Published in final edited form as:

Med Decis Making. 2012 ; 32(3): 397–399. doi:10.1177/0272989X12441398.

Accounting for Biases when Linking Empirical Studies and Models

Jeremy D. Goldhaber-Fiebert, PhD

Stanford Health Policy, Centers for Health Policy and Primary Care and Outcomes Research,
Stanford University, Stanford, CA 94305-6019

Jeremy D. Goldhaber-Fiebert: jeremygf@standord.edu

Comparative effectiveness and cost-effectiveness analyses have become increasingly common and important [1–3]. Such analyses aim to assess multiple interventions by comparing their long-term outcomes and costs for real-world patient populations and care delivery settings. To inform current decisions, empirical studies of sufficient size and duration to assess long-term outcomes are often infeasible. Computer-based mathematical models can play a valuable role by providing the needed estimates. Such models generate outcomes based on inputs derived from shorter-term empirical studies. Therefore, appropriate methods for linking empirical studies to mathematical models underpin rigorous analyses. An important concern is that estimates from empirical studies contain a variety of potential biases which if not corrected can influence modeled outcomes and resulting policy recommendations.

With recent developments such as the establishment of the Patient Centered Outcomes Research Institute [1,2], the methods literature on combining the strengths of empirical studies and mathematical models to support comparative effectiveness and cost-effectiveness analyses has grown rapidly. The article by Gastineau Campos and colleagues in this issue of *Medical Decision Making* contributes to this area [4]. The authors describe an approach for assessing and correcting potential biases in estimates from a randomized controlled trial (RCT) of cervical cancer screening technologies. They then provide an illustrative example of the importance of this approach by showing how such bias corrections might alter cervical cancer screening policy conclusions reached with biased estimates. Their approach is designed for situations where individual-level data from a single RCT are available. It represents a valuable tool for other researchers facing similar situations.

Their approach contributes to a consistent set of best practices for selecting, adjusting for bias, and incorporating empirical evidence into decision-analytic mathematical models -- a need that has been previously acknowledged [5,6]. Such best practices can aid researchers who confront a number of typical situations when selecting empirical study estimates to use in their model-based analyses. These situations can be characterized by: 1) number of available studies (one vs. more than one); 2) study types (randomized controlled trials (RCT) vs. observational studies); 3) available information (summary estimates from published articles vs. individual-level study data).

Ideally, when multiple studies are available, decision-analytic mathematical models should incorporate their combined information. Systematic review and meta-analytic methods, including Bayesian synthesis and meta-regression techniques are well developed [7,8]. One important source of potential bias when combining study estimates is failure to account for study differences that can influence outcomes, an area where meta-regression and other similar techniques can be particularly useful. Another common challenge in combining all available, high-quality studies without introducing bias is accounting for that fact that

studies report different and difficult-to-compare measures. For example, it may be important to assess the impact of diabetes control on health outcomes, but separate studies have measured diabetes control in terms of fasting plasma glucose, oral glucose tolerance tests, or HbA1c. Mapping the effect of a unit change in one measure to that of another requires careful analysis [9].

In attempting to avoid bias, analysts often prefer randomized trials to observational studies, though this choice is not always obvious. The rationale is that randomization balances unobserved covariates producing an unbiased estimate of the effect of the intervention. However, RCT estimates are only unbiased for the often highly selected patient population in which they are conducted. Incorporating them into a model to consider outcomes for a patient population that differs from the trial population risks introducing serious bias. Observational studies often are conducted in larger, less selected patient populations over longer time periods and may provide estimates that are more applicable and generalizable – potentially more appropriate for inputs into a model. However, observational studies may suffer from non-random selection issues that introduce confounding biases. Attempts to correct such biases in observational studies often employ propensity score methods [10]. Propensity score methods estimate effects after balancing observed covariates, but they only provide unbiased causal inference and effect estimates if the strong and non-verifiable assumption of also balancing unobserved covariates is met [11]. In the end, the choice may rest upon a subjective assessment of the relative magnitude and import of non-generalizability versus confounding.

In situations where RCTs are preferred or the only available study is an RCT, researchers may still be worried about biases in study estimates, as Gastineau Campos and colleagues illustrate. For example, when randomization does not result in balance for all observed covariates, it is possible that differences in treatment effects may be due to this imbalance. Even when randomization is successful, if study blinding is incomplete, selection issues can arise in which the *de facto* treatment assignments are not balanced with respect to observed or unobserved covariates, again influencing the observed treatment effects. Gastineau Campos and colleagues provide one practical approach to these challenges and also cite methods developed by Eddy [12]. A particularly challenging case is when biased selection leads to differential follow-up and hence differential missingness of outcomes across arms. While imputation methods are often employed in such situations [13,14], incorrect assumptions about the process generating the missingness can introduce bias. Manski describes bounding methods that require fewer strong assumptions about the generating process of missingness but often lead to large uncertainty intervals [15].

When individual-level data from multiple studies are available, one could imagine employing bias correction methods for each individual study and then using standard meta-analytic methods to combine their estimates. It is unclear whether the bias correction approaches could be directly incorporated into individual-level meta-analytic techniques [16] to maximize the available information while simultaneously reducing biases. This is a potentially important area of future research.

When multiple studies are available but their individual-level data are not, direct adjustments for bias may not be possible. Consistency across multiple studies may then be the best feasible criteria for inclusion in model-based analyses. Direct comparison of study outcomes to assess consistency is often possible. When it is not, the model itself may play a role in the assessment – examining whether inputs from one study lead model-generated outcomes that are consistent with other studies. It is still possible that both the input study and the comparator study are similarly biased, in which case consistency assessment is not helpful. However, when a larger number of studies are available, the possibility of correlated

biases seems less likely. This approach applies techniques developed for model evaluation and external validation [17,18].

The rigor and applicability of comparative effectiveness and cost-effectiveness analyses can benefit from methods that appropriately link unbiased study estimates to decision-analytic mathematical models. Gastineau Campos and colleagues provide a cautionary tale to decision analysts about assessing empirical studies prior to including their unadjusted estimates in mathematical models. Their bias correction approach contributes to the growing body of methods literature in this area.

References

1. http://www.gao.gov/press/pcori_2011jan21.html
2. <http://www.pcori.org/funding-opportunities/past-funding-opportunities/>
3. Neumann PJ, Fang CH, Cohen JT. 30 years of pharmaceutical cost-utility analyses: growth, diversity and methodological improvement. *Pharmacoeconomics*. 2009; 27(10):861–72. [PubMed: 19803540]
4. Campos NG, Castle PE, Schiffman M, Kim JJ. Policy implications of adjusting randomized trial data for economic evaluations: a demonstration from the ASCUS-LSIL Triage Study. *Med Decis Making*. 2012 May-Jun;32(3):400–27. [PubMed: 22147881]
5. Weinstein MC, O'Brien B, Hornberger J, Jackson J, Johannesson M, McCabe C, Luce BR. ISPOR Task Force on Good Research Practices--Modeling Studies. Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices--Modeling Studies. *Value Health*. 2003 Jan-Feb;6(1):9–17. [PubMed: 12535234]
6. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, Woolacot N, Glanville J. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. *Health Technol Assess*. 2004 Sep; 8(36):iii–iv. ix–xi, 1–158. [PubMed: 15361314]
7. Egger, M.; Smith, G Davey; Altman, Doug, editors. *Meta-analysis in Context*. London: BMJ Books; 2001. *Systematic Reviews in Health Care*.
8. Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res*. 2001 Aug; 10(4):277–303. [PubMed: 11491414]
9. Danaei G, Lawes CM, Vander Hoorn S, Murray CJ, Ezzati M. Global and regional mortality from ischaemic heart disease and stroke attributable to higher-than-optimum blood glucose concentration: comparative risk assessment. *Lancet*. 2006 Nov 11; 368(9548):1651–9. [PubMed: 17098083]
10. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70(1):41–55.
11. Pearl, J. *Causality: Models, Reasoning, and Inference*. 2. New York: Cambridge University Press; 2009.
12. Eddy DM. The Confidence Profile Method: A Bayesian method for assessing health technologies. *Operations Research*. 1989; 37:210–228. [PubMed: 10292450]
13. Rubin DB. Inference and missing data. *Biometrika*. 1976; 63(3):581–92.
14. Efron B. Missing data, imputation, and the bootstrap. *JASA*. 1994; 89:463–79.
15. Manski, CF. *Identification for Prediction and Decision*. Cambridge: Harvard University Press; 2007.
16. Simmonds MC, Higgins JP, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clin Trials*. 2005; 2(3):209–17. [PubMed: 16279144]
17. Eddy, DM.; Hollingworth, W.; Caro, JJ.; Tsevat, J.; McDonald, KM.; Wong, JB. DRAFT - Model Transparency and Validation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group - Part 4. http://www.ispor.org/workpaper/modeling_methods/DRAFT-Modeling-Task-Force-Validation-and-Transparency-Report.pdf
18. Goldhaber-Fiebert JD, Stout NK, Goldie SJ. Empirically evaluating decision-analytic models. *Value Health*. 2010 Aug; 13(5):667–74. [PubMed: 20230547]