

## Outline for Discussion at NCES Roundtable on Imputation

1. Brief discussion of Bamberg (2011) presentation
  - Types of applications for multiple imputation
    - o Traditional (will come back to NHANES DXA imputation later)
    - o Bridging (and other types of combining information)
      - Note uncongeniality issues reported in Rubin and Schenker (1987, *JOS*)
    - o Measurement error
  - Topics for future research
    - o Flexible models and methods
    - o Diagnostics for imputation models
    - o “Portability” of bridging models when the two surveys have different contexts
    - o “Uncongeniality” between imputation model and analysis model
    - o Methods for reflecting complex sample designs in imputation models
2. “Hot-deck imputation versus multiple imputation”
  - Not really the issue, because multiple hot-deck imputation possible
    - o To reflect variability more fully, draw bootstrap sample from complete data before creating each set of imputations
      - Rubin and Schenker (1986, *JASA*; 1991, *Statistics in Medicine*))
  - Two big issues
    - o Single imputation versus multiple imputation
    - o Hot-deck versus explicit-model-based imputation
  - Hot-deck versus explicit-model-based imputation
    - o Hot deck
      - Imputes values that have actually occurred
      - Less parametric flavor => possible robustness
        - See Schenker and Taylor (1996, *Computational Statistics and Data Analysis*)
    - o Explicit-model-based
      - Easier to explain the model
      - Handles general patterns of missing data better
      - Can include more variables as predictors (e.g., by omitting high-order interactions)
        - Can improve prediction and make missingness at random more plausible
3. Some issues of interest for NHES imputation
  - Single imputation versus multiple imputation
    - o So far, differences in variance estimates not major (note low item nonresponse rates)
    - o See if there are classes of analyses for which differences are larger
  - Possible advantages of explicit-model-based imputation over hot-deck imputation
    - o Handles general patterns of missing data better
      - Predictors (analogous to “boundary variables”) can have missingness
      - Note that “random imputation” (used for “boundary variables”) probably okay for marginal distributions, but may attenuate multivariate analyses
    - o Can include more variables as predictors
      - Could reduce bias and decrease variance
      - No need to worry about number of donors in cells
      - Note that there is a bias/variance trade-off associated with number of donors, collapsing cells etc. (see Schenker and Taylor 1996 for some relevant work)
  - Effects of manual imputation and post-imputation edits
    - o Any attenuation of the positive effects of the prior imputation?