Bayesian Analysis (2007) **2**, Number 1, pp. 37–42

Comment on Article by Dominici et al.

David Ruppert* and Raymond J. Carroll†

We thank the authors for a very thought-provoking paper. They address a problem of great importance and have proposed an interesting and ingenious solution. It is quite challenging to develop a model that is suitable for the complex data structure presented by this problem, and the authors should be congratulated for their success.

We believe in the philosophy of using whatever works best, and for complex problems such as this one, a fully Bayesian analysis seems much more satisfactory than anything else we know of. As the authors show, a Bayesian analysis allows one to multiply impute missing data, the true values of mismeasured covariates, and counterfactuals and to adjust inference for uncertainty in the imputations. Figure 4 shows the importance of the latter. As we will discuss shortly, there are a few places where the authors could have improved their analysis by taking further advantage of Bayesian techniques. For example, Bayesian modeling would allow for a more data-driven choice of the spline model. We also believe that their measurement error model (9) could be improved by a fully Bayesian analysis rather than using a regression that is fit outside the Gibbs sampler.

The authors use counterfactuals to relate treatment effects on birth weight to treatment effects on survival. This is a very interesting technique and relatively new to us. However, as discussed at the end of these comments, we wonder to what extent the results, which are based on counterfactuals, really prove a causal relationship between these treatment effects.

The paper starts with a simple analysis based upon model (1) for $W_{it_i}^{obs}$ that is conditional upon covariates including the outcome Y_i^{obs}. Because models such as (1) do not distinguish between cause and effect, e.g., do not say whether $W_{it_i}^{obs}$ influences Y_i^{obs} or vice versa, we find them less satisfactory than hierarchical Bayesian models such as equation (4). The authors appear to be in agreement with us here.

A common approach taken in the measurement error literature is to develop a hierarchical model with three basic components. Using, for the moment, the notation of Carroll et al. (2006), the three parts of the model are

- an "exposure model" for the true values, $\boldsymbol{X}$, of the error-prone covariates conditional upon correctly measured covariates, $\boldsymbol{Z}$,
- a "measurement error model" for the surrogates, $\boldsymbol{W}$, given $(\boldsymbol{Z}, \boldsymbol{X})$, and
- an "outcome model" for the responses, $\boldsymbol{Y}$, given $(\boldsymbol{Z}, \boldsymbol{X})$, which is assumed to be

*School of Operations Research & Industrial Engineering,Cornell University, Ithaca, NY, mailto:dr24@cornell.edu

†Department of Statistics, Texas A&M University, College Station TX, mailto:carroll@stat.tamu.edu

 ba0005