Author’s response to reviews

Title: Bayesian adjustment for measurement error in continuous exposures in an individually matched case-control study

Authors:

Gabriela Espino-Hernandez (gaby@stat.ubc.ca)
Paul Gustafson (gustaf@stat.ubc.ca)
Igor Burstyn (igor.burstyn@drexel.edu)

Version: 2 Date: 13 March 2011

Author’s response to reviews: see over
Re: BMC Medical Research Methodology, MS 1193068125031207

Dear Editors:

We thank the two reviewers for their insightful comments, and we have revised the paper in light of them. Our responses are below.

Yours Sincerely,
Gabriela Espino-Hernandez
Paul Gustafson
Igor Burstyn

Response to reviewer #1

Comment 1 concerning Bayesian estimation of the measurement error variance:
We agree that it would be better to be fully Bayesian with how we treat the measurement error covariance matrix. However, given the complicated relationship by which the quality control data inform this matrix, we are stymied at present on how to do this. We now remark in the Conclusions section that fully Bayesian modeling would be preferred, at least when the form of the validation data is simpler.

Comment 2 concerning the sensitivity analysis and MCMC performance: We now report on this in the Results section. The MCMC performance degrades slightly as the assumed measurement error magnitude increases, but the method doesn’t ‘break’ in any sense.

Comments 3-5: These typos have been corrected.
Comment 6 on run times: Computing times are now reported in the Results section.

Response to reviewer #2

Comment 1 on measurement error models: We now try to stress that the Bayesian approach is modular, i.e., a different measurement error model could be substituted in without any big change in approach. Regarding multiplicative error, we now try to stress more that our example assumes additive error for log-PFA-concentration, hence is indeed making a multiplicative error assumption for concentration.

Comment 2 on flexible measurement models: We have added remarks and citations on flexible measurement models, in the last section. It seems that most of this work emphasizes univariate exposure, so it may not be so easy to extend to multivariate exposure. We also note that the log of the measured PFA concentrations do not exhibit very much skewness, so we suspect the multivariate normal assumption is not too risky for these data.

Comment 3 on order: We have reorganized the methods section, as suggested. The reorganized sections’ title have been highlighted.

Comment 4 on other methods: We took a close look at this issue, to the point of implementing a version of regression calibration on the PFA data. We wrote this up, and it is appended to this letter. Ultimately we lean against including this in the paper, because it makes the paper long and unwieldy. We do now raise the issue in the Conclusions section though, emphasizing that for our context regression calibration isn’t an easier-to-implement option (nor do there appear to be other easier options). One needs variance component estimates from a random effect model for multivariate unbalanced data – so one still needs WinBUGS or some other such heavy-computing tool.
**Comment 5 on alternative data structures:** Our original wording was too vague. We now try to be more explicit about the possibility of adapting to other (simpler!) data sources informing the measurement error variance.

**Appendix to cover letter: Regression calibration technique**
Regression calibration is one of the most commonly used techniques to correct for measurement error in explanatory variables. The basis of this technique is to use additional information, for instance validation data, in order to replace $X$ in the disease model by the regression of $X$ on $W$. In order to avoid misleading results, this technique requires a linear homoscedastic relationship between $X$ and $W$.

In reference to the epidemiological matched case-control study on PFAs [3], the only validation data available corresponds to percentages of recovery for each exposure as a result of performing a quality control procedure on the PFAs in ppb concentrations. Based on this information, in Appendix I, we presented an estimation of the measurement error variability $\Sigma$.

Using the limited validation data available for this particular study, in the form of the measurement error variability $\Sigma$, the regression calibration technique was implemented as follows:

First, the vector of true exposures was estimated by fitting the exiting model in two stages. That means, we assume that the vector of surrogate exposures for the $j$–th subject from the $i$–th matched set follows a multivariate normal distribution around the vector of exposures. Meanwhile, the vector of exposures follows a multivariate normal distribution around the vector of exposure means of the corresponding matched set. That results in

$$W_j \mid X_j, \Sigma \sim N_p(X_j, \Sigma)$$

and

$$X_{ij} \mid \mu, V \sim N_p(\mu, V).$$
Under a Bayesian scheme, the true exposures can be estimated as the estimated posterior mean of $X_{ij} | W_{ij}, \Sigma, \mu, V_w$, as follow

\[
\hat{X}_{ij} = E[X_{ij} | W_{ij}, \Sigma, \mu, V_w]
= E[X_{ij} | W_{ij}, \Sigma, \hat{\mu}, \hat{V}_w]
= \left[\Sigma^{-1} + \hat{V}_w^{-1}\right]^{-1}\left[\Sigma^{-1} W_{ij} + \hat{V}_w^{-1} \mu\right]
\]

In order to estimate the within-covariance matrix $V_w$ and the vector of exposure means of the corresponding matched set $\mu$, a random-effect model was used as follow. Assume,

\[
W_{ij} | \mu, A \sim N_P(\mu, A), \text{ with } A = \Sigma + V_w
\]

and

\[
\mu | \mu_B, V_B \sim N_P(\mu_B, V_B)
\]

Since this a multivariate random-effect model fitted to unbalanced data, there are limited options for software to do the fitting. We again turn to the WinBUGS software, version 1.4.3 [24] and two MCMC chains of length 55,000 were run, using different initial values. Making use of the last 50,000 MCMC iterations, the posterior mean of $\mu$ and $A$ were obtained. The convergence to the posterior distributions and mixing of the two chains were assessed from the trace, autocorrelation, and the Gelman-Rubin convergence statistic plots.

Therefore, the within-covariance matrix $V_w$ and the vector of exposure means of the corresponding matched set $\mu$, were estimated as

\[
\hat{V}_w = \hat{A} - \Sigma,
\]

and

\[
\hat{\mu} = E[\mu | W_{ij}, A, \mu_B, V_w]
= \left[\hat{V}_B^{-1} + n_i A_i\right]^{-1}\left[\hat{V}_B^{-1} \mu + n_i A_i \mu\right]
\]
By substituting, $\hat{V}_W$ and $\hat{\mu}_i$ into equation (A.1), the vector of estimated true exposures $\hat{X}_{ij}$ was obtained for each $j-th$ subject from all the $i-th$ matched sets.

Second, according to the spirit of regression calibration technique, the conditional logistic regression was carried out replacing $X_{ij}$ by $\hat{X}_{ij}$. Two models are considered in the analysis, a simpler model and a model adjusted by the confounding variables. The models were implemented using the statistical package R, version 2.11.1 and the results are displayed in Table A.1.

**Table A.1 – Regression calibration method: Estimated ORs and their 95% confidence intervals**

Point estimated and 95% confidence intervals of the ORs for the simple model and the model adjusted for confounding variables obtained with the regression calibration technique.