Representation of Uncertainty in Dose Reconstruction

Implications for Epidemiological Analysis

Daniel O. Stram

University of Southern California

BEIR 8 Planning Session/Beebe Symposium

Nov 17, 2014

Keck School of Medicine of USC

Background

- Dose reconstruction systems have become increasingly sophisticated
  - In estimation of dose
  - In representation of uncertainties where both
    - Independent errors in dose estimates, and
    - Shared errors in dose estimates

  need to be considered
- One approach to represent dosimetric uncertainty is by providing multiple realizations
  - Typically these realizations have a Bayesian interpretation
    - They are (attempts) to sample from the posterior distribution of true dose given "everything known" about dose determinants
    - Ideally they would represent both shared and independent uncertainties
  - They raise a natural question: how do these dosimetric uncertainties get translated into epidemiologic uncertainties?

Possible answers

- Modified single imputation
  - Use "best" estimate of dose in a single fit to estimate dose response parameter (a slope *b*)
  - Then use the multiple realizations to fix up the variance of the estimate of *b*
- Traditional Multiple Imputation (MI, Little and Rubin)
  - Fit the model to each realization of dose to estimate both *b* and the nominal *Var(b)*
  - Average the resulting estimates *b* and of *Var(b)* over the realizations
  - Add the observed variance of the estimates *b* to the average *Var(b)*.
- Likelihood averaging, Monte Carlo EM,

Complications with MI

- As described by Little & Rubin MI requires that missing data (here true dose) be sampled conditionally on all available information, including outcome of interest (e.g. cancer incidence)
- Failing to condition on the outcome of interest leads to serious biases towards the null in the MI estimates

    Resampling methods could potentially solve this problem (at least for simple problems)

Conversion to a Berkson error model

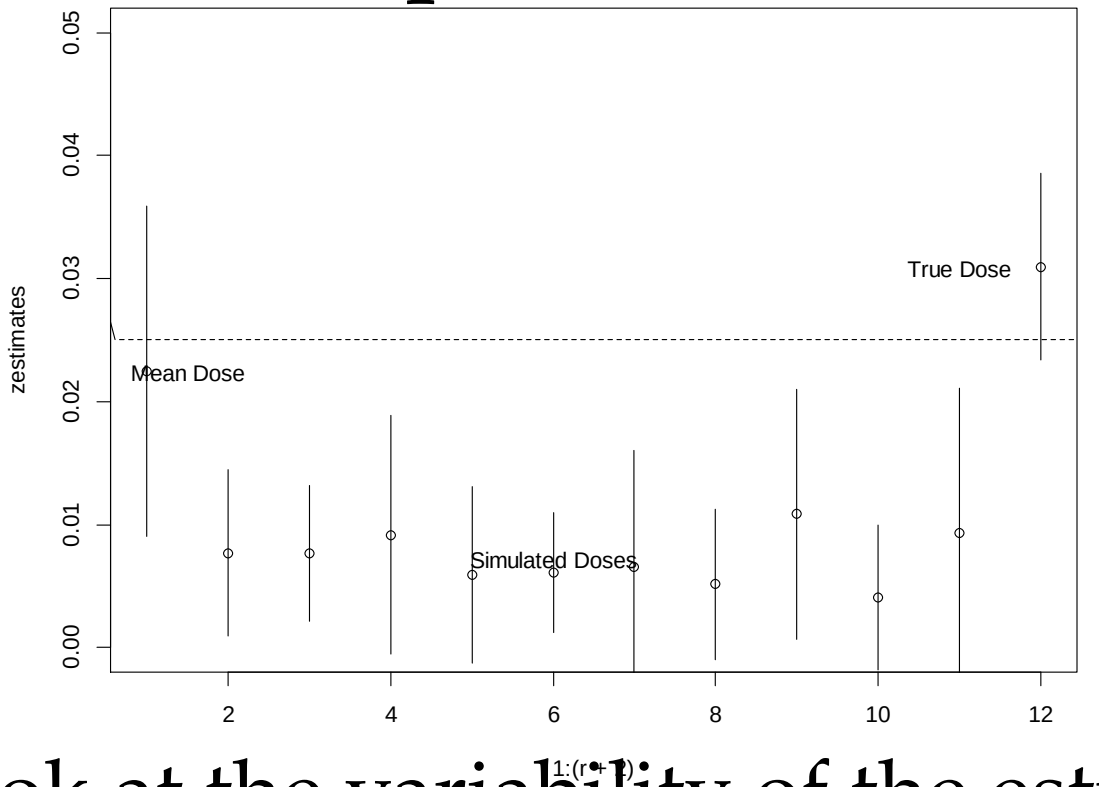
- Calculate the posterior mean **Z** by averaging over the realizations
  - Assume that enough realizations are available so that we can ignore error in estimating **Z**
- Then we have true X having mean Z, E(X|Z)=Z (the Berkson model) and for linear dose response model the estimate of *b* will be unbiased
- But the errors X-E(X|Z) are correlated which will increase (generally) the variance of the estimate of *b* over the nominal variance reported by the regression software used

Berkson Simulation Experiment

1. First choose mean dose **Z**
2. Generate true dose, $\mathbf{X_t}$, distributed around **Z** that has correlated multiplicative errors
3. Use this true dose and a (nearly) linear model to generate binary outcomes **Y**
4. Sample many realizations $\mathbf{X_r}$ from the same model that generated the true dose
5. Explore various approaches to using **Z** and $\mathbf{X_r}$ to estimate the dose response model used in (3)

A typical simulation, point estimates and nominal 95 percent CIs for slope *b* are shown

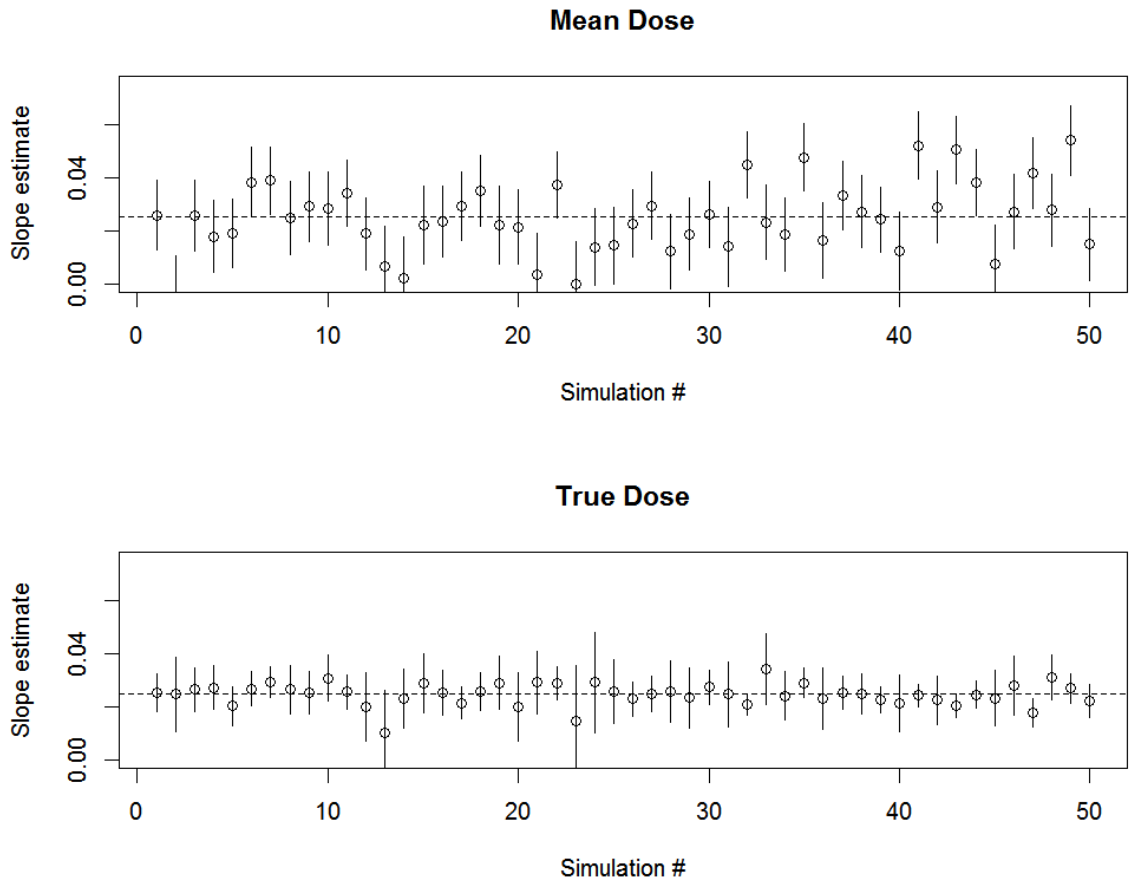
Clearly we don't want to use the average of the models fit to the simulated doses as our estimate of slope term b



- Next let's look at the variability of the estimate using the mean dose **Z**, over many simulations

- While nearly unbiased, the nominal confidence intervals for the estimate using the mean dose are not accurate

Only 70 percent coverage of true value



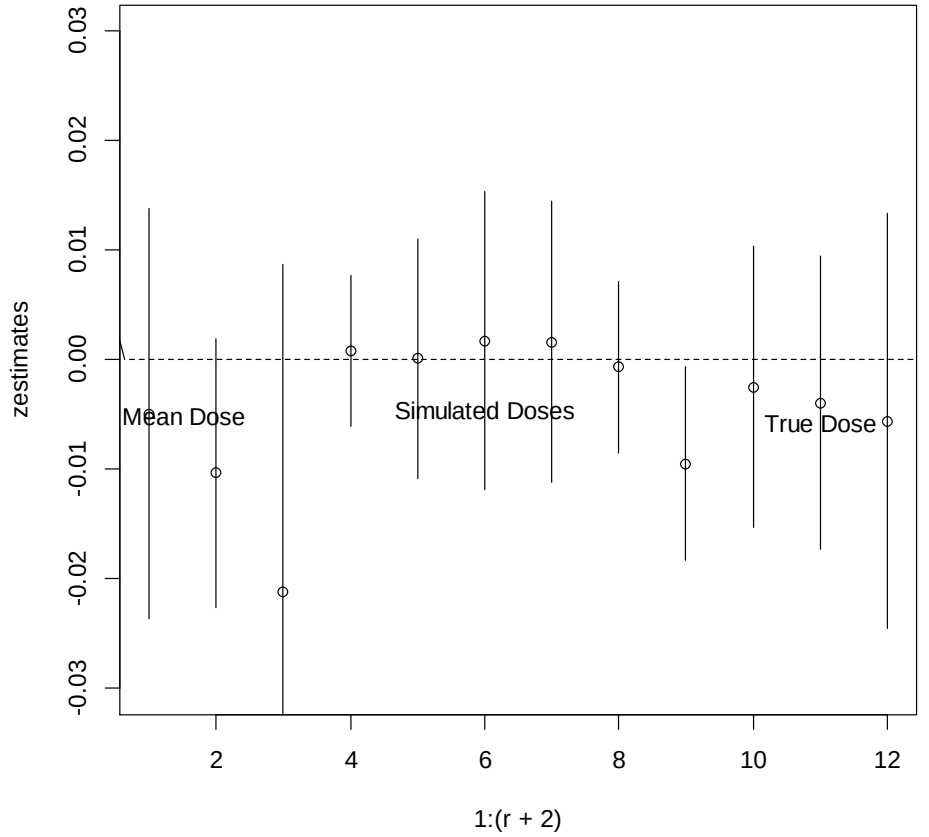
Possible solution

- Estimate the variance of the slope estimate over the replications and add this to the nominal variance
  - Seems to be an example of the MI method; however true multiple imputation requires conditioning on all data including Y
  - We know it is inadequate for another reason: **it's behavior under the null!**

Behavior of multiple imputation dose estimates under null hypothesis (slope=0)

Now the simulated doses provide unbiased slope estimates (of zero)

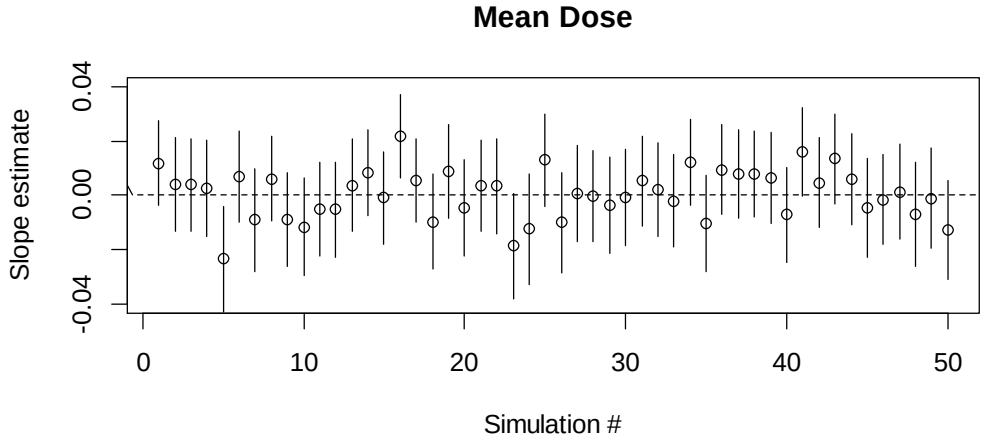
But note that the slope estimates are still pretty variable.



Do we need to add this extra variability to the variance estimate obtained using the mean dose?

Behavior of the mean dose estimates over many replications, but under the null hypothesis slope=0

Now the behavior of the confidence intervals on the mean dose estimate is fine coverage = 95 percent



Evidently we don't need to add any additional variability to the estimator under the null.

This feature is important … (see below)

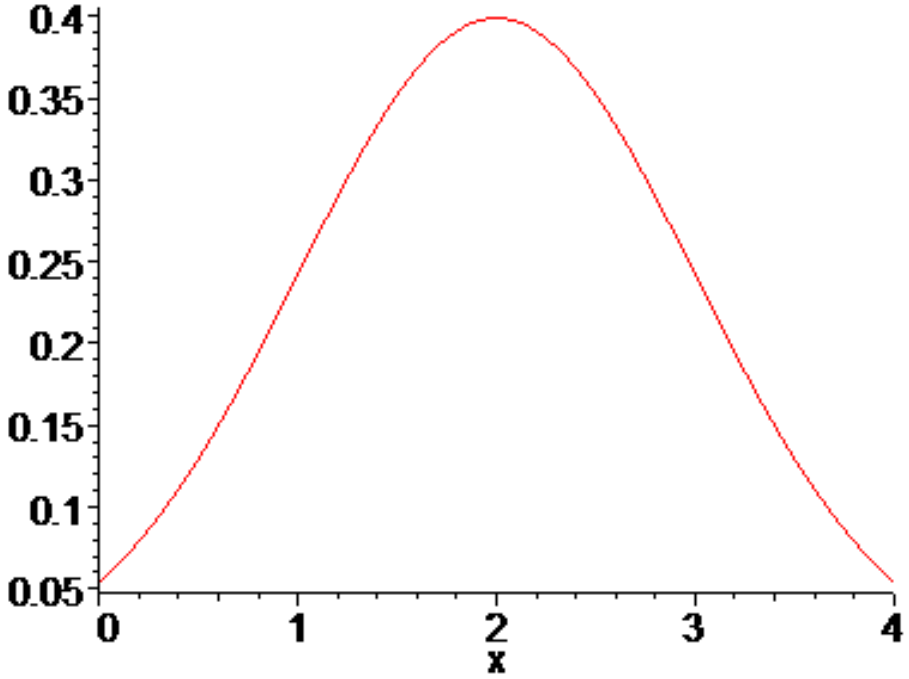
What about a better multiple imputation approach?

- The Rubin and Little MI procedure would say that we need to draw samples of the true doses from its distribution CONDITIONAL on Y as well
- Proposal: Use a likelihood-based resampling scheme to mimic this procedure.

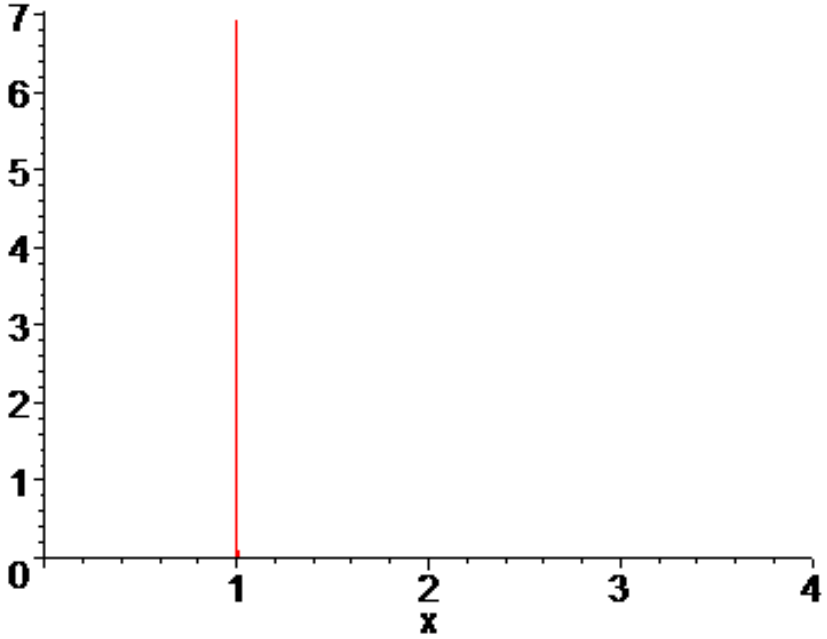
When should this work? (i.e. provide samples from the distribution of interest)

- When we can sample from the prior and the likelihood is not too extreme. For example this will work fine for univariate f(x) with shape

Like this



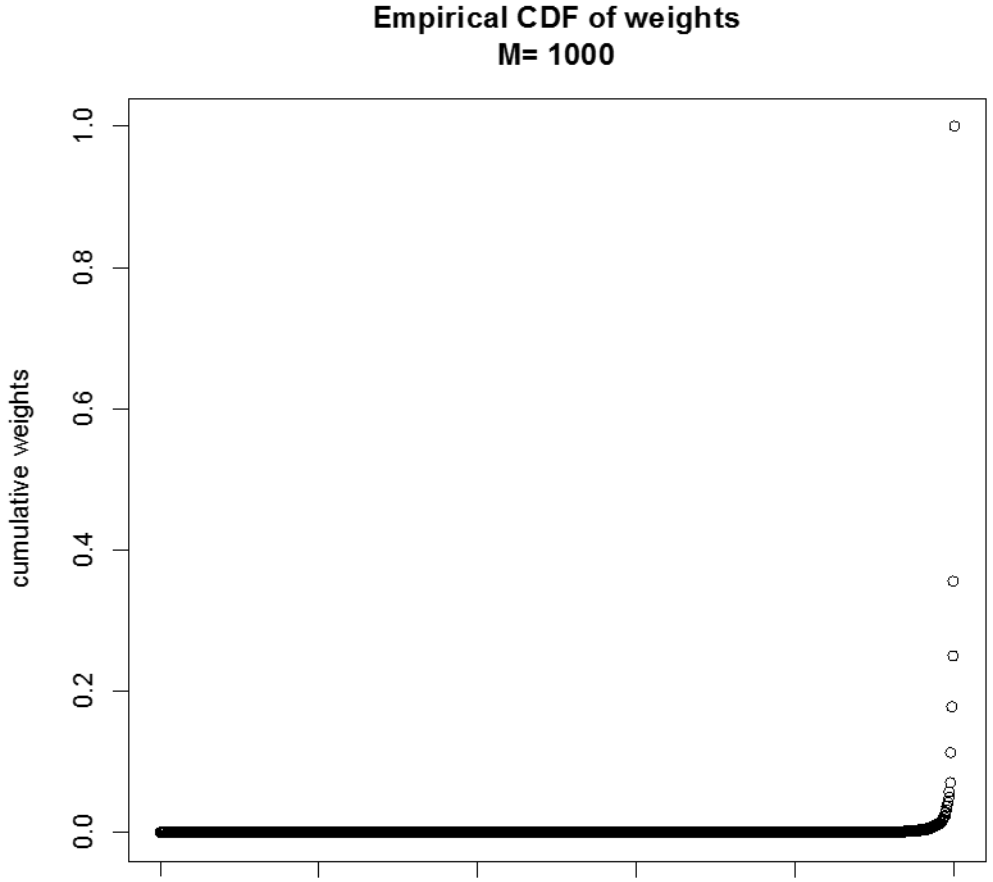
But not like this!



What happens in the previous simulation?

- Computing likelihood based weights in this fashion for N=1,000 individuals and M=1,000 replications

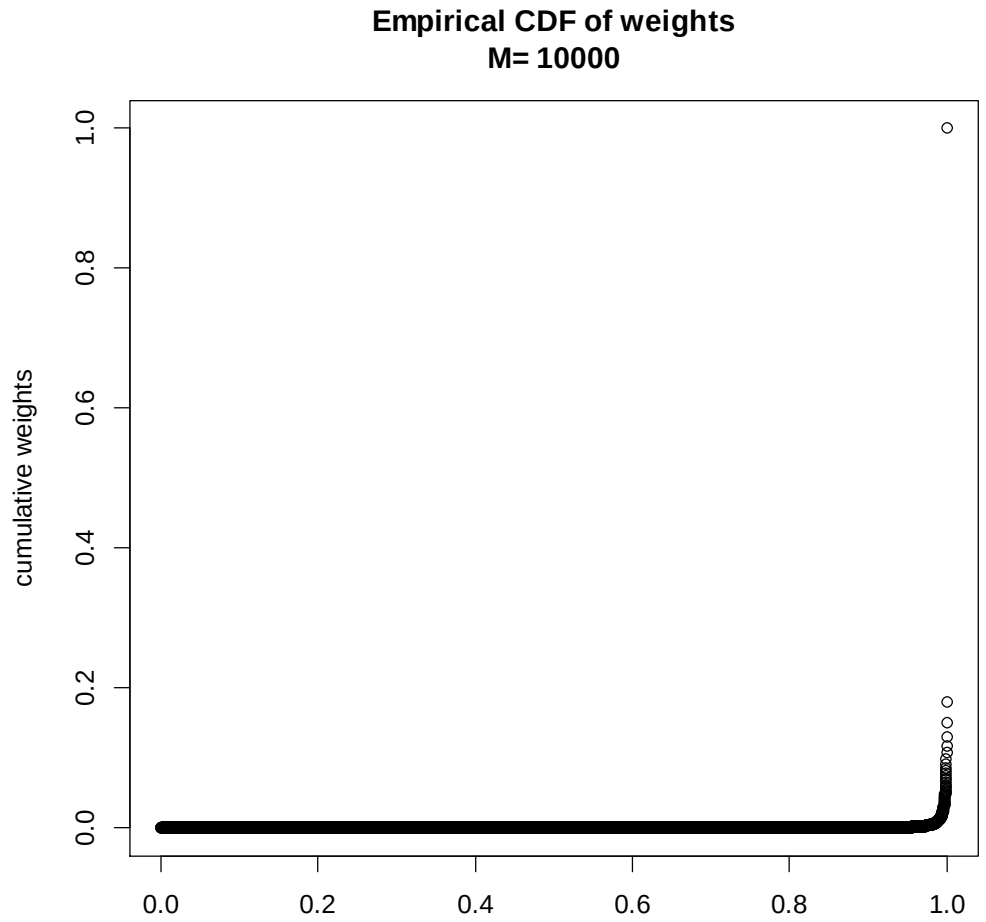
64% of the probability is concentrated on a single replication.



Would seem to imply very little correction needed

Effect of increasing number of replications by factor of 10 – the problem seems to be getting even worse!

80 percent of probability assigned to a single realization



Curse of dimensionality

With a strong dose response the likelihood becomes more and more concentrated into a small region of the N-dimensional space as N increases.
The exploration of this space is inadequate even with LARGE numbers of replications

Modified Single imputation

- For certain models we can directly adjust the estimated variance of parameter estimates. These models include
  - Normal linear regression
  - Poisson linear regression – which forms the basis for much analysis in radiation epidemiology including survival analysis

For these models

$$Var(\hat{\beta}) = I_w^{-1} + b^2 I_w^{-1} M^T Var(X \mid Z) M I_w^{-1}$$

- Where I_w is the usual information matrix
- Adjustment term depends upon b (and drops out under the null) as well as $Var(X|Z)$ **which can be estimated from the realizations**
- $\boldsymbol{M}$ is a matrix of known form (function of covariates and parameters)

Error Correction Method

- Calculate the mean dose **Z** and the variance matrix $Var(X|Z)$ by averaging the realizations
- Perform the usual analyses using **Z** in place of **X**
- When happy with the model add

$$b^2 I_w^{-1} M^T Var(X \mid Z) M I_w^{-1}$$

to the variance of the parameter estimates

Some other implications

- Testing for non-zero dose response can be done "as usual" (ignoring shared errors) since the term

$$b^2 I_w^{-1} M^T Var(X \mid Z) M I_w^{-1}$$

  disappears under the null
- The standard errors are most sensitive to highly shared multiplicative uncertainties
- This approach can be extended naturally to survival data and prolonged exposures (as in the Mayak Workers Study)

Other models (e.g. those strongly nonlinear in dose) require other methods

- Likelihood averaging as described by Stayner et al (2007)
  - Can be considered in simple cases, but for strong dose response model, the variation of likelihood contributions are extreme
- Monte-Carlo EM algorithm
  - Again requires sampling from conditional distribution given outcome of interest

Acknowledgements

- Funding
  - Mayak Workers Study (my work funded by DOE Russian Health Studies Program, Program Officer Barrett Fountos)
  - Atomic Veterans Study/Million Workers (NCI/DOE)
- People
  - Dale Preston, Bruce Napier, Ken Kopecky, Mikhail Sokolnikov, John Boice, Harold Beck, John Till, Andre Bouville, Duncan Thomas