May, 2017

“All Generalizations Are Dangerous, Even This One.” —Alexandre Dumas

Laura Balzer

Available at: https://works.bepress.com/laura_balzer/45/
COMMENTARY

“All Generalizations Are Dangerous, Even This One.”
—Alexandre Dumas

Laura B. Balzer

In public health and medicine, there is a tension between internal and external validity.1–12 Many interventions are known to be efficacious at the individual level, but less is known about (1) their impact once scaled to a population level, (2) effect modification by both measured and unmeasured factors, and (3) which intervention components should be implemented universally and which adapted to local context. The field of human immunodeficiency virus (HIV) prevention and treatment provides an illustrative example. Randomized trials and observational studies have shown that immediate initiation of antiretroviral therapy for an HIV-positive individual improves his or her health and prevents transmission between couples and from mothers to children.13–16 Four community randomized trials aim to examine the impact of “universal test and treat” (population-wide HIV testing with immediate antiretroviral therapy initiation for all HIV-positive persons) on HIV incidence in several countries in Eastern and Southern Africa.17–20 These trials are pragmatic in that they aim to learn about effectiveness and implementation in real-world conditions. Nonetheless, the specific components of their interventions, their implementation, and their impact are expected to vary within and across trials. Given promising interim results,21 open questions remain about nation-wide rollout and the heterogeneity in expected impact.

In this issue of Epidemiology, Lesko et al.12 highlight the distinction between estimating the effect for the study units and the effect for the target population. Concretely, the sample average treatment effect22,23 is the mean difference in the counterfactual (potential) outcomes for the enrolled units, whereas the population average treatment effect is the expected difference in the counterfactual (potential) outcomes for the population from which the study units were selected. It is worth emphasizing that sample and population effects are fundamentally different causal parameters—even when the units are drawn as a simple random sample from the target population, there is only one version of the treatment, and there is no interference.1,2,11 In other words, if all the assumptions for both internal and external validity held, the sample and population effects are likely to be different.

Consider the simulation study conducted by Lesko et al.,12 hereafter called “the authors.” There is one value of the population average treatment effect, calculated analytically or by Monte Carlo simulations as −5.5%. In contrast, the sample average treatment effect is a data-adaptive parameter; with each new selection of study units, a new value of the sample effect is obtained. To illustrate this point, we replicate the authors’ simulation 5,000 times. For increasing enrollment sizes, we draw a simple random sample of
The sample effect is also an appealing causal parameter but has received less attention in public health literature. As discussed by the authors, we commonly assume the existence of a real or hypothetical target population from which the study units were selected and about which we wish to make inferences. Concretely defining this population is challenging. In contrast, the sample effect avoids all assumptions about a “vaguely defined super-population of study units” and is simply the intervention effect for units at hand. In the Sustainable East Africa Research in Community Health (SEARCH) trial, for example, the sample effect corresponds to the average difference in the counterfactual cumulative HIV incidence under the test and treat strategy and under the standard of care for the \( n = 32 \) study communities. In this example, the sample effect captures the impact for approximately 320,000 people living in rural Uganda and Kenya. As Lesko et al. and others discuss, generalizing this intervention effect to a wider population (e.g., all of Uganda and Kenya) or transporting it to a different population (e.g., Boston or San Francisco) requires additional assumptions and distinct estimators. Finally, the sample effect will be estimated with at least as much precision as the population effect, especially under pair matching. In particular, if there is heterogeneity in the intervention effect by measured or unmeasured factors, a given study will have more power to detect a sample than a population effect. In other words, the price for generalizing the sample to the population is higher variance. Altogether, the interpretability, relevance, and increased precision from specifying the sample average treatment effect as the target of inference make this causal parameter an appealing alternative to the population average treatment effect.

The authors’ presentation is focused on the enrollment (sampling) mechanism, and their estimators are derived under the potential outcomes framework. An alternative would be to consider the structural causal model of Pearl. Recall the authors’ notation with \( W \) as the set of characteristics influencing enrollment, \( S \) as indicator of being selected into the study, \( A \) as an indicator of receiving the exposure, and \( Y \) as the outcome. Specifically, we consider a binary exposure with \( A = 1 \) for the intervention and \( A = 0 \) for the control. For simplicity, we define the exposure and outcome \((A, Y)\) to be 0 for units not selected \((S = 0)\). Then the following causal model would describe a study (observational or randomized) wherein units are enrolled as a function of baseline covariates and the exposure is assigned as a function of baseline covariates and enrollment:

\[
W = f_W(U_W) \\
S = f_S(W, U_S) \\
A = f_A(W, S, U_A) \\
Y = f_Y(W, S, A, U_Y)
\]

Here \( \{U_W, U_S, U_A, U_Y\} \) denote the corresponding set of background or unmeasured factors that have some joint distribution \( \mathbb{P}_{U_i} \). This approach for representing selection into the study (a threat to external validity) could easily be extended to include missing data after enrollment (a threat to internal validity) as well as postenrollment covariates that influence the exposure assignment (additional confounders, another threat to internal validity; Appendix A). Furthermore, this causal model is defined at the unit level, and implicitly assumes no interference. (This framework could also be extended to handle interference.) The authors’ simulation example is one possible data-generating process, compatible with this structural causal model. In a randomized trial (such as considered by the authors), the unmeasured factors contributing to the intervention assignment \( U_A \) are independent of the others and the covariates \( W \) do not impact randomization \( A \).

We assume that the above causal model provides a description of the data-generating process under existing conditions and under specific interventions. (See Appendix 1 in Bareinboim and Pearl for a short introduction.) Counterfactual outcomes are generated by intervening on this causal model. To define the sample effect, we intervene to set the exposure \( A = a \) to generate the counterfactual outcome \( Y_i(a) \): the outcome that would have been observed if unit \( i \) received exposure level \( A = a \). Then the sample average treatment effect (SATE) is defined as the average difference in these counterfactual outcomes among enrolled units \((S = 1)\):

### TABLE 1. Summary of the Sample Effect (in %) Over 5,000 Simulations Under Simple Random Sampling of Size \( n = \{100, 500, 2,000\} \) and the Authors’ Biased Sampling Scheme of Size \( n = 2,000 \)

<table>
<thead>
<tr>
<th>Sample Average Treatment Effect</th>
<th>Minimum</th>
<th>Median</th>
<th>Maximum</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 100 ) simple</td>
<td>-14.8</td>
<td>-5.8</td>
<td>9.2</td>
<td>0.10</td>
</tr>
<tr>
<td>( n = 500 ) simple</td>
<td>-10.2</td>
<td>-5.6</td>
<td>0.2</td>
<td>0.02</td>
</tr>
<tr>
<td>( n = 2,000 ) simple</td>
<td>-8.0</td>
<td>-5.5</td>
<td>-2.9</td>
<td>0.006</td>
</tr>
<tr>
<td>( n = 2,000 ) biased</td>
<td>-12.8</td>
<td>-10.4</td>
<td>-8.3</td>
<td>0.005</td>
</tr>
</tbody>
</table>

© 2017 Wolters Kluwer Health, Inc. All rights reserved.

www.epidem.com | 563

---

**Epidemiology • Volume 28, Number 4, July 2017**
where \( n \) denotes the total number of units in the study. To define the population average treatment effect (\( PATE \)) in the context of biased sampling, we consider a hypothetical intervention to enroll the entire target population (i.e., set \( S = 1 \)) and assign the exposure \( A = a \). We denote the counterfactual outcome under this joint intervention as \( Y(s = 1, a) \). The \( PATE \) is then given by the expected difference in these counterfactual outcomes across the target population of interest:

\[
PATE = \mathbb{E}[Y(1, 1) - Y(1, 0)].
\]

For illustration, we repeat the authors’ simulation study 5,000 times: (1) generate a target population of 50,000 units; (2) from that population, draw a biased sample of \( n = 2,000 \) units; and (3) for each sample, calculate the \( SATE \) as the average difference in the counterfactual outcomes for the enrolled units (R code in Appendix C). As shown in Table 1, the sample effect under the biased sampling scheme ranges from −12.8% to −8.3% and is −10.4% on average. The sample effect is larger on average than the population effect (−5.5%), because units at higher risk of the outcome are selected into the study. Practically, this may suggest that instead of rolling out the intervention to the entire population, the greatest impact could be obtained by targeting the intervention to high-risk groups. Likewise, the effect heterogeneity may suggest alternative parameters of interest, such as the conditional average treatment effect, an intermediate connection between the authors’ assumptions and estimators (Appendix B).1,28,29

The structural causal model representation also emphasizes that a rich toolkit of estimators could be used to correct for biased sampling, which is presented by the authors as a threat to external validity. The nonparametric estimators, implemented by the authors and here, will break down when many covariates or a single continuous covariate influence unit selection (and/or the exposure mechanism in an observational setting). As an alternative, we could immediately implement augmented inverse probability weighting or targeted maximum likelihood estimation.35,38,39 These methods are double robust and can incorporate data-adaptive (machine learning) algorithms to relax parametric modeling assumptions while retaining valid statistical inference.

An important open question, not addressed by Lesko et al.12 nor in this commentary, is generalizability and transportability when the intervention (or its specific components) must be adapted to local context. We should be wary of assuming the sample effect is immediately generalizable to the population. We should also be wary of assuming that a one-size-fits-all intervention is best.
ACKNOWLEDGMENTS
The author thanks Maya Petersen and Victor DeGruttola for their expert advice.

ABOUT THE AUTHOR
LAURA B. BALZER is a postdoctoral fellow in Biostatistics at the Harvard T.H. School of Public Health. She is a methodologist with substantive interests in global health, community-based participatory research, and social determinants of health. Laura is an expert causal inference and semiparametric estimation. With Maya Petersen, she was awarded the 2014 Causality in Statistics Education Award from the American Statistical Association.

REFERENCES
APPENDIX A

Let $Z$ denote the set of postenrollment characteristics influencing exposure assignment and $\Delta$ be an indicator that a unit has its outcome measured (i.e., is not loss to follow-up). For simplicity, we define the postenrollment characteristics, the exposure, the missing data indicator, and the outcome $(Z, A, \Delta, Y)$ equal to 0 for units not enrolled ($S = 0$). Then the following structural causal model would describe a study (observational or randomized) wherein units are enrolled as a function of baseline covariates, the exposure is rolled out as a function of baseline and postenrollment characteristics, and missingness on the outcome is not random:

$$W = f_W(U_W)$$

$$S = f_S(W, U_S)$$

$$Z = f_Z(W, S, U_Z)$$

$$A = f_A(W, S, Z, U_A)$$

$$\Delta = f_\Delta(W, S, Z, A, U_\Delta)$$

$$Y = f_Y(W, S, Z, A, \Delta, U_Y)$$

Here $(U_W, U_S, U_Z, U_A, U_\Delta, U_Y) \sim P_U$ denote the corresponding set of background or unmeasured factors with some distribution. Let $Y(s = 1, a, \delta = 1)$ denote the counterfactual outcome for a given unit under a hypothetical intervention to ensure its enrolled (i.e., set $S = 1$), assign the exposure $A = a$, and ensure its outcome is measured (i.e., set $\Delta = 1$). Then the PATE is defined as

$$\mathbb{E}[Y(1, 1, 1) - Y(1, 0, 1)].$$

Under the sequential randomization and positivity assumptions,32 the corresponding statistical estimand could be estimated with a variety of methods, including longitudinal parametric G-computation,40 longitudinal inverse probability weighting,33,41 and longitudinal targeted maximum likelihood estimation.35,42

APPENDIX B

In 2002, Abadie and Imbens28 proposed the conditional average treatment effect as

$$CATE = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[Y_i(1) - Y_i(0) | W_i].$$

where $i$ indexes the $n = 2,000$ units selected for the study. The conditional effect is interpreted as the average intervention effect given the covariates of the study units and is equal to $-10.4\%$ under this biased sampling scheme:

$$\frac{1}{2000}[320 \times (-0.05) + 480 \times (-0.05) + 480 \times (-0.05) + 720 \times (-0.05 - 0.15)] = -10.4\%$$

APPENDIX C

Simulation studies were conducted in R-3.3.2.43 Full R code and the resulting dataset are available at https://github.com/LauraBalzer/On-Generalizability.