Reply to the Discussion of "Estimating the Distribution of Dietary Consumption Patterns"

Raymond J. Carroll

The discussants' repeat remarks they raised in the review of the original paper upon which my article is based (Zhang et al., 2011b), so I will be brief. As part of their review we also produced extensive Supplementary Material:

1. In a time when it is routine to see models with thousands of latent variables and small sample sizes, it seems far-fetched at best to call ours "*highly complex*." We have 6 latent variables, a few regression parameters, two 19-dimensional covariance matrices and typically hundreds of observations, and thousands in our particular application. If that is "*highly complex*," what is being discussed in this issue must be worthy of international awards.

2. Our model is fully parametric, not "*semiparametric*."

3. "*How, without something like sensitivity analysis, are we to know that it is valid*?" As mentioned in my article, and the original paper, unlike massive latent variable problems, our model can be checked, because the simple bivariate submodels for every combination of dietary components can be fit by other means using standard weighted likelihoods, and we have confirmed that the bivariate sub-fits from our model agree with the direct fits to the submodels. See Section 4 of the *Annals of Applied Statistics* article where this issue is discussed in more detail.

4. We are sorry that the Bayesian sample survey methodology is a "*mess*." This is too bad, on many fronts. Our method would be fully Bayesian if there were no sampling weights needed.

5. We had an important, practical problem to be solved. The percentage of children with alarmingly bad diets is grossly overestimated by the standard single 24 hour recall. Should we have waited for the "*mess*" to be cleared up before solving it and being pure Bayesians in the process? I would love to do it as fully Bayesian, but one needs to remember that, as said in the article, it was Bayesian thinking that enabled the model to be fit to begin with, for example, the latent variables we use are a standard Bayesian formulation, MCMC computation, etc.

6. Section 3.3 of Zhang et al. (2011b) discusses weighting. We showed that the weights were not important for model fitting, and conjectured it was "*because the covariates we use are major players in determining the sampling weights*."

7. "*pseudolikelihood*": Yes, indeed, see the "*mess*" above, the abstract and Section 4 of my article, and Section 4 of the *Annals of Applied Statistics* paper.

An Offer: Understanding the distributions of usual intake in a population is important. Bayesian thinking allowed us to propose and fit our model, the methodology is fully Bayesian in nonsurvey problems, and is being used in such contexts to analyze the risk of usual dietary intake on disease.

I would be happy to work with any expert in Bayesian survey methodology to extend our work to be properly Bayesian in the context of NHANES, with proper Bayesian uncertainty statements. This problem is not going away, and regularly the analyses need to be updated.

Raymond J. Carroll is Jill and Stuart A. Harlin Chair and Distinguished Professor of Statistics, Nutrition and Toxicology, Department of Statistics, Texas A&M University, 3143 TAMU, College Station, Texas 77843-3143, USA (e-mail: carroll@stat.tamu.edu).