

Outline for Discussion at NCES Roundtable on Imputation

Nathaniel Schenker
January 26, 2018

1. Brief Discussion of Bamberg (2011) presentation

Nathaniel Schenker
January 26, 2018

➤ **Types of Applications for Multiple Imputations**

- Traditional (will types of combining information)
- Note uncongeniality issues come back to NHANES DXA imputation later)
- Bridging (and other types of combining information)
 - Note congeniality issues reported in Rubin & Schenker (1987, *JOS*)
- Measurement error

➤ **Topics for Future Research**

- Flexible models and methods
- Diagnostics for imputation models
- “Portability” of bridging models when the two surveys have different contexts
- “Uncongeniality” between imputation model and analysis model
- Methods for reflecting complex sample designs in imputation models

2. Hot-Deck Imputation vs Multiple Imputation

Nathaniel Schenker
January 26, 2018

➤ **Not Really the Issue, because Multiple Hot-Deck Imputation Possible**

- To reflect variability more fully, draw bootstrap sample from complete data before creating each set of imputations
 - Rubin & Schenker (198, *JASA*; 1991, *Statistics in Medicine*)

➤ **Two Big Issues**

- Single imputation vs multiple imputation
- Hot-Deck vs Explicit-Model-Based imputation

2. Hot-Deck Imputation vs Multiple Imputation

Nathaniel Schenker
January 26, 2018

➤ **Hot-Deck vs Explicit-Model-Based imputation**

Hot-Deck

- Imputes values that have actually occurred
- Less parametric flavor +> possible robustness
 - See Schenker & Taylor (1996, *Computational Statistics & Data Analysis*)

➤ **Explicit-Model-Based**

- Easier to explain 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

Nathaniel Schenker
January 26, 2018

➤ **Single Imputation vs Multiple Imputation**

- So far, differences in variance estimates not major (note low item nonresponse rates)
- See if there are classes of analyses for which differences are larger

➤ **Possible Advantages of Explicit-Model-Based Imputation Over Hot-Deck Imputation**

Handles general patterns of missing data better

- Predictors (analogous to “boundary variables”) can have missingness
- Note that “random imputation” (used for “boundary variables”) probably ok for marginal distributions, but may attenuate multivariate analyses

3. Some Issues of Interest for NHES Imputation

Nathaniel Schenker
January 26, 2018

➤ **And 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 & Taylor 1996 for some relevant work)

➤ **Effects of Manual Imputation and Post-Imputation Edits**

- Any attenuation of the positive effects of the prior imputation?