New Directions for Extending Research on Interviewer Effects in Surveys

GESIS Seminar:
June 12, 2017

Brady T. West, PhD
Survey Research Center
Institute for Social Research
University of Michigan – Ann Arbor
bwest@umich.edu

Seminar Objectives

- Provide an overview of recently published, ongoing, and needed future research on the effects that interviewers can have on the quality of both survey data and survey paradata (and related survey estimates)
  - Good list of possible dissertation topics (red)… =)
- Provide initial answers to **four research questions** that have received only minimal attention in the literature to date

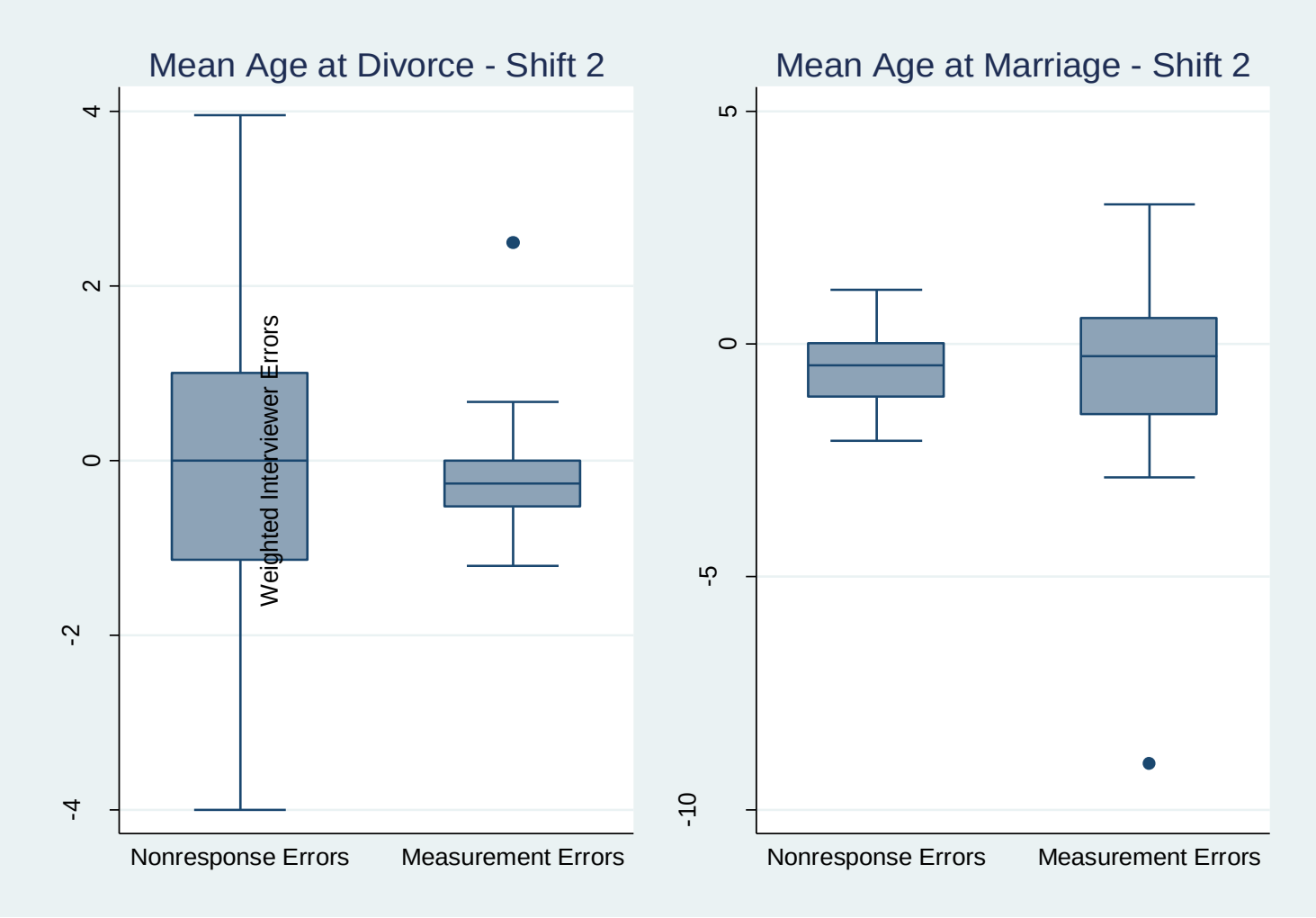
Research Question #1

- What is the relative decomposition of interviewer variance in a variety of survey variables in terms of sampling error variance, nonresponse error variance, and measurement error variance among interviewers?
  - And how can we study these decompositions in the absence of **interpenetrated** sample designs?

West and Olson (2010)

- Analyzed data from a **telephone** survey where true values were available on a sampling frame (official divorce records)
  - **Question:** To which interviewer should a telephone survey nonrespondent be assigned?
  - Interpenetrated assignment within calling shifts
- Found evidence of significant total interviewer variance for selected variables being driven by nonresponse error variance among IWERs

West and Olson (2010), cont'd



West, Kreuter, and Jaenichen (2013)

- Replication of West and Olson (2010), focusing on a **face-to-face** economic panel survey in Germany (PASS)
  - Larger interviewer effects expected
  - Interpenetrated assignment not feasible / realistic; models adjusted for PSU-level covariates
  - This adjustment was not a panacea, and **is a necessary area for future research**

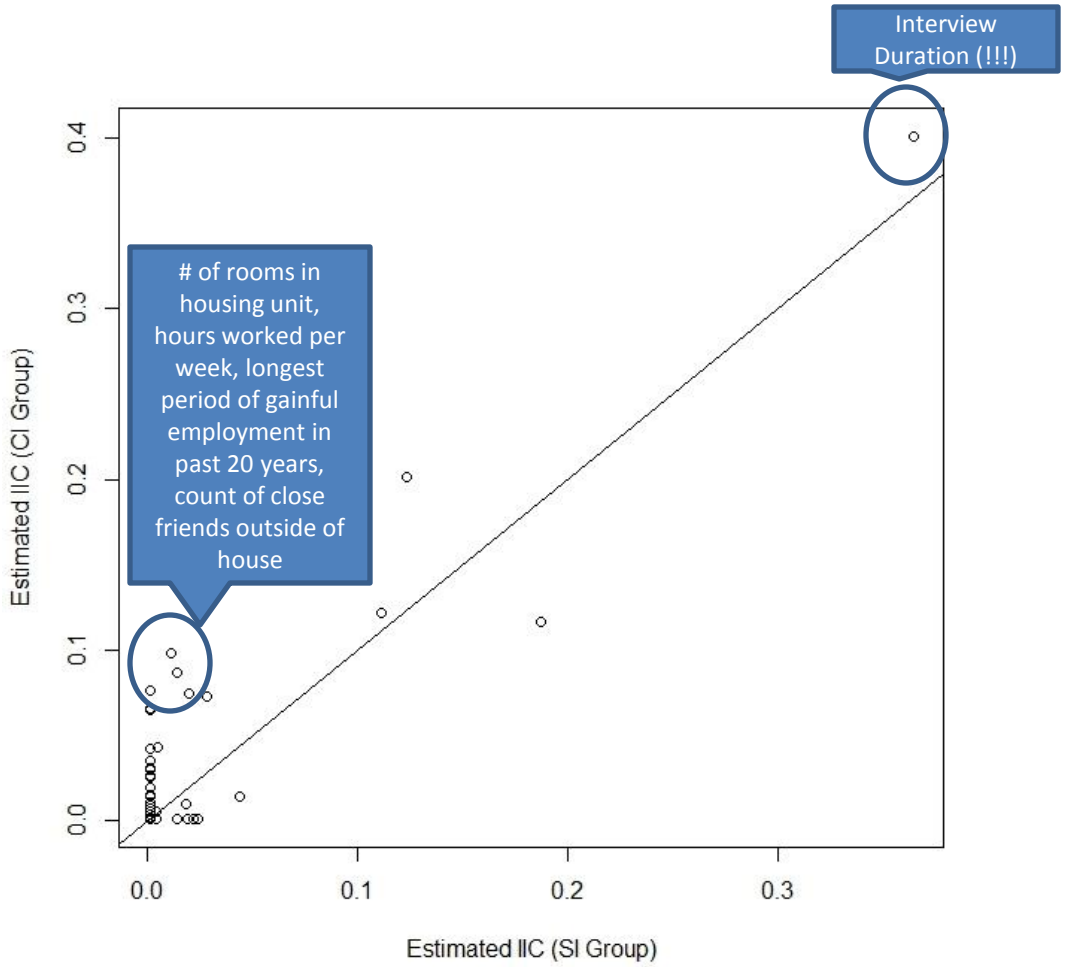
West, Kreuter, and Jaenichen (2013)

- Out of 11 key survey variables, 5 were found to have significant total interviewer variance
  - Variance in age reports was driven by nonresponse error variance
  - Variance in benefit receipt reports was driven by a mixture of nonresponse error and measurement error variance
  - Other three variances were driven primarily by measurement error variance

West, Conrad, Kreuter, and Mittereder (2017a; 2017b R&R)

- NSF-supported, face-to-face survey experiment in Germany examining the decomposition question for **conversational interviewing (CI)** and **standardized interviewing (SI)**; *n* = 1,850, 30 interviewers per group (60 total)
- Makes use of truly interpenetrated sample assignments in each group, in addition to official record information
- **GAP 1:** Does the CI technique introduce more interviewer variance? (Use testing approaches of West and Elliott, 2014)
- **Answers:** CI tends to increase interviewer variance, but only slightly in most cases (5 out of 55 survey / paradata variables analyzed had significantly higher interviewer variance for CI)

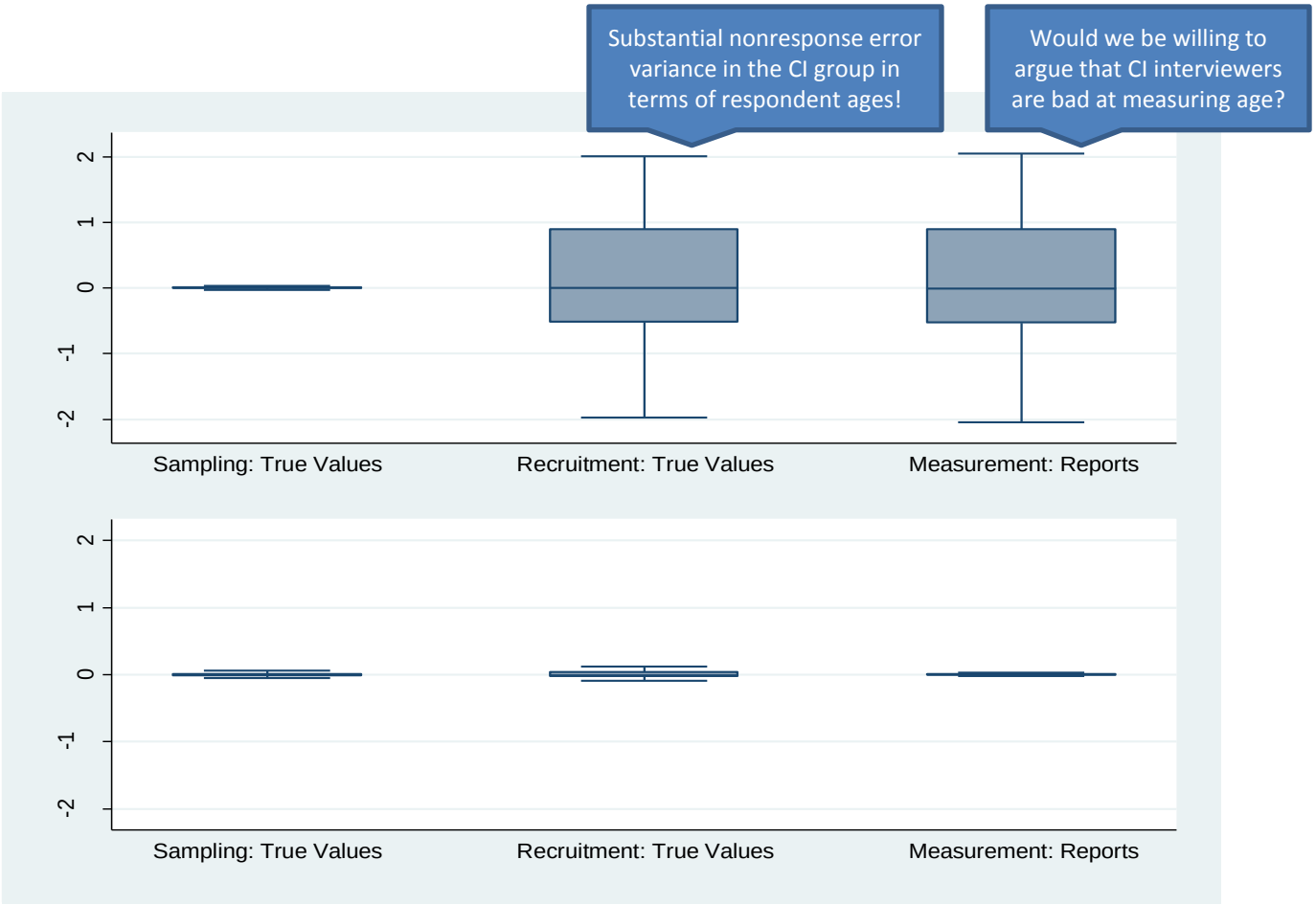
West, Conrad, Kreuter, and Mittereder (2017a; 2017b R&R)



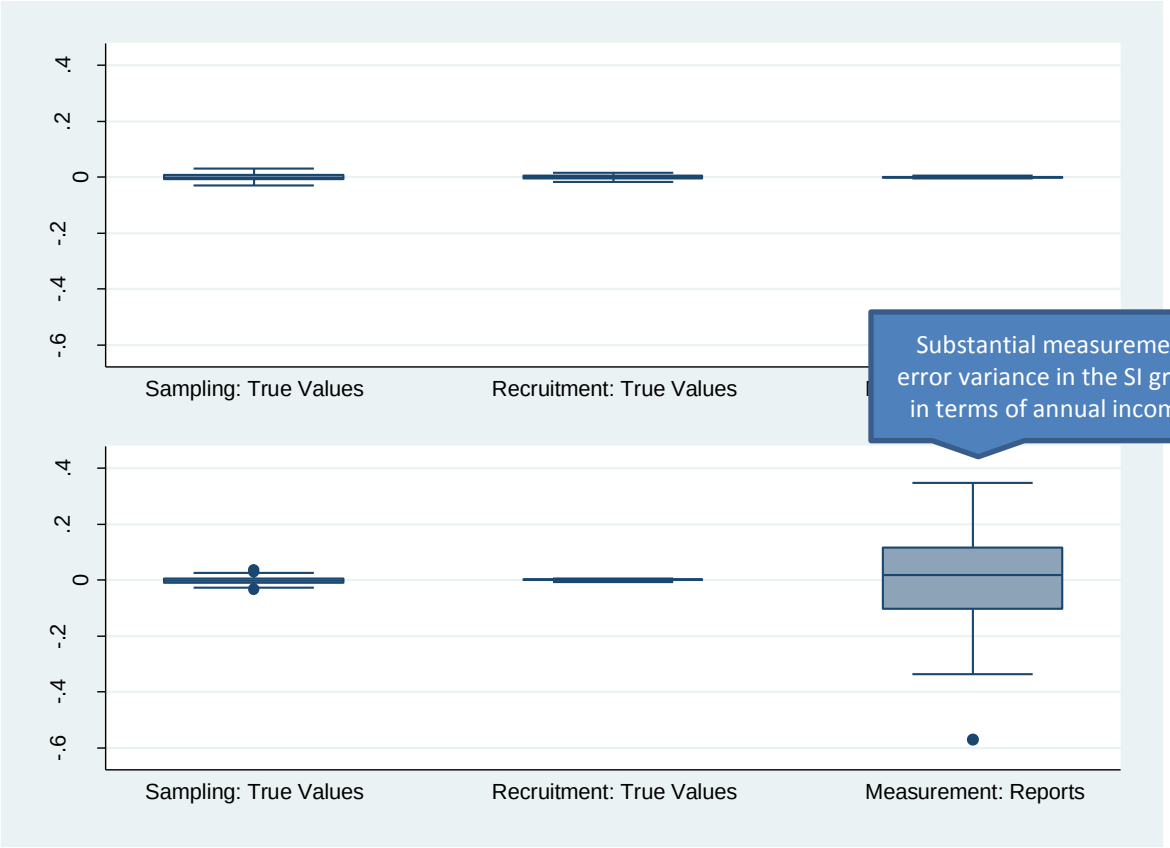
West, Conrad, Kreuter, and Mittereder (2017a; 2017b R&R)

- CI was also found to shift response distributions significantly and in the direction of higher response quality for 14 of the 55 variables
- Based on items with administrative data available, the increase in interviewer variance from CI was not large enough to offset gains in estimated MSE
- **GAP 2:** What about the decomposition of interviewer variance for each of the two groups?
- Focus on 3 items in particular with substantial interviewer variance based on respondent reports

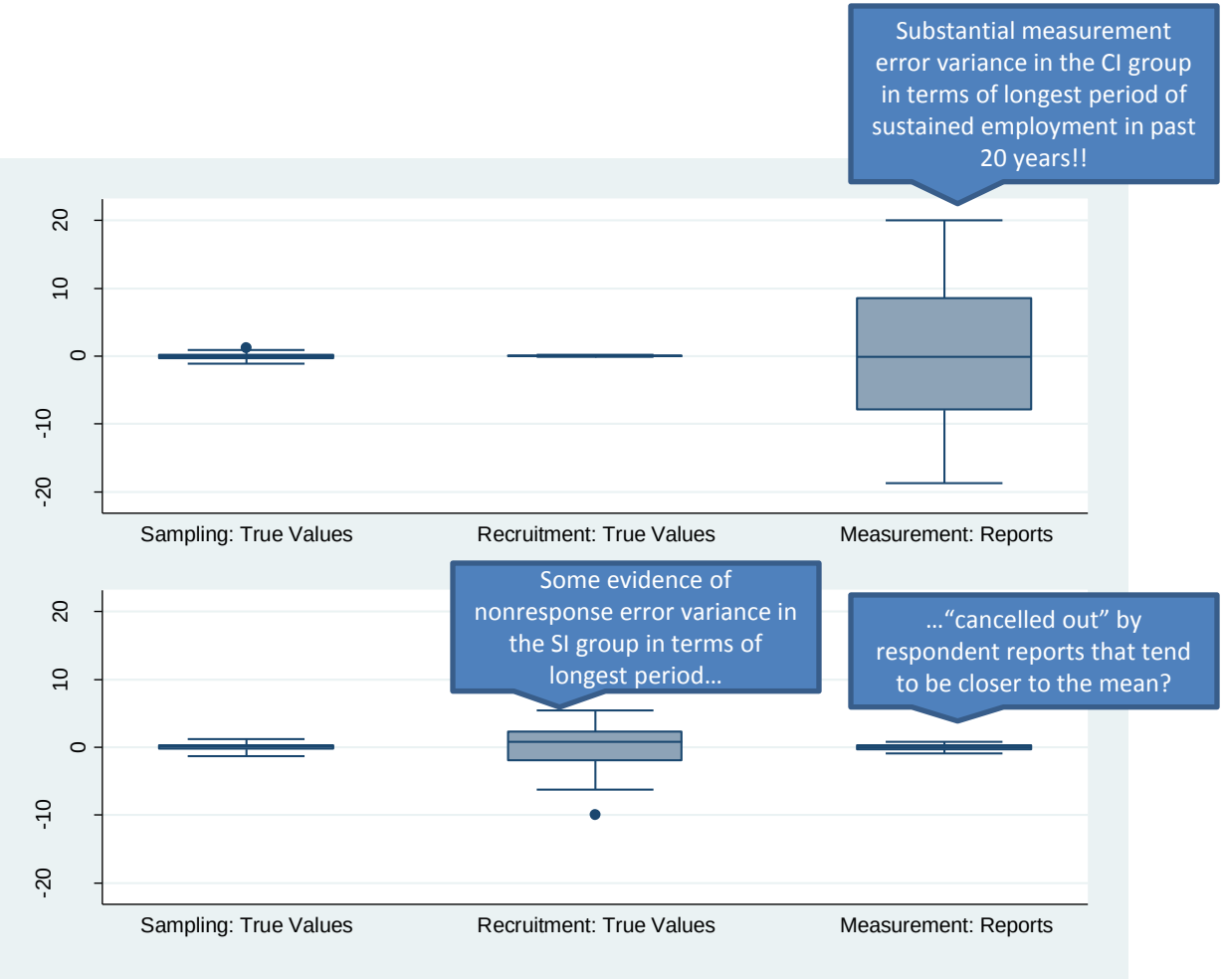
West, Conrad, Kreuter, and Mittereder (2017b R&R)



West, Conrad, Kreuter, and Mittereder (2017b R&R)



West, Conrad, Kreuter, and Mittereder (2017b R&R)



West, Conrad, Kreuter, and Mittereder (2017b R&R)

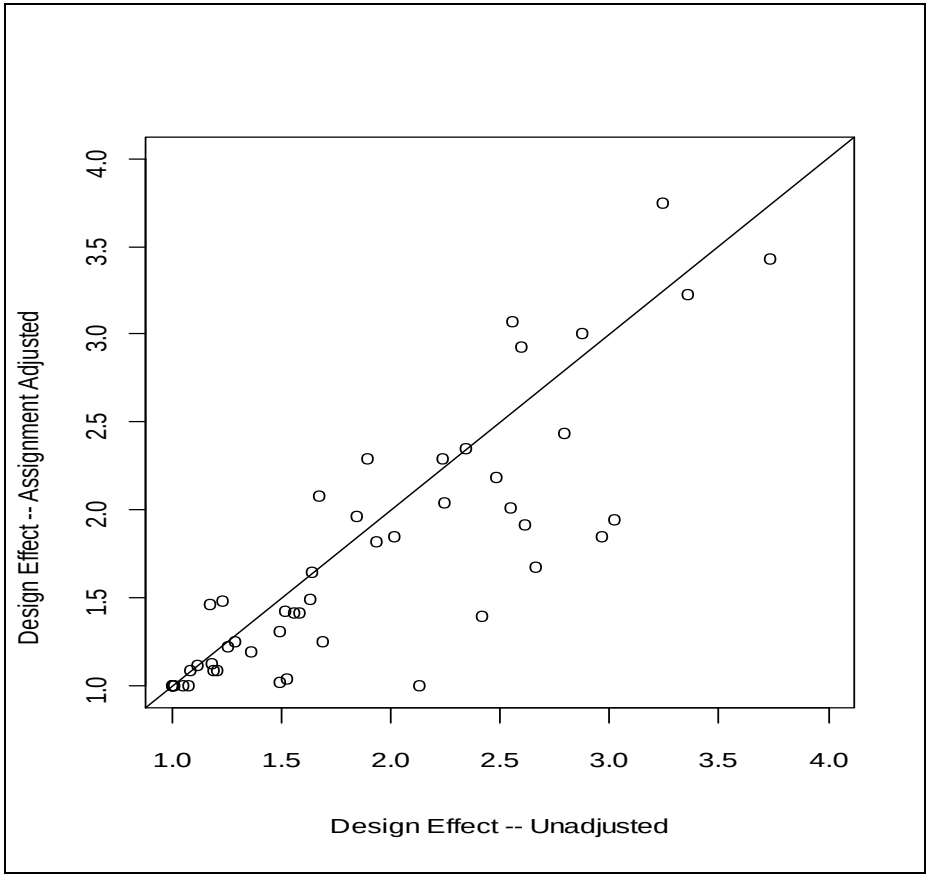
- **Conclusions / Current Thinking:**
  - Survey managers cannot ignore the possibility of nonresponse error variance among interviewers on key correlates of survey measures of interest (e.g., age); should be monitored "live"
  - SI is not entirely free from significant measurement error variance
  - CI can introduce substantial increases in measurement error variance; uneven implementation? Additional re-training?

Future Research Directions: RQ #1

- **How do we handle lack of interpenetration?**
- Current work with Mike Elliott on two ideas:
  - **"Anchoring"**, or adjusting estimates of interviewer variance for *within-subject* correlations of residuals on an *anchor* known to be free from interviewer measurement error (e.g., age) and another known to be correlated with the "anchor" but prone to interviewer effects (e.g., health)
  - **"Assignment Propensity Adjustment"**, or a weighting approach designed to adjust survey estimates for the probability of being assigned to a particular interviewer based on known covariates free from interviewer error

Anchoring Illustration: BRFSS

- Several decreases across states in estimates of multiplicative interviewer effects on variances of mean self-reported health from “anchoring” on age



Future Research Directions: RQ #1

- **How do we efficiently model measurement error variance and nonresponse error variance among interviewers simultaneously?**
  - **Idea:** First, impute responses for non-respondents using administrative / proxy / auxiliary data for the full sample
  - Next, use multilevel modeling to simultaneously estimate interviewer variance in nonresponse errors (a predictor with 1 / -1 coding for R / NR) and measurement errors (imputed for NR)
  - Estimate the *covariance* of these two error sources as well
  - Idea is currently at the proposal stage (open for discussion!)
  - Initial simulation studies and empirical analyses provide good support for this idea, and suggest that simpler approaches (ignoring the covariance) may overstate estimates of interviewer variance

Future Research Directions: RQ #1

- Variance decompositions for **regression coefficients** (Research Question #4 today)
- Do different interviewer-administered **modes** introduce different variance components or have different decompositions? (e.g., CATI vs. Texting; Yan et al., 2013)

Research Question #2

- What effects can interviewers have on the quality of survey paradata, such as
  - interviewer observations,
  - records of call attempts, or
  - post-survey observations,

  and what design or training strategies can be used to reduce these effects?
- Error-prone paradata are problematic for nonresponse adjustments (West, 2013a, 2013b), and also potentially responsive survey designs…

West and Kreuter (2013)

- Demonstrate significant interviewer variance in the quality of observations of household features (based on completed rosters)
- Identify predictors of observation accuracy at the respondent, interviewer, and area levels
- Show that significant variance in observation accuracy remains, motivating future investigations of alternative strategies being used to make observations
- Similar results reported by Sinibaldi et al. (2013)

West and Kreuter (2013), cont'd

Figure 1: Variance in Observation Accuracy Among 94 NSFG Interviewers.

West and Kreuter (2015)

- Tests the idea of supplying interviewers asked to make judgments of selected features with a list of **observable** predictors of the feature being judged
- Providing this information to all interviewers in the last two quarters of the NSFG (Cycle 7) was shown to reduce false positive rates for sexual activity judgments
- Idea needs further testing based on randomized experiments; the key is helping interviewers with obs!
- Could also replace interviewer observations with predictions in "hard" areas; **idea also needs evaluation**

West and Kreuter (2017, R&R)

- NSFG Interviewers vary substantially in their ability to correctly judge traits of potential respondents (currently sexually active?) without prior training
- One *possible* source of this variance is alternative observational strategies being used in the field
- Open-ended justifications provided by interviewers for 3,992 current sexual activity judgments were coded and aggregated to the interviewer level
- Unique clusters of interviewers were evident based on 13 coded indicators; the clusters varied significantly in terms of judgment accuracy when adjusting for area-level features (e.g., those who detect relationship status in screening interviews are more accurate)

West and Li (2017, R&R)

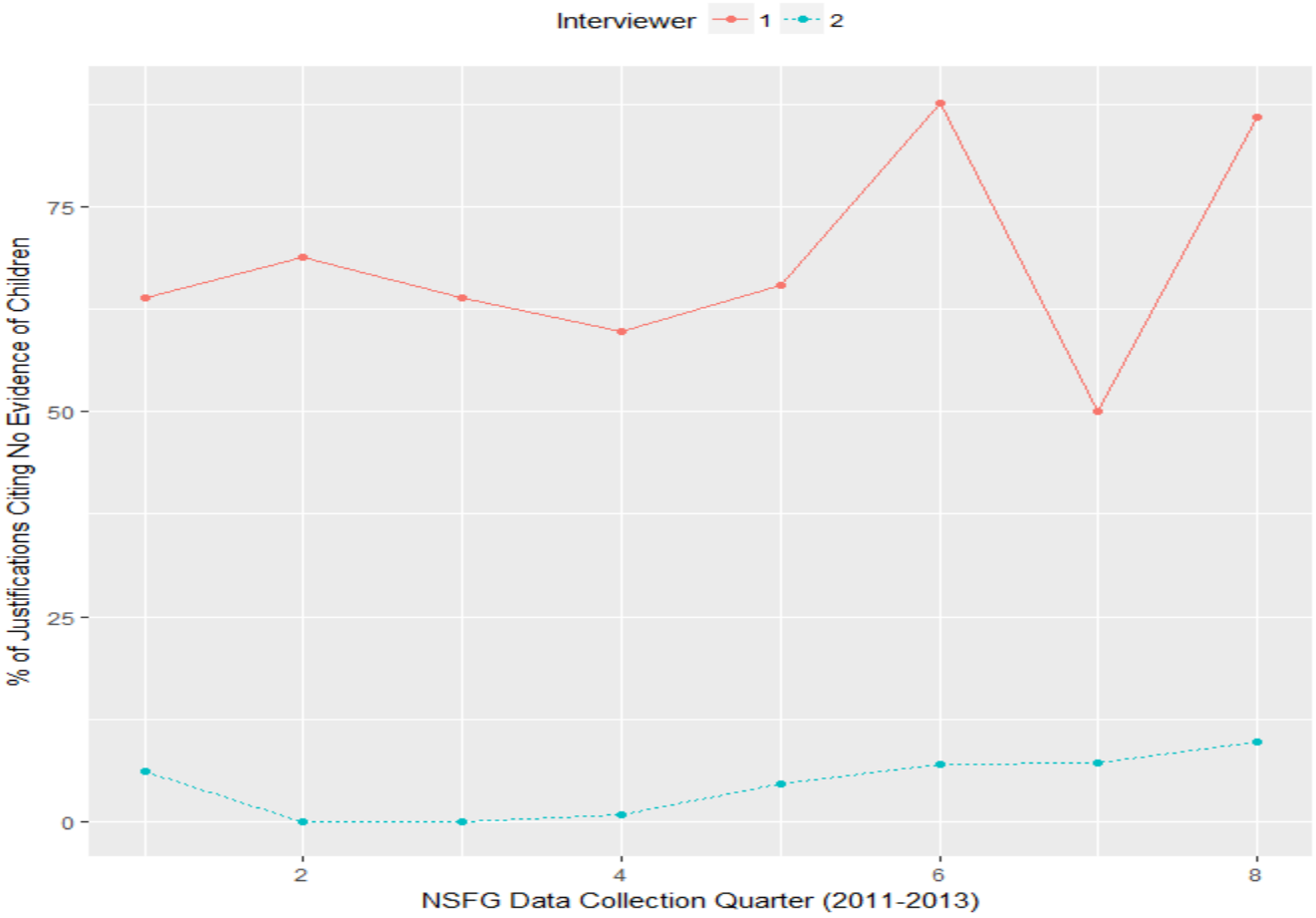
- An NIH-supported research project looking to build on the results of West and Kreuter (2013) and West and Kreuter (2017, R&R)
- More than 45,000 NSFG interviewer justifications for their observations on presence of young children and estimated response propensity (RP) were coded
- What cues did the interviewers report in their justifications, and did these cues explain interviewer variance in accuracy on these two observations?

West and Li (2017, R&R)

- Interviewers tended to vary substantially in the types of cues used, and a given interviewer tended to use the same cues across the data collection period
- This was true regardless of the type of area being worked by the interviewer!
- RP estimates tended to be more accurate in urban areas, while young children estimates tended to be more accurate in rural areas

West and Li (2017, R&R)

- Note this graph showing trends in the % of justifications indicating a general “no evidence of children” cue, for 2 IWERs in the largest MSAs



West and Li (2017, R&R)

- For RP estimates, the types of cues used explained *54% of the unexplained variance in accuracy among interviewers*
- For young children estimates, the types of cues used explained *42% of the unexplained variance in accuracy among interviewers*
- The detailed results (available upon request) provide direct practical guidance for future training
  - Some examples: take note of housing unit and area features; check with neighbors; note general lack of evidence of children (not toys!); and never guess!

Future Research Directions: RQ #2

- Literature also exists showing evidence of substantial interviewer variance in *post-survey observations* and *call records*
- A joint NIH proposal with Ting Yan and Frauke Kreuter has been submitted to explore this problem in more detail
- Future studies are needed looking at interviewer variance in the quality of post-survey observations in general (e.g., Wang, West and Liu, MAPOR 2013)

Research Question #3

- Do interviewer observations add predictive power in panel surveys, above and beyond key survey measures collected at previous waves?

West, Kreuter, and Trappmann (2014)

- Evidence of unexplained interviewer variance in the accuracy of observations on two economic features of households in Germany, from the PASS survey
  - Consistent with all prior studies discussed
- Interviewer observations of income and receipt of unemployment benefits do in fact add predictive power to models for current-wave measures of these variables, **above and beyond prior-wave measures**
  - The observations pick up *changes over time* that are not accounted for by using prior-wave measures

West, Kreuter, and Trappmann (2014)

- Results are important for improving nonresponse adjustments; given the focus on panel cases, there was relatively low attrition, so observations did not also predict response propensity
- In this case, we would only expect reductions in variance, not bias; the paper demonstrates this
- Also shows evidence of more effective observational strategies based on a small qualitative study (consistent with the studies for **Research Question #2**)
- **Future Work:** Replications of this study are needed in other panel surveys with higher attrition rates and/or less available auxiliary information

Research Question #4

- How do interviewers affect the quality of estimated regression coefficients?
- The vast majority of the literature on interviewer effects has considered the effects of interviewers on simple descriptive statistics
- Exceptions (all focusing on respondents!):
  - Davis and Scott (1995): domain comparisons
  - Beullens and Loosveldt (2014): latent constructs
  - Beullens and Loosveldt (2016): the ESS

Fischer et al. (2017, R&R)

- We designed a large simulation study to look at interviewer effects (from both recruitment and measurement) on Y (the DV) and X (the IV) in a simple linear regression model, and how these effects can impact estimates of simple linear regression coefficients
- We also analyzed administrative data from the earlier German CI/SI study to look at these same kinds of effects using real data

Fischer et al. (2017, R&R)

- Key Findings:
  - **Most important factors:** relationship of Y with response propensity (see Groves 2004), and interviewer variance in reported values of X
  - Real data application echoed these results
  - A standard adjustment based on the inverse of the reliability ratio that we are all familiar with (Fuller, 1987), incorporating the estimated interviewer variance in X, can correct much of the bias in the estimated coefficient
    - **Easy to compute, so use it!**
  - Need est. of simple response variance for full correction

Future Research Directions: RQ #4

- Do interviewer effects on Y and X play a similar role in generalized linear models?
- What if multiple X (predictor) variables are affected by interviewers (in a multiple regression context)? What correction should be used in that case?
- Replications are needed in other contexts where administrative data are available on the sampling frame

Questions / Discussion

- I am entirely open to ideas and suggestions on any of these topics, and hope to have a good discussion of them!
- For electronic copies of any papers or further discussion, please email **bwest@umich.edu**
- Thank you for attending the seminar today!

References

- Beullens, K. and Loosveldt, G. (2014). Interviewer Effects on Latent Constructs in Survey Research. Journal of Survey Statistics and Methodology, 2, 433-458.
- Beullens, K. and Loosveldt, G. (2016). Interviewer Effects in the European Social Survey. *Survey Research Methods*, 10(2), 103-118.
- Davis, P. and Scott, A. (1995). The Effect of Interviewer Variance on Domain Comparisons. *Survey Methodology*, 21(2), 99-106.
- Fischer, M., West, B.T., Elliott, M., and Kreuter, F. (2017, Under Review). The Impacts of Interviewer Effects on Regression Coefficients. Submitted to *Journal of Survey Statistics and Methodology*.
- Sinibaldi, J., Durrant, G.B., and Kreuter, F. (2013). Evaluating the Measurement Error of Interviewer Observed Paradata. *Public Opinion Quarterly,* 77, 173-193.
- Wang, Y., West, B.T., and Liu, M. (2013). Interviewer perceptions of Survey Data Quality. *Presented at the 38th Annual conference of the Midwest Association for Public Opinion Research (MAPOR), Chicago, IL, 11/23/2013.*
- West, B.T. (2010). A Practical Technique for Improving the Accuracy of Interviewer Observations: Evidence from the National Survey of Family Growth. NSFG Paper No. 10-013. November 2010.
- West, B.T. (2013a). An Examination of the Quality and Utility of Interviewer Observations in the National Survey of Family Growth. *Journal of the Royal Statistical Society, Series A (General)*, 176(1), 211-225.

References

- West, B.T. (2013b). The Effects of Error in Paradata on Weighting Class Adjustments: A Simulation Study. Chapter 15 in *Improving Surveys with Paradata: Analytic Use of Process Information*. Wiley Publishing.
- West, B.T., Conrad, F.G., Kreuter, F., and Mittereder, F*. (2017a, In Press). Can Conversational Interviewing Improve Survey Response Quality without Increasing Interviewer Variance? Journal of the Royal Statistical Society (Series A).
- West, B.T., Conrad, F.G., Kreuter, F., and Mittereder, F*. (2017b, R&R). Nonresponse and Measurement Error Variance among Interviewers in Standardized and Conversational Interviewing. Submitted to *Journal of Survey Statistics and Methodology*.
- West, B.T., and Elliott, M.R. (2014). Frequentist and Bayesian Approaches for Comparing Interviewer Variance Components in Two Groups of Survey Interviewers. *Survey Methodology*.
- West, B.T. and Kreuter, F. (2015). A practical technique for improving the accuracy of interviewer observations of respondent characteristics. *Field Methods.*
- West, B.T. and Kreuter, F. (2013). Factors Affecting the Accuracy of Interviewer Observations: Evidence from the National Survey of Family Growth (NSFG). *Public Opinion Quarterly*, 77(2), 522-548.
- West, B.T. and Kreuter, F. (2017b, R&R). Strategies for Increasing the Accuracy of Interviewer Observations of Respondent Features: Evidence from the U.S. National Survey of Family Growth. Submitted to *Methodology*.

References

- West, B.T., Kreuter, F., and Jaenichen, U. (2013). Interviewer Effects in Face-to-face Surveys: A Function of Sampling, Measurement Error or Nonresponse? *Journal of Official Statistics*, 29(2), 277-297.
- West, B.T., Kreuter, F., and Trappmann, M. (2014). Is the collection of interviewer observations worthwhile in an economic panel survey? New evidence from the German Labor Market and Social Security (PASS) study. *Journal of Survey Statistics and Methodology,* 2(2), 159-181.
- West, B.T. and Li, D. (2017, R&R). Sources of Variance in the Accuracy of Interviewer Observations. Submitted to *Sociological Methods and Research*.
- West, B.T. and Olson, K. (2010). How much of interviewer variance is really nonresponse error variance? *Public Opinion Quarterly*, 74(5), 1004-1026.
- Yan, H., West, B.T., Conrad, F.G., and Schober, M.F. (2013). Comparing Interviewer Variance Components Between Telephone and Text Message Interviews. *Presented at the 38th Annual conference of the Midwest Association for Public Opinion Research (MAPOR), Chicago, IL, 11/23/2013.*