same distribution may not be plausible, since all the factors that affect the true value of effectiveness and cost, such as students' SES status, teachers' profiles and school leadership, arguably vary from site to site. Again, let Ej and Cj represent (the linear transformations of) the estimated average treatment effect and average cost for site j. To capture the site-to-site variability, a hierarchical model is expressed as follows.

$$\begin{bmatrix} E_j \\ C_j \end{bmatrix} \sim MVNormal \left[\begin{pmatrix} \theta_{Ej} \\ \theta_{Cj} \end{pmatrix}, \quad \Delta\Lambda_\delta\Delta \right] \qquad \text{[Likelihood]}$$

$$\begin{bmatrix} \theta_{Ej} \\ \theta_{Cj} \end{bmatrix} \sim MVNormal \left[\begin{pmatrix} u_E \\ u_C \end{pmatrix}, \quad \tau\Lambda_\tau\tau \right] \qquad \text{[Multivariate normal prior for } \begin{bmatrix} \theta_{Ej} \\ \theta_{Cj} \end{bmatrix}\text{]}$$

$$\Delta \sim Gamma(2, 1) \qquad \text{[Gamma prior for each element of } \Delta\text{]}$$

$$\Lambda_\delta \sim LKGCorr(\eta = 1) \qquad \text{[LKJ prior for correlation matrix } \Lambda_\delta\text{]}$$

$$u_E \sim N(0, 10) \qquad \text{[Normal hyperprior for } u_E\text{]}$$

$$u_C \sim N(0, 10) \qquad \text{[Normal hyperprior for } u_C\text{]}$$

$$\tau \sim Gamma(2, 1) \qquad \text{[Gamma hyperprior for each element of } \tau\text{]}$$

$$\Lambda_\tau \sim LKGCorr(\eta = 1) \qquad \text{[LKJ hyperprior for correlation matrix } \Lambda_\tau\text{]}$$

Usefulness / Applicability of Method:

Demonstration of the usefulness of the proposed methods using hypothetical or real data.

As a demonstration, I will apply the methods and models to the site-level effectiveness and cost data of two programs that share the objective of increasing high school completion rate: New Chance and JOBSTART. Implemented in 16 sites across the country between 1989 and 1992, New Chance is a residential demonstration project targeting at 16-to 22year-old mothers who had first given birth as teenagers, had dropped out of high school, and were receiving cash welfare assistance (Quint et al., 1997). JOBSTART is a non-residential demonstration program targeting at 17-to 21-year-old, economically disadvantaged dropouts. It was implemented in 13 sites across the country between 1985 to 1988 (Cave et al., 1993). Both programs provided academic tutoring, vocational education, and job assistance to their participants. The impact evaluations (designed as randomized block trials) and cost analyses of both programs were conducted by MDRC (Cave et al., 1993; Fink and Farrell, 1994; Quint et al., 1997). Levin et al. (2012) adjusted both the effect and cost data to increase the comparability of these two programs, and this paper will base on the adjusted data.

Site-level EC ratio: Table 1 reports the mean value and 95% confidence interval of the posterior predictive distribution of the site-level EC ratio for New Chance and JOBSTART, estimated by the complete pooling model and the hierarchical model respectively. With regard to the site-level EC ratio, the mean estimates of the same program are not significantly different across models; the 95% confidence interval estimated by the hierarchical model is slightly larger than that generated by the complete pooling model, given that 1) the site-to-site variability is incorporated into the model; and 2) each parameter is less likely to be estimated precisely as the number of parameters to estimate increases.
<insert Table 1 here>

Weighted EC ratio: Table 2 reports the mean value and 95% confidence interval of the posterior predictive distribution of the weighted EC ratio for New Chance and JOBSTART, estimated by the complete pooling model and the hierarchical model respectively. As it shows, the distribution of the weighted EC ratio is more concentrated than that of the site-level EC ratio. It is consistent with our expectation since the weighting process averages out both effectiveness and cost and tends to eliminate the extreme values. For both New Chance and JOBSTART, the two models also generate dissimilar posterior predictive distributions of the weighted EC ratio, indicating that accounting for the site-level variation makes a difference in the estimation.
<insert Table 2 here>

Program comparison
To visualize the comparison of the two programs, I plot the posterior predictive distributions of the two estimators for both programs together, all generated by the hierarchical model. As shown in Figure 1, for both estimators, JOBSTART has a larger mean value and a larger variance than New Chance; but there is also a small probability that an estimate for JOBSTART is smaller than an estimate for New Chance. It implies that in terms of the best guess to what would happen in efficiency, JOBSTART is much better than New Chance; it is very unlikely to happen that JOBSTART performs worse than New Chance, although the worst-case scenario of JOBSTART can be worse than that of New Chance. In conclusion, JOBSTART is preferred to New Chance in terms of efficiency as measured by both estimators.
<insert Figure 1 here>

Conclusions:
Description of conclusions, recommendations, and limitations based on findings.
To respond to the methodological challenge of capturing the uncertainty of an efficiency ratio in cost-effectiveness analysis, this paper synthesizes and evaluates various methods used to quantify uncertainty derived from either sampling variation or incomplete information, and proposes a Bayesian approach that can be used to process the available site-level effectiveness and cost information. Compared to other methods, the Bayesian approach has at least two advantages with regard to informing and guiding the decision making in educational practice. First, it provides direct answers to questions that decision makers are most interested in when they encounter a choice problem related to resource allocation: the best guess on what would happen in terms of efficiency if a program is implemented at a specific site once, and the best-and worst-scenarios. Second, its validity does not depend on the number of observations available. This feature is extremely attractive when site is the unit of analysis and the datasets available usually have limited number of observations in the educational context.

Appendices

Not included in page count.

Appendix A. References

Briggs, A. H., O'Brien, B. J., and Blackhouse, G. (2002). Thinking outside the box: Recent advances in the analysis and presentation of uncertainty in cost-effectivenessstudies. *Annu. Rev. Public Health*, 23:377–401.

Cave, G., Bos, H., Doolittle, F., and Toussaint, C. (1993). JOBSTART final report on a program for school dropouts. Technical report, Manpower Demonstration Research Cooperation, New York, NY.

Fink, B. L. and Farrell, M. E. (1994). *New Chance: The cost analysis of a comprehensive program for disadvantaged young mothers and their children.* Technical report, Manpower Demonstration Research Cooperation, New York, NY.

Levin,H.M.andBelfield, C.R.(2014). Guiding the development and use of cost-effectiveness analysis in education. *Journal of Research on Educational Effectiveness*, pages 1–19.

Levin, H. M., Belfield, C. R., Hollands, F., Bowden, A. B., Cheng, H., Shand, R., Pan, Y., and Hanisch-Cerda, B. (2012). *Cost-effectiveness analysis of interventions that improve high school completion.* Technical report, Teachers College Columbia University, New York, NY.

Quint, J., Bos, J., and Polit, D. (1997). New Chance: *Final report on a comprehensive program for young mothers in poverty and their children.* Technical report, Manpower Demonstration Research Cooperation, New York, NY.

Appendix B. Tables and Figures

Table 1 Mean value and 95% confidence interval of the posterior predictive distribution of the site-level EC ratio

	Mean	2.5th percentile	97.5th percentile
New Chance (complete pooling model)	0.055	-0.121	0.233
New Chance (hierarchical model)	0.055	-0.132	0.248
JOBSTART (complete pooling model)	0.131	-0.229	0.927
JOBSTART (hierarchical model)	0.132	-0.283	1.016

Table 2 Mean value and 95% confidence interval of the posterior predictive distribution of the weighted EC ratio

	Mean	2.5th percentile	97.5th percentile
New Chance (complete pooling model)	0.054	0.003	0.107
New Chance (hierarchical model)	0.052	0.010	0.096
JOBSTART (complete pooling model)	0.108	-0.003	0.267
JOBSTART (hierarchical model)	0.125	0.009	0.259

Figure 1 Comparison of New Chance and JOBSTART

(a) Site-level EC ratio

(b) Weighted EC ratio