Author’s response to reviews

Title: A Simple Method for Analyzing Data from a Randomized Trial with a Missing Binary Outcome

Authors:

Dr Stuart G Baker (sb16i@nih.gov)
Laurence S Freedman (lsf@actcom.co.il)

Version: 2 Date: 16 Apr 2003

PDF covering letter
We thank Dr. Follman for the careful reading and helpful comments which have substantially improved the manuscript.

1. We now added a hat to the variance and explicitly write the steps in deriving (8):

"Using the delta method, and noting that \( \hat{\Delta} = d_1 w_1 + d_2 w_2 + \ldots + d_{h-1} w_{h-1} + d_h (1 - \sum_{s=1}^{h-1} w_s) \), we obtain

\[
\hat{\text{var}}(\hat{\Delta}) = \sum_{s=1}^{h} \left( \frac{\partial \hat{\Delta}}{\partial w_s} \right)^2 \hat{\text{var}}(d_s) + \sum_{s=1}^{h-1} \left( \frac{\partial \hat{\Delta}}{\partial w_s} \right)^2 \hat{\text{var}}(w_s) \\
= \sum_{s=1}^{h} w_s^2 \sum q_{sz} (1 - q_{sz}) / n_{sz^+} + \sum_{s=1}^{h-1} (d_s - d_h)^2 w_s (1 - w_s) / N_+,
\]

(8)

where \( w_h = 1 - \sum_{s=1}^{h-1} w_s \)."

2. Along the same lines we write the equation as:

\[
\Delta_s = pr(Y = 1 \mid Z = 1, s, x) - pr(Y = 1 \mid Z = 0, s, x) \\
= pr(Y = 1 \mid Z = 1, s, x, R = 1) - pr(Y = 1 \mid Z = 0, s, x, R = 1) \text{ from (9)}
\]

(11)

We also added a similar equation for (19).

3. Equation (19) is based on parametrization that specifies an effect of \( Z = 1 \) instead of \( Z = 0 \) (rather than vice versa). Therefore because of this asymmetry with respect to \( Z \) in the parametrization, we do not think that \( \phi_s \) would be the same if we had conditioned on \( Z = 1 \).

4. We changed the description of (19) to "effect of \( X \) on subjects in the control group with observed outcomes", as suggested. We also added (following up on point 2 above)

"By virtue of the MAR assumption in (9), we could also write \( \psi_s = pr(Y = 1 \mid Z = 0, s, X = 1) - pr(Y = 1 \mid Z = 0, s, X = 0) \), which is the effect of \( X \) on all subjects in the control group."

5. We changed the phrase to, "measures the degree of confounding between \( X \) and \( Z \) among subjects with observed outcomes (i.e. \( R = 1 \))"

6. We added a table similar to the proposed table. For readers who prefer graphical displays, we also added a figure.

7. We agree. We rewrote this part as follows.
"One might argue that if \( x \) were a strong unobserved inherited gene, \( \psi_{max} \) would be close to 1. However because, "eligible subjects had no history of colorectal cancer, surgical resection of adenomas, bowel resection, the polyposis syndrome, or inflammatory bowel disease" [14], it is unlikely that many subjects had an unobserved high-penetrance gene related to the recurrence of adenomas. We therefore we believe that unobserved factors that might affect both adenoma recurrence and missingness could have an effect of similar magnitude as observed baseline covariates. Thus to obtain a plausible lower bound for \( \psi_{max} \), we suggest estimating \( \psi_0 \), as defined in (19), based on observed covariates".

8. Done. We replaced "treatment" with "randomization".

9. We corrected the typo in (A.1).

10. There was a typo. We change \( k_{zs} \) to \( \kappa_{zs} \), which was defined previously.