**Associate Editor**

(1) The reviewers appear to have different views on the bias of omitted variables in shifting-case analyses. Reviewer 2 (#4) suggests reminding readers about the omitted variable bias in meta-regression. On the other hand, Reviewer 1 (#9) believes that the bias in shifting-case analyses should compare estimates of these coefficients to the true parameter values in the restricted model. There are many potential covariates in a meta-regression, e.g., year of publications, mean age of participants, and duration of intervention. If we always assume that the model with all covariates is correct, all restricted models with fewer covariates are misspecified. I think that this issue needs further discussion.  
  
(2) Another related issue is the substance abuse interventions example, which includes Group 1 Hi-Int. and Group 2 Hi-Int. as two covariates. It gives readers an impression that we should always include both covariates in the analysis. However, if the covariates are year of publications, mean age of participants, and duration of intervention, does it make sense to include them only one at a time? Moreover, the effect sizes in the example are correlated; this may complicate the results because the paper assumes independent effect sizes (see also Reviewer 1, #1, #2, and #3).  
  
(3) The shifting-case analysis (SCA) seems the same as the pairwise deletion in data analysis. It will be helpful to mention it so that readers from other disciplines find it easier to follow.  
  
(4) Figure 4 suggests that the omitted variable bias and missingness bias are always in opposite directions. Is it true?  
  
(5) The current version does not follow the required format specified in the Author Guidelines. For example, the AMA reference style must be used. Please refer to the Author Guidelines for details.

**Reviewer 1**

1. Section 2. In the substance abuse interventions example, it is mentioned that “Often the same group (typically the control group) in a study was used in multiple contrasts...”. For a control group, how would a covariate indicating high- versus low-intensity intervention defined? It seems that this would be undefined, rather than missing?
2. Section 2. Page 3. The discussion contrasting coefficients for X1 in complete-case analysis (Column 1 in Table 2) and the coefficient for X1 in Shifting-case Group 1 (Column 2 in Table 2; and similarly for X2) is misleading. They are not the same parameter and are usually not expected to the same unless under some special cases. They are not comparable. Whether they are close or different does not necessarily mean any estimate is inaccurate. Both can be correct for their corresponding parameter of interest.
3. Section 3. Page 4. It would be helpful to make it clearer that the model formulation (1), individual and joint likelihood contributions (2) and (3) correspond to the setting where there are a total of k studies, and each study contributes one Ti. The main technical results seem to focus on this setting (i.e., Ti ’s are independent) while the example in Section 2 corresponds to the setting where Ti ’s can be correlated.
4. Page 4, line 2 after model (1), typo in “which is true of some effect sizes”?
5. Section 3.2, page 6, line 49-50, the coefficients βS is not necessarily a subset of the full vector of coefficients β, the full model with all covariates and a restricted model with a subset of covariates can both hold, the restricted model would correspond to the full model with the remaining variables marginalized out. βS and the corresponding subset of β in the full model are not directly comparable.
6. Section 4, page 9, I fail to understand expression (11). Let T denote an estimator for a target parameter θ, then bias of an estimator is defined as E(T) − θ. I don’t understand the sentence “The bias term δij refers to the bias induced in effect estimate i due to conditioning on missing pattern Rj ”. In the current meta-regression setting, the effect estimate Ti are obtained irrespective of Xi . The potential bias due to missingness in Xi is only relevant when considering the coefficient of Xi in the meta-regression model relating Ti and Xi.

This is a really good point, and something that got us to think through our language more thoroughly. Essentially, delta is the bias induced by looking only at Ti for which Xi are observed. Particularly if Xi are observed more frequently for larger Ti, then if we only use those Ti to estimate θ, we will likely overestimate θ (since we would be using the larger Ti to do so). We have revised that section to read:

1. Section 4, page 9, please define the terms in equation (12). What is mj? what are fmj (·)’s? With p covariates, there will be 2 p missing patterns. How would model (12) work? Are there additional assumptions required to model the missingness mechanisms?
2. Section 5.1, page 12, the wi ’s in (18) and (19) are undefined.
3. Section 6. Please see my comment #2. In linear regression models, coefficients in a restricted model (i.e., based on a subset of covariates) are not necessarily a subset of the coefficients of the same variables in the full model with the presence of other covariates. The bias in shifting-case analyses should compare estimates of these coefficients to the true parameter values in the restricted model. More specifically, use linear regression models as a simple example:

E(Y | X1, X2) = α + β1X1 + β2X2 (1)

E(Y | X1) = α ∗ + β ∗ 1X1 (2)

E(Y | X2) = ˜α + β˜ 2X2 (3)

Note that β ∗ 1 in general is not equal to β1, and β˜ 2 is not equal to β2. There are full distributions such that all models (1)-(3) are correct. When we fit models (2) and (3), the target parameters are β ∗ 1 and β˜ 2, not the original β1 and β2. They have different interpretations. The conditions the authors try to identify are those that yield β ∗ 1 = β1, and β˜ 2 = β2. When these conditions are not met, that does not mean βˆ S will be biased.

This is a really good point. Our thinking was that SCA are at times used when someone wants to fit the full model (Y | X1, X2), but due to missingness patterns lacks the degrees of freedom to do so, and so they fit complete-case models for (Y | X1) and (Y | X2). If the original intent is the full model, then the omitted variable bias described in () and () would seem to apply. However, if instead only the marginal relationships between Y and X1 and Y and X2 are of interest, then omitted variable bias may be less important. We have drawn attention to this distinction in the revision on page XX: “”

In general, since the bias of an SCA depends on the intent of the analyst, this would seem to be something that may not be known unless analysis plans or hypotheses are registered prior to data collection.

**Reviewer 2**

1. The manuscript should include additional discussion around the appropriateness/reasonableness of using the log-linear distribution for the underlying selection models. How might the conclusions and recommendations for practice vary under alternate selection models, and under what real-world scenarios might log-linear models be more or less appropriate?

This is a great question. In general, it is impossible to know if a selection model is properly specified, since the precise missingness mechanism is almost never truly known in practice. However, we feel the use of log-linear models is reasonable here for a few reasons.

1. Link probably doesn’t matter much.
2. General result is very flexible.
3. Logit links are naturally associated with common understandings of the odds of missingness, in fact, the log odds ratio is a common effect size in MA.

Since analyses condition on *R* ∈ **R** for some collection of missingness patterns **R**, their bias will depend on the distribution of the event *R* ∈ **R**. There are certainly alternative link functions we could use in the GLM framework, but approximations for other link functions are messier (e.g., for probit link functions) but wind up depending on the same general factors (i.e., missingness rate and correlation of missingness with effect size). Though we restrict our focus to the GLM framework for selection/missingness, our general result relies on a very flexible model that can account for nonlinear relationships (or interaction terms) between *T*, *v*, or *X* and the probability of selection. Thus, our general result encompasses a wide class of possible missingness mechanisms. Finally, most available software that make likelihood-based adjustments for missing covariates in regression models assume a log-linear selection.

2. It seems that additional attention may be needed in the discussion section around the finding that bias will be larger in CCA when there are larger values of between-studies heterogeneity. This seems particularly relevant and noteworthy given that meta-regression is often of greatest interest when there is a large amount of between-studies variance (i.e., this is the exact scenario in which many analysts will pursue estimation of meta-regression models).

3. Are the results/conclusions reported here sensitive to any other design or analysis features, such as effect size metric, estimator used for tau-squared, measurement level of the covariate with missing data, and marginal/joint distribution(s) of the covariate(s) with missing data?

Another excellent set of questions. These do not assume any specific effect size metric, so long as the effect estimates are normally distributed. This is true for a large number of metrics on which we conduct meta-analyses (e.g., *z*-transformed correlations), and is a very accurate large sample approximation for others. When the assumptions of our underlying model are violated (effect estimates are not normal or their variance is not known), then these results may not be accurate.

The results here are not specific to any single variance component (tau-squared) estimator. In fact, the use of the weight matrix in (4) assumes that tau is known. If it is estimated with bias or substantial uncertainty, we expect bias to be even worse! [NOTE: Add language about this!]

The results assume that covariates are measured at the effect estimate level, but make no assumptions about their underlying distribution. As such the general result is somewhat difficult to unpack, and hence we use examples to elucidate the key properties of our general result for bias.

4. Even though omitted variable bias is not a concern in CCA when considering bias that may emerge from dropping missing data, it is probably worth reminding readers that omitted variable bias is always a concern in meta-regression (regardless of the approach used to address missing data for covariates). Without such clarification, the conclusions as reported could be misinterpreted to imply that omitted variable bias is only a concern in SCA.

5. The application of Cohen’s rules of thumb for interpreting effect size magnitude is not necessary, and potentially misleading to readers; this should probably be dropped entirely. These rules of thumbs are acontextual and were never intended to be used as universal rules of thumb for interpreting effect size magnitude. It seems better to let the results speak for themselves regarding the resulting bias that may emerge under different scenarios, rather than citing these problematic rules of thumb that are not empirically based (nor empirically validated).