2-26-2019

Designing Studies for Comparing Interviewer Variance Components in Two Groups of Survey Interviewers

Brady T. West
University of Michigan, Ann Arbor, bwest@umich.edu

Designing Studies for Comparing Interviewer Variance Components in Two Groups of Survey Interviewers

Brady T. West

Survey Research Center

Institute for Social Research

University of Michigan-Ann Arbor

bwest@umich.edu

Example of Motivating Research Question

- Standardized interviewing (*SI*) is widely used to ensure consistent administration of survey content and believed to minimize interviewer effects
- A body of literature exists indicating that conversational interviewing (*CI*), designed to ensure respondent comprehension, can decrease response bias (e.g., Conrad and Schober, 2000, POQ); but critics wonder about an…
- **Open Question:** Does CI produce higher interviewer variance in survey responses than SI?
  - Uneven implementation, variance in wording, etc. may introduce more variance in responses across interviewers

Study Design: **Key Points**

- Original FTF data collection in 15 large geographic areas in Germany
- Simple random samples of 480 currently-employed adults drawn from each of the 15 areas (**geographic representation**)
  - Adults had history of at least one unemployment spell
  - Samples drawn from government database (IEB) of official employment histories in each area (**possible validation data**)
  - *n* = 7,200 in full sample; **multiple (4) interviewers per area**
- **60 Interviewers each assigned 120 cases at random**
  - Interpenetrated design, **after conditioning on area effects**

Study Design: **Key Points**

- **Two interviewers in each area were rigorously trained in CI, and the other two were rigorously trained in SI (two groups, assignment not confounded with area)**
- Data Collection Period: April 2014 - October 2014
- Interviewers administered a 30-minute CAPI instrument
- The instrument included questions that we judged to require complex response processes, related to housing conditions, employment histories, and social networks
- **Many questions were explicitly constructed to enable response validation using data on the IEB frame**

Study Design: **Power Analysis**

- Need to power study to be able to detect realistic differences in interviewer variance components between two independent groups of survey interviewers (**in a multilevel model**); more on this soon!
- No "canned" software for this task: **need simulation**
- See the SAS macro at: https://github.com/bradytwest/SimStudiesSAS/blob/master/var_comp_power.sas
- The macro accepts expected differences, desired counts of interviewers in each group, and respondents per interviewer, and then empirically simulates power for normal or binary outcomes
- Needed 1,800 respondents total for this study (about 30 per interviewer)

Analytic Approaches

- **Multilevel linear, logistic, and ordinal models for each survey variable**, with fixed effects of the CI technique and 14 of the 15 areas (**necessary control!**), and **random interviewer effects**
- Models allow the interviewer and residual variance components (for continuous items) to vary for the two groups; for example (*i* = interviewer, *j* = respondent):

Note that the interviewer and residual variance components for the two groups are allowed to vary!

$$y_{ij} = \beta_0 \ldots REA_i = p] + u_{1i} I[CI_i = 1] + u_{2i} I[SI_i = 1] + \varepsilon_{ij}$$

$$u_{1i} \sim N(0, \tau^2_{CI}),\ u_{2i} \sim N(0, \tau^2_{SI}),\ \varepsilon_{ij} \sim N(0, \sigma^2_{CI}) \text{ if } CI_i = 1,\ \varepsilon_{ij} \sim N(0, \sigma^2_{SI}) \text{ if } SI_i = 1$$

- Differences in variance components tested using frequentist (LRT) or Bayesian methods outlined by West and Elliott (2014, *Survey Methodology*)

Frequentist Approach: LRTs

- Classical likelihood ratio test of constrained null hypothesis that two variance components are equal (**easy!**)
- **Limitations:**
  - Likelihood ratio tests rely on asymptotic theory: generally small samples of interviewers!
  - Likelihood ratio tests are not appropriate when using pseudo-likelihood methods
  - No accounting for uncertainty in estimating features of prior distributions for parameters
  - Negative estimates of variance components possible
  - Not possible to compute a confidence interval for the difference in the variance components

Bayesian Approach

- Specify **prior distributions** for parameters of interest:

$$\beta_0 \sim N(0,100)$$
$$\beta_1 \sim N(0,100)$$
$$\tau_1^2 \sim Uniform(0,10)$$
$$\tau_2^2 \sim Uniform(0,10)$$
$$\sigma_\varepsilon^2 \sim Uniform(0,10)$$

- Proper, diffuse, and noninformative, as recommended by Gelman (2006) for multilevel models

Bayesian Approach, cont'd

- Uses Gibbs Sampler with Adaptive Rejection Sampling methodology (as implemented in BUGS) to simulate draws from joint posterior distribution of parameters in model; could use Stan / brms / etc.
- Inferences about difference in variance components based on posterior distribution of **differences in draws of variance components**, denoted by $\tau_1^{2(d)} - \tau_2^{2(d)}$
- 2,500 burn-in draws, 3 Markov chains using random normal and uniform draws to start

Advantages of Bayesian Approach

- More appropriate for small samples of clusters (interviewers in this context)
- Does not rely on asymptotic theory for inferences
- Enables computation of posterior credible sets for differences in variances with natural interpretation
- Accounts for uncertainty in estimation of parameters of prior distributions

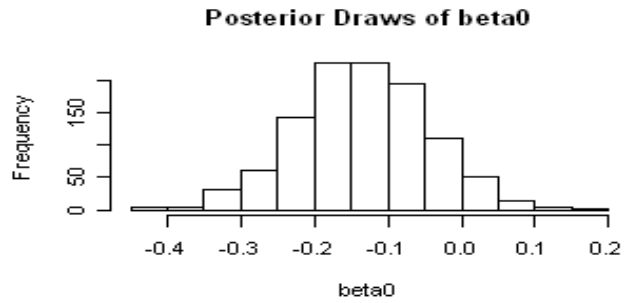
Example: Married and Non-Married Interviewers in the National Survey of Family Growth (West and Elliott, 2014)

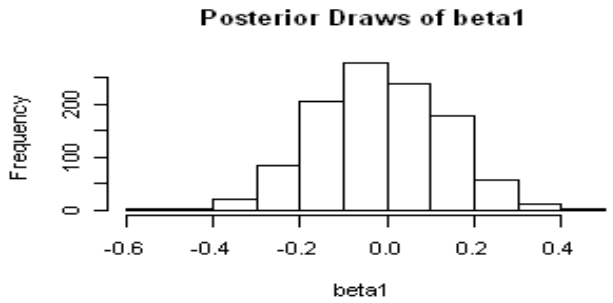
- No fixed effect of marital status on expected value of **parity**; evidence of overdispersion
- Estimated variance components for parity reports (SE / PSD):

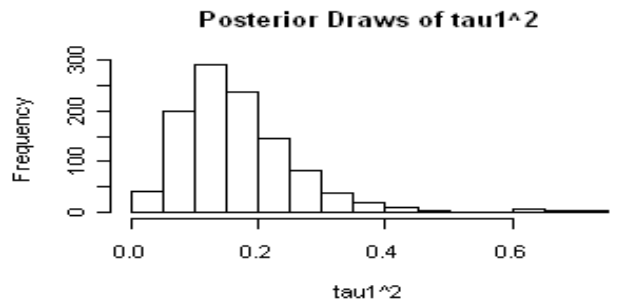
  Frequentist → M = 0.126 (0.060), NM = 0.003 (0.024)

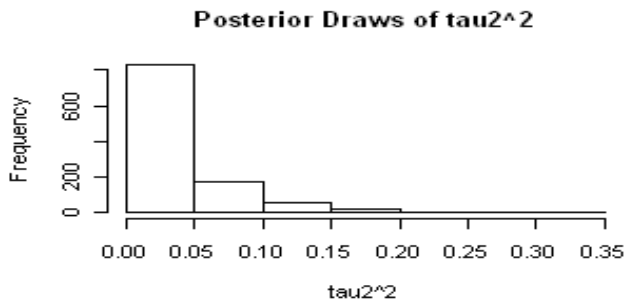
  Bayes → M = 0.151 (**0.092**), NM = 0.023 (**0.040**)
- LRT of equality of variance components for married and non-married interviewers: $p = 0.041$
- Bayesian 95% credible set: (-0.029, 0.360)
- **Marginal evidence of a difference…examine plots!**

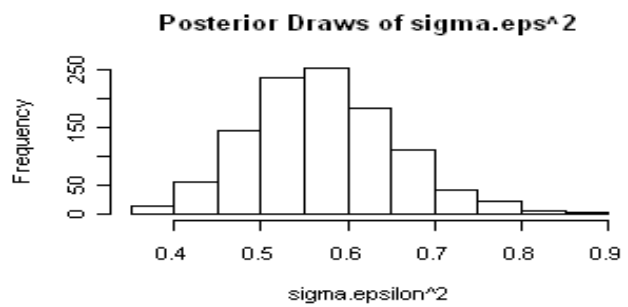
Posterior Simulations

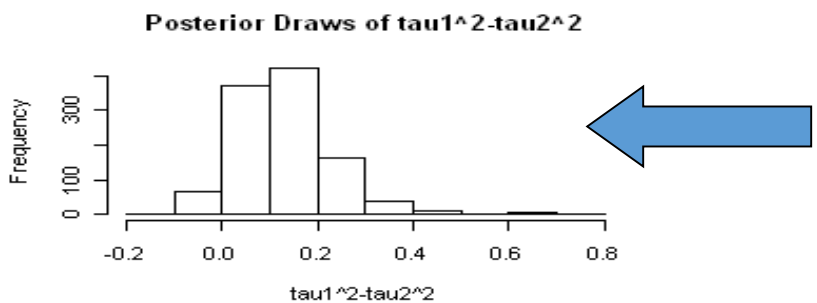








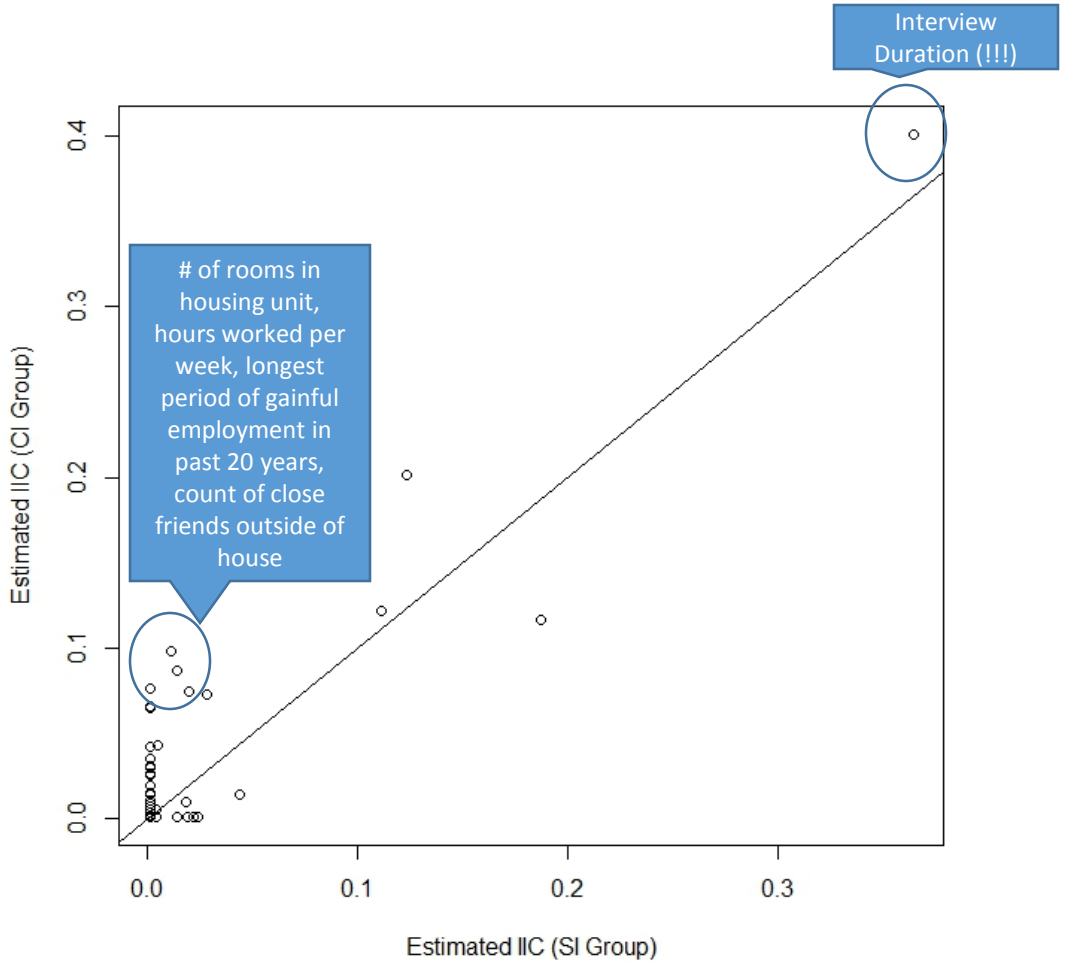




Back to the Motivating Example: Results / Interpretation

Does CI increase the *influence* of *Is*?

West, Conrad, Kreuter & Mittereder (*2018, JRSS-A*)



Does CI increase the *influence* of *I*s?

- Not much, if at all:
  - **Significant** increases in variance components due to the use of CI are rare (5/55 items)
  - When they occur, improved accuracy due to CI more than offsets them, resulting in smaller MSEs
- CI improved quality of reporting relative to SI, consistent with previous findings, *without notably increasing interviewer effects*

A Total Survey Error Perspective

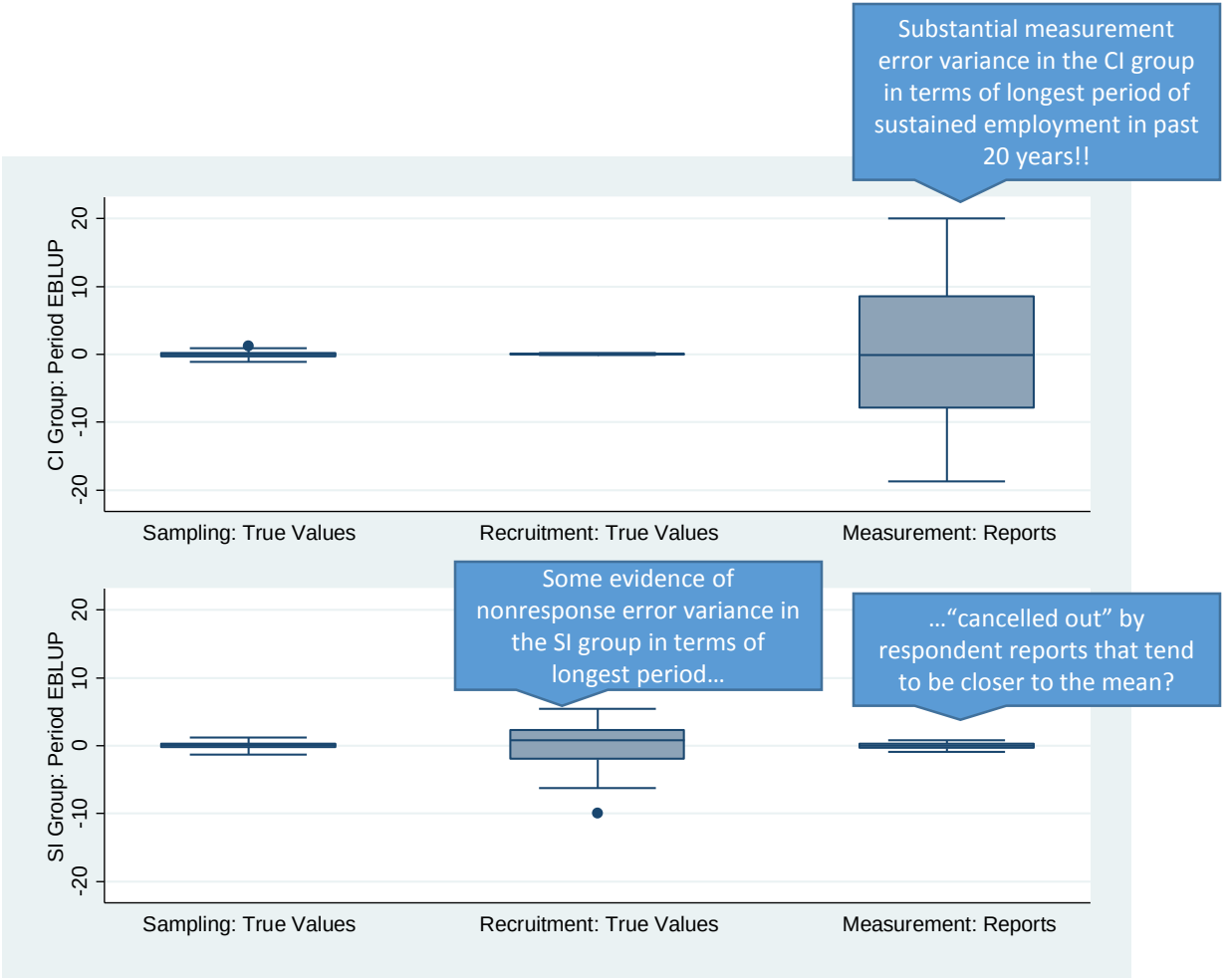
- Recent work (West and Olson, 2010, POQ; West et al., 2013, JOS) has attempted to decompose interviewer variance into sampling error variance, nonresponse error variance, and measurement error variance
- What do these decompositions look like for conversational and standardized interviewing?
- Consider results from the same study in Germany (West et al., 2018, *JSSAM*): compare interviewer variance at each stage
- Focus on 3 items in particular, with: a) admin data available from the IEB database, and b) substantial interviewer variance based on respondent reports

Respondent Age

Longest Period of Sustained Employment in Past 20 Years

CI Group

SI Group



Conclusions

- Survey managers cannot ignore the possibility of nonresponse error variance among interviewers on key correlates of survey measures of interest (e.g., age); **should be monitored "live"**
- SI is not entirely free from significant measurement error variance; should also be monitored in a "live" fashion (e.g., ongoing computation of EBLUPs)
- CI can introduce substantial increases in measurement error variance; uneven implementation? Additional re-training?
- *Careful design can lead to interesting comparative studies!*
- **Papers mentioned are all available upon request!**