

Exploring the association between question characteristics, respondent characteristics, interviewer characteristics and survey data quality

Aaron Maitland, Westat

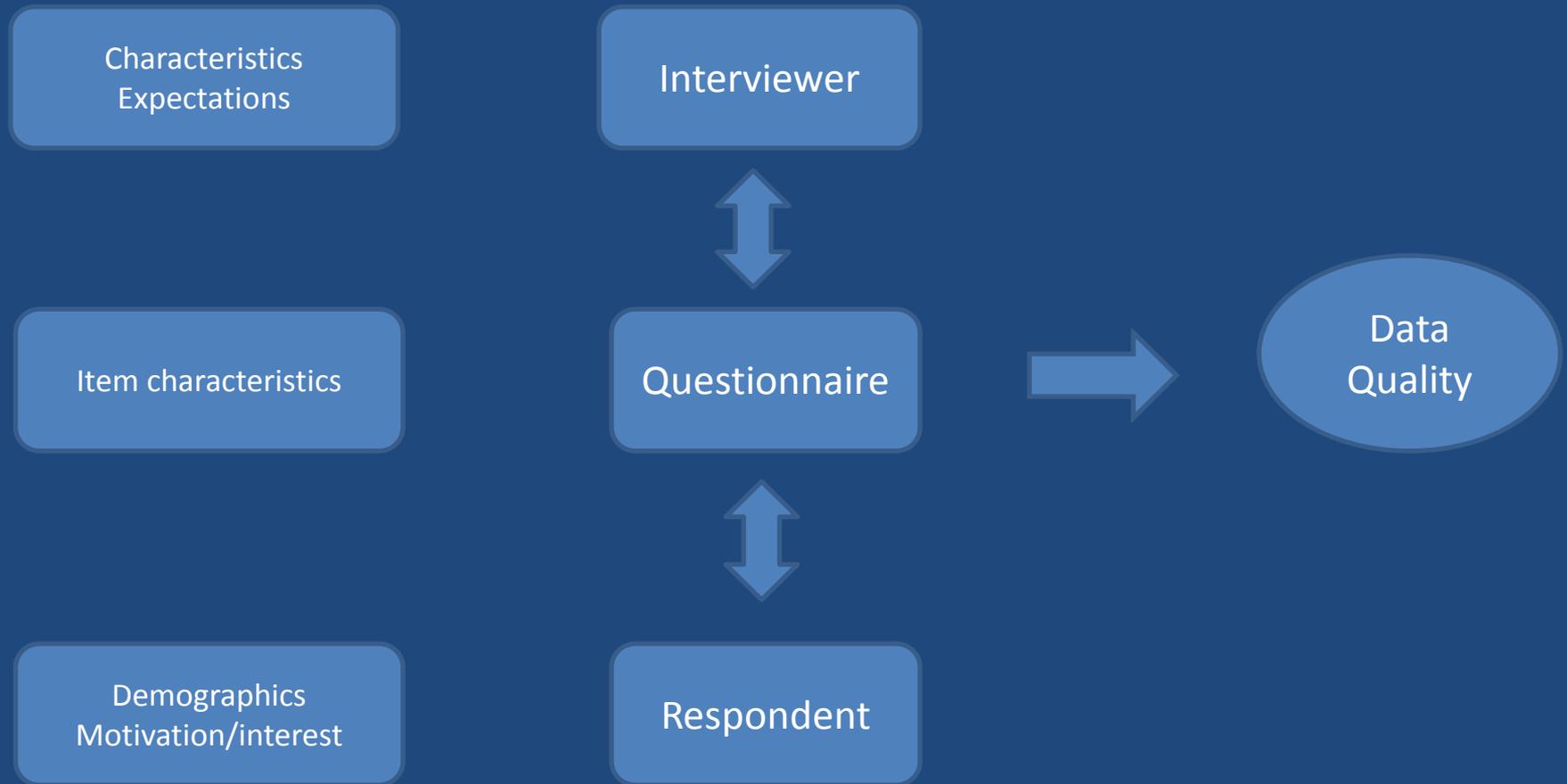
Heather Ridolfo, NASS

James Dahlhamer, NCHS

Antuane Allen, NCHS

Dynesha Brooks, NCHS

Background and Motivation



The Real Motivation

- This study was motivated by observations of face to face interviews in the field.
- Interviewers have the dual roles of recruiting respondents and administering questions.
- Interviewers want the interview to go smoothly to obtain a completed interview.
 - Tendency to avoid potentially difficult or embarrassing situations
 - Not always obvious when these situations might arise

Research Questions

- How much of the variability in item nonresponse is due to interviewers, respondents, and the items?
- What characteristics are related to item nonresponse?

Methods

- 2010 National Health Interview Survey (NHIS) sample adult interview
- Set up hierarchical data file with items nested with respondents nested within interviewers.
- Mixed models that include random effects for the interviewer and respondent. Residual variability is due to the items.
- Analyses run in SAS PROC GLIMMIX
- Current presentation focuses on ten percent sample for model building.

Variables

- Item characteristics
 - Predictors of nonresponse: item difficulty, unfamiliar technical terms, item sensitivity, content, response format
 - Paradata: interviewer instruction, show cards, help screen, edits, optional text, question fills, question type, questionnaire section
- Respondent characteristics
 - Demographics: age, race/ethnicity, gender, education, marital status, employment status, income, health status, region of residence
 - Paradata: hard refusal, gatekeeper, privacy/content concerns, time concerns, case reassigned, number of contacts, telephone interview, non-English interview
- Interviewer characteristics
 - Cooperation rate in previous quarter
 - Item nonresponse rate in previous quarter

Random Effects

Table 1. Intraclass correlation coefficients.

Interviewer	.06
Respondent	.17

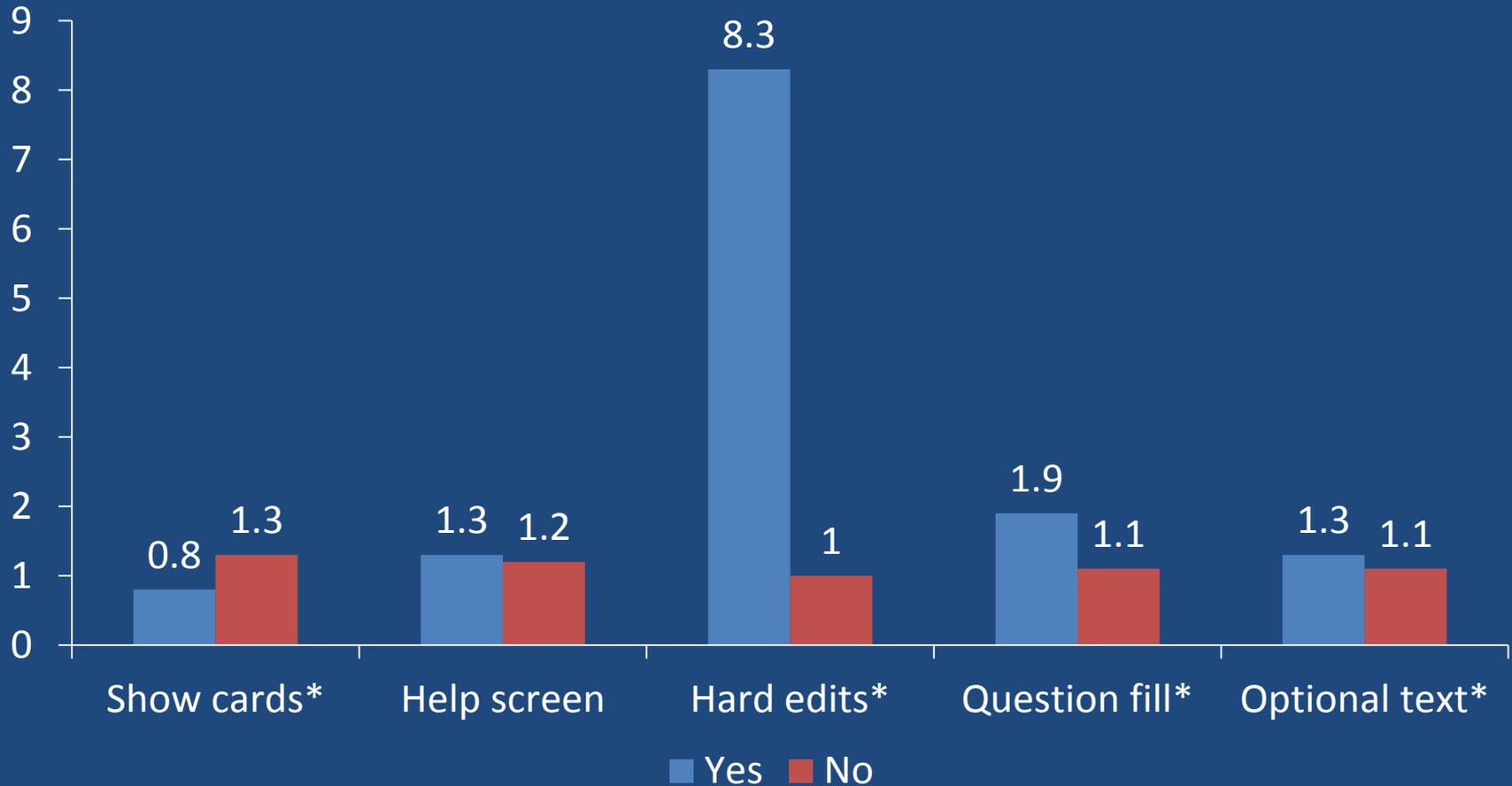
Table 2. Range of missing data by level.

Level	Minimum	Maximum
Interviewer	0	.07
Respondent	0	.08
Question	0	.22

Bivariate Analyses: Item Difficulty and Sensitivity

Flesch-Kincaid reading ease***	Item nonresponse rate
Very easy	1.6
Easy	0.6
Fairly easy	0.7
Standard	0.9
Fairly difficult	1.3
Difficult/very difficult	2.7
Question sensitivity index (range 10-30)***	Item nonresponse rate
10	1.2
11-13	0.4
14-17	1.2
18+	3.6
***p<.001	

Bivariate Analysis: Item Characteristics

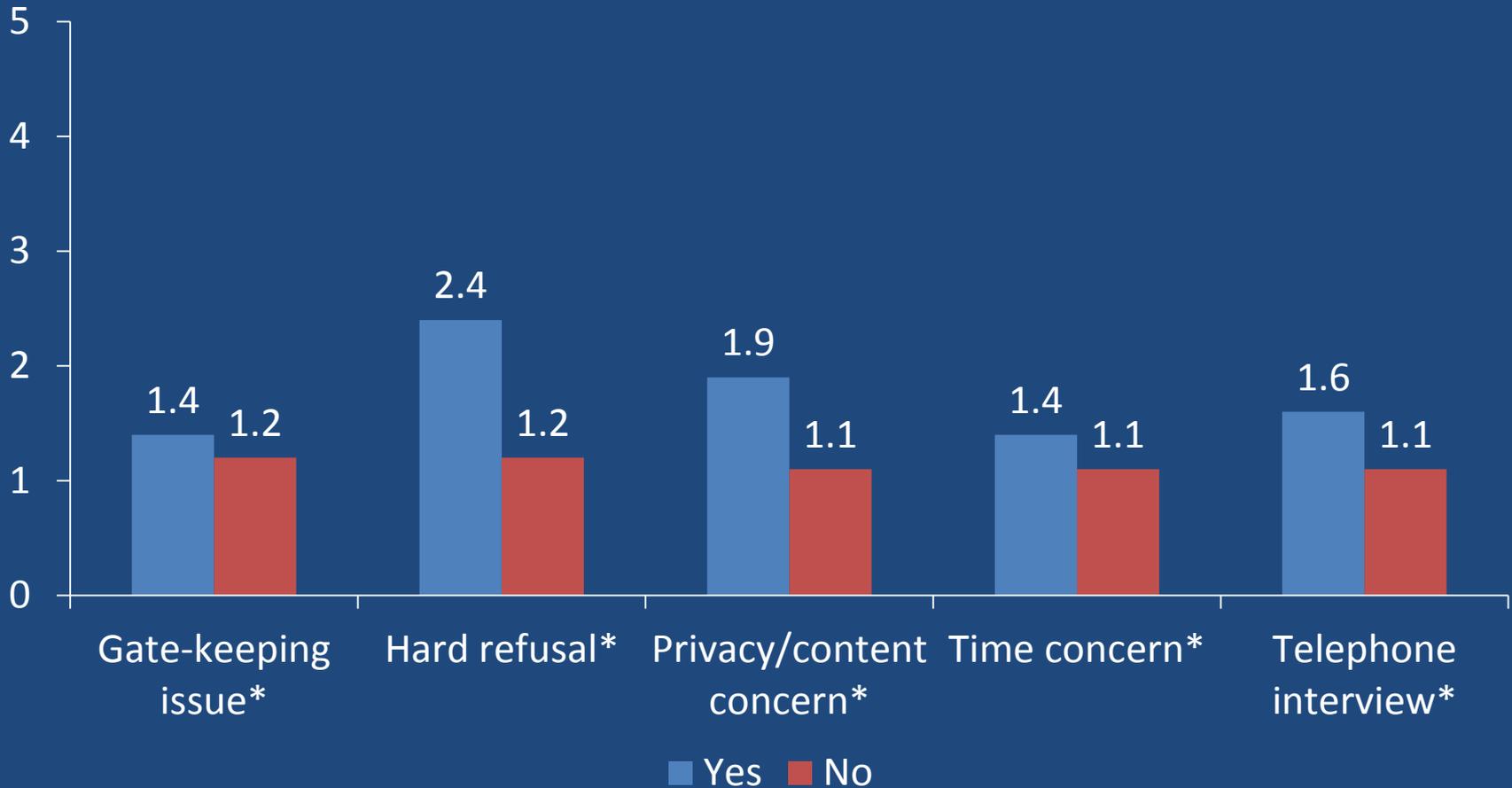


*P<.001

Bivariate Analyses: Respondent Characteristics

Age***	Item nonresponse rate
18-24	0.9
25-44	1.2
45-64	1.1
65+	1.7
Education***	Item nonresponse rate
Less than high school	1.4
High school	1.3
Some college	1.1
Bachelor's degree or more	1.0
***p<.001	

Bivariate Analysis: Respondent Paradata



*P<.001

Bivariate Analyses: Interviewer Characteristics

Interviewer cooperation rate***	Item nonresponse rate
< 72%	1.6
73% - 86%	1.2
87% - 96%	1.1
97%+	1.1
Interviewer item nonresponse rate***	Item nonresponse rate
0%	0.9
.10% - .39%	0.9
.40% - .64%	1.1
.65% - 1.24%	1.3
1.25%+	1.9
***p<.001	

Table 2. Summary of multilevel model fixed effects (Effect on the odds of item nonresponse).

Level	Increase	Decrease	Other
Item	Edits	Interviewer instruction	Difficulty
	Optional text	Show card	Sensitivity
	Question fill	Help screen	
	Time		
Respondent	Age 65+		
	Low education		
	Missing income		
	Hard refusal		
	Privacy/content concern		
	Telephone		
	Non-English		
Interviewer	Higher item nonresponse rate in previous quarter		

Discussion

- Found small proportion of variability in item nonresponse to be between interviewers
- Puzzling findings for question difficulty and sensitivity
- Paradata findings
 - Encouraging that help screens, show cards, and instructions help
 - Edits, fills, and optional text may need further investigation.
 - Less cooperative respondents have lower data quality
- Findings with respect to mode and language may suggest focusing question design efforts

Next Steps

- Improve specification of the model
- Try to identify if particular types of items are more subject to interviewer effects
- Considering additional outcome variables as measures of data quality