Reducing Bias in Estimates of Per Protocol Treatment Effects
A Secondary Analysis of a Randomized Clinical Trial

Stephen R. Cole, PhD; Jessie K. Edwards, PhD; Paul N. Zivich, PhD; Bonnie E. Shook-Sa, DrPH; Michael G. Hudgens, PhD; Jeffrey S. A. Stringer, MD

Introduction

Intent-to-treat effect estimates from randomized trials are influenced by participant adherence to the treatment protocol. When protocol adherence is imperfect, estimation of per-protocol effects is recommended.\(^1,2\) Frequently, per-protocol effects are estimated by censoring or excluding participants who deviate from a protocol (Table). This approach provides valid estimates if protocol deviations are noninformative (ie, completely at random within treatment groups).

However, protocol deviations can be informative, especially when deviations are caused by factors associated with risk for the outcome. Reporting per-protocol effects without accounting for the possible bias induced by censoring protocol deviations is analogous to reporting an observational study with no control for confounding. To minimize the potential for such bias, one can account for protocol deviations using a per-protocol estimator that standardizes for observed variables with inverse probability weights,\(^3\) generalized computation,\(^4\) or comparable approaches.

Methods

In this secondary analysis of a randomized clinical trial, we illustrate how failing to account for informative protocol deviations can lead to biased estimates using data from a trial of 17 alpha-hydroxyprogesterone caproate (17P) to reduce preterm birth among HIV positive pregnant people. This study was approved by the local institutional review board, and participants provided written informed consent. The study followed the Consolidated Standards of Reporting Trials (CONSORT) reporting guideline (see Supplement 1 for further details). Data were analyzed with SAS version 9.4 (SAS Institute) in February 2023.

<table>
<thead>
<tr>
<th>Source(^a)</th>
<th>Primary outcome</th>
<th>Treatment groups</th>
<th>Estimate (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Intent-to-treat</td>
</tr>
<tr>
<td>Goldenberg et al, 2022</td>
<td>Symptomatic recurrent venous thromboembolism by 1 y</td>
<td>Anticoagulant therapy of 6 wk (207 participants) or 3 mo (210 participants)</td>
<td>−1.1 (−5.2 to 2.3)</td>
</tr>
<tr>
<td>Grillot et al, 2023</td>
<td>Successful tracheal intubation on first attempt</td>
<td>Neuromuscular blockers (575 participants) or remifentanil (575 participants)</td>
<td>−6.1 (−11.6 to −0.5)</td>
</tr>
<tr>
<td>Nyang’wa et al, 2022</td>
<td>Death, treatment failure or stop, loss to follow-up, or tuberculosis recurrence by 72 wk</td>
<td>Bedaquiline, pretomanid, linezolid, and moxifloxacin (151 participants) or standard of care (152 participants)</td>
<td>−30 (−46 to −14)</td>
</tr>
<tr>
<td>Reis et al, 2022</td>
<td>Hospitalization or emergency department visit due to COVID-19 by 28 d</td>
<td>Ivermectin for 3 d (679 participants) or placebo (679 participants)</td>
<td>0.90 (0.70 to 1.16)(^b)</td>
</tr>
<tr>
<td>Turkova et al, 2022</td>
<td>Treatment failure, loss to follow-up, or death by 72 wk</td>
<td>Standard first-line pediatric antituberculosis treatment for 16 (602 participants) or 24 (602 participants) wk</td>
<td>−0.6 (−2.3 to 1.4)</td>
</tr>
<tr>
<td>Van der Vaart et al, 2022</td>
<td>Patient-reported improvement at 24 mo</td>
<td>Pessary treatment (218 participants) or surgery (222 participants)</td>
<td>−6.1 (−12.7 to 0)(^c)</td>
</tr>
<tr>
<td>Varghese et al, 2023</td>
<td>Death by 28 d, persistent complications at day 7, or persistent fever at day 5</td>
<td>Intravenous azithromycin and doxycycline (266 participants) or doxycycline (265 participants)</td>
<td>−13.3 (−21.6 to −5.1)</td>
</tr>
</tbody>
</table>

Abbreviation: NA, not available.

\(^a\) Further information about articles is available in the eAppendix in Supplement 2.

\(^b\) Reis et al provides a risk ratio; all other sources provide risk differences.

\(^c\) One-sided CIs.
Results

In the Improving Pregnancy Outcomes with Progesterone (IPOP) trial, 580 HIV-positive pregnant people were randomly assigned to weekly injections of 17P (399 participants) or placebo (401 participants) to reduce preterm birth (before 37 weeks) between 2018 and 2020. Nearly all participants (760 participants; 95%) adhered to a protocol allowing only 1 missed injection. The estimated intent-to-treat effect was a risk ratio (RR) of 1.00 (95% CI, 0.64-1.55). At randomization, 363 participants (45%) had a mid-trimester cervical length less than 4 cm, which was associated with a 2-fold risk of preterm birth.

To illustrate the potential for bias arising from differential protocol deviations, suppose contrary to fact, cervical length was also a predictor of protocol deviation, such that 152 of 186 (81%) participants in the 17P group with a cervix less than 4 cm deviated from the protocol but only 95 of the remaining 614 (15%) participants deviated from the protocol. Under such a scenario, if we censor protocol deviants as did the studies in the Table, then the unstandardized per-protocol RR estimate comparing 17P vs placebo in the 553 uncensored participants is 0.74 (95% CI, 0.41-1.32), suggesting a lower risk of preterm birth in the 17P group, as the authors hypothesized. But the standardized per-protocol RR estimate, accounting for cervical length using inverse probability weights, is 1.02 (95% CI, 0.52-2.02). This standardized result is as expected because the estimated intent-to-treat RR was 1 and actual protocol deviations were minimal.

Discussion

When protocol deviations are not completely at random within treatment groups, an unstandardized per-protocol effect estimator can be biased, even grossly, as shown in the example. The example enjoys all the benefits of counterexamples but has the limitations of being partially contrived and having restricted scope. Using postrandomization information on protocol adherence to censor participants requires accounting for possible induced bias. Standardized per-protocol effect estimators will generally be valid if the common causes of protocol deviations and outcomes are measured and accounted for appropriately. However, typical regression models fail to account for postrandomization variables appropriately even when a sufficient set of variables is measured. But standardization methods such as inverse probability weighting and generalized computation can provide valid per-protocol effect estimates when a sufficient set of variables is measured.

Although there is 1 intent-to-treat effect (for the observed amount of adherence), there are as many per protocol effects as there are possible definitions of protocol deviation. In the 17P example previously mentioned, participants were classified as deviating from protocol when they missed more than 1 weekly injection, but protocol deviations could have been defined less stringently as, for example, more than 3 missed injections. In conclusion, we recommend randomized trials report the estimated intent-to-treat effect, and when protocol deviations occur, a series of standardized per-protocol effect estimates which vary the stringency of the definition of a protocol deviation.
Health, University of North Carolina, Chapel Hill (Shook-Sa, Hudgens); Department of OB/GYN, School of Medicine, University of North Carolina, Chapel Hill (Stringer).

**Author Contributions:** Dr Cole had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Concept and design:** Cole, Edwards, Zivich, Shook-Sa, Hudgens.

**Acquisition, analysis, or interpretation of data:** Cole, Edwards, Stringer.

**Drafting of the manuscript:** Cole, Hudgens.

**Critical review of the manuscript for important intellectual content:** Cole, Edwards, Zivich, Shook-Sa, Stringer.

**Statistical analysis:** Cole, Shook-Sa.

**Obtained funding:** Cole, Edwards, Stringer.

**Administrative, technical, or material support:** Cole, Edwards, Zivich.

**Supervision:** Hudgens.

**Conflict of Interest Disclosures:** Dr Edwards reported receiving grants from the National Institutes of Health (NIH) National Institute of General Medical Science outside the submitted work. Dr Shook-Sa reported receiving grants from NIH, grants from Doris Duke Charitable Foundation, and grants from North Carolina Department of Health and Human Services outside the submitted work. Dr Hudgens reported receiving grants from NIH outside the submitted work. No other disclosures were reported.

**Funding/Support:** NIH grants R01AI157758 (to Drs Cole, Edwards, Shook-Sa, Hudgens, and Stringer), R01HD087119 (to Drs Stringer and Cole), P30AI050410 (to Drs Hudgens, Stringer, Shook-Sa, and Cole), T32AI07001 (to Dr Zivich).

**Role of the Funder/Sponsor:** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Data Sharing Statement:** See Supplement 3.

**Additional Contributions:** We thank Mathew L. Maciejewski, PhD (Duke University) and Enrique F. Schisterman, PhD (University of Pennsylvania) for expert advice. They were not compensated for their contributions.

**REFERENCES**


**SUPPLEMENT 1.**

Trial Protocol

**SUPPLEMENT 2.**

eReferences

**SUPPLEMENT 3.**

Data Sharing Statement