Examining Interviewer Behavior in Handling 'Difficult' Cases

Prepared for FedCASIC
March 28, 2012

Wendy Hicks, Aaron Maitland, and Brad Edwards

Motivation for Study

- What do we know about the contextual factors related to interviewer behaviors?

- Can we use CARI to predict the situations in which interviewers might deviate from standard/expected behaviors?

Developing Our Hypothesis…

- Behavior coding studies have not found a strong or consistent link between interviewer behavior and data quality
  - Hess, Singer and Bushery (1999)
  - Schaeffer and Dykema (2011)

- However, studies using survey/frame data do link interviewers to both measurement error and nonresponse:
  - West and Olson, 2010
  - Tourangeau, Kreuter and Eckman (in press)

Specific Hypothesis

- Interviewers behavior differs by the perceived difficulty or cooperativeness of the case
  - More difficult the case, the more likely interviewers will deviate from expected (standardized) interview protocol
    - shortened protocol, reduced content
  - Conversely, interviewers will more often maintain the standardized interview protocol with less difficult cases

Study Design

- National Health and Aging Trends Study (NHATS)
  - Longitudinal design
  - Sample of Medicare beneficiaries (65+)
  - CAPI interview – about 2 hours average length
  - Baseline data collected May through August, 2011
  - Recorded a subset of core and filtered questions
- Used contact history data to divide into two levels of ‘cooperation’– Difficult or not
- Analyzed CARICode data by Difficulty group

Westat CARICode Approach

- As long as R gives consent, all interviews recorded
- Each interviewer's first completed case selected for coding, and then sampled thereafter
  - Small number of interviewers have first two selected
- For NHATS, started at 25% sample, then dropped to 10% sample per interviewer
- Each coded case is assigned a "result code"
- Result codes are tracked within and across interviewers

CARICode Data – Defining Result Codes

- 'Overall' score - Aggregates performance across all coded questions within an interview
  - Validation Risk – highest priority code (excluded from this analysis)
  - Inaudible – less than half of the recordings in the interview were audible
  - Question Administration issue - At least 20% of audible recordings were coded as 'changed meaning'
  - Probing issue – Interviewer did not probe in at least 20% of audible recordings in which the respondent initially gave an invalid answer
  - Professionalism concern – biased feedback/commentary, taking cell phone calls, etc

Skill Issues Identified:

Skill issue	Count	%
All cases sampled for coding	1,670	(100)
Administration Issue	66	4.0
Probing Issue	72	4.3
Professionalism concern	159	9.5
Multiple Issues *(admin + other)*	362	21.7
Total cases with any issue	659	39.4%

Categorizing into Difficulty Groups

- Used Electronic Record of Contacts (EROCs) to categorize
- Difficult Case (n=517)
  - Ever refused (refusal conversion case)
  - 2 + broken appointments
  - 75th percentile in number contacts/attempts
- Not difficult – the remainder (n=1,153)

Results: Any Type of Skill Issue by Difficulty of Case

Overall Interviewer Skill Issue	Difficult Case (N=517)	Not Difficult (N=1,153)
Yes	36.7	40.7
No	63.3	59.3
	100%	100%

Chi-square: 2.30

Results: Skill Issue by Case Difficulty, at Question Level

	Percent of Q recordings with an interviewer error in reading the question		
Question	Difficult Case (N=517)	Not Difficult (N=1,153)	Chi-square
Health Conditions (HC6)	2.9	4.2	1.56
Overnight stay (HC7)	12.9	14.2	1.25
Senior housing (HT4)	9.1	11.5	2.06
Who does laundry (HA1)	22.1	27.5	5.47*
Shower, bath (SC7)	5.8	8.8	4.29*
Well-Being (WB)	1.0	2.6	4.64*
Economic wellness (EW7)	20.1	23.2	1.99

* ($p<0.05$)

Results: Do Respondents Differ by Case Difficulty?

- Maybe this pattern is reflective of differences in respondents by the two groups
  - Less difficult, longer avg. interview time (Conrad, et al 2008)
  - In this population, less difficult may be less ‘able’ as well

- Looked at respondent data
  - Compared respondent attributes by “difficulty”
  - Compared respondent behavior by “difficulty”, per core question

Results: Respondent Attributes by Case Difficulty

	Self Rated Health		Self Rated Memory	
Rating	Difficult Case (N=1,153)	Not Difficult (N=517)	Difficult Case (N=1,109)	Not Difficult (N=483)
Excellent	12.3	12.0	11.7	12.2
Very Good	25.8	27.3	31.2	31.9
Good	31.7	32.1	34.9	36.0
Fair	21.7	22.2	18.2	17.2
Poor	8.5	6.4	4.1	2.7
	100%	100%	100%	100%

Results: R Issue by Case Difficulty, at Question Level

	Percent of Q recordings where R initial response not a valid answer		
Question	Difficult Case (N=517)	Not Difficult (N=1,153)	Chi-square
Health Conditions (HC6)	5.4	5.5	~0.00
Overnight stay (HC7)	2.3	2.9	~0.00
Senior housing (HT4)	4.5	4.7	~0.00
Who does laundry (HA1)	3.9	4.9	~0.00
Shower, bath (SC7)	3.7	3.3	~0.00
Well-Being (WB)	30.0	35.0	4.05*
Economic wellness (EW7)	3.3	2.1	2.18

* ($p<0.05$)

Why Would 'Not Difficult' Cases Result in More Skill Issues?

- Sample for coding at the interviewer level may influence results
  - All interviewers' first case
  - Variability in interviewer skills/behaviors may reflect differences in interviewer productivity
    - Difficult cases not randomly assigned

- Repeat analysis –
  - Separate the "first" case from remaining for all interviewers
  - Control for differences in interviewer productivity

Results: % Skill Issues by Difficulty, by first or later case

Case type	N	Difficult Case	N	Not Difficult	Chi-square
First case only	61	54.1	179	41.3**	4.92**
All other cases	456	34.4	974	40.6*	3.00*

* p=0.08 ** p=0.03

Suggests that as gain more experience with the interview, the interviewers' behavior changes (Olson & Peytchev, 2007)

Creating Interviewer Productivity Groups

- Divided each interviewer into one of three groups
  - Low productivity: # of completed cases in 25th percentile
  - Average productivity: # of completed cases in 26th -74th percentile
  - High productivity: # of completed cases in 75th percentile

- Each coded case had interviewer productivity flag

Results: Skill Issue by Difficulty–Controlling for Productivity

	Percentage of Difficult or Not, with an Interviewer Skill Issue, by Productivity category of Interviewer				
Productivity category	N	Difficult Case	N	Not Difficult	Chi-square
Low	(39)	59.0%	(116)	37.9%	5.27*
Average	(295)	34.6%	(620)	44.8%	8.67*
High	(183)	35.5%	(417)	35.3%	.00

* ($p<0.05$)

Skill Issue by Difficulty–Controlling for Productivity, Case

	Percentage of Difficult or Not, with an Interviewer Skill Issue, by Productivity category of Interviewer , ignoring first case				
Productivity category	N	Difficult Case	N	Not Difficult	Chi-square
Low	(24)	58.3%	(69)	37.8%	3.00*
Average	(265)	32.4%	(531)	44.6%	10.87**
High	(167)	34.7%	(374)	35.3%	.00

* (p=0.08), ** (p=0.001)

What do these Data Suggest?

- Interviewer performance/behavior may be related to the perceived difficulty of a case
- Direction of relationship seems related to productivity
  - Seems independent of respondent attributes or behaviors
  - Relationship strengthens if account for the “learning curve”
- Training implications?

Next Steps in our Understanding

- Account for interviewer experience (and other attributes, such as ‘converters’),
- Further analysis of the CARI data, at an exchange level, by Difficulty group
- Apply to a general population survey
- Affect data quality?

QUESTIONS OR COMMENTS:

WendyHicks@Westat.com

A special thank you to **Andrew Mercer and Drew Kistler** for assistance with processing the contact record data.

Wording of Individual Questions

HC6:

Is there another serious disease or illness that I have not asked about that a doctor has told you that you have?

HC7:

Have you had an overnight hospital stay within the last 12 months, that is since (DATE)?

Wording of Individual Questions

HT4:

I have recorded that you live in a (structure type). Is your home part of a retirement community or a senior housing community?

Wording of Individual Questions

HA1:

(Showcard)

Let’s start with your laundry. By laundry we mean cleaning your clothing, sheets and towels.

Which answer best describes how your laundry got done in the last month? Did you always do this by yourself, always do it together with someone else, did someone else always do it for you, or did it vary?

Wording of Individual Questions

SC7:

In the last month, how did you usually clean yourself up? By taking a shower, bathing in a tub, or washing up some other way?

IF NEEDED: Do not include whirlpool baths you take for therapy.

Wording of Individual Questions

WB:

Sometimes people feel older or younger than their age. During the last month, what age did you feel most of the time.

Wording of Individual Questions

EW:

Family members often help each other out financially. The next questions are about last year, ending December 31.

Last year, did you receive any financial help or financial gifts from (children or other) relatives, either regularly – like every month – or just every so often as needed?

Results: R Issue by Case Difficulty, at Question Level

	Percent of Q recordings with an interviewer error in reading the question, by workload		
Individual Question	Difficult Case (n=517)	Not Difficult (n=1,153)	Chi-square
Overnight stay (HC7)			
Small	31.0	13.0	6.46*
Average	10.5	14.8	3.19
Large	10.9	13.7	.85
Who does laundry (HA1)			
Small	48.7	31.0	3.99*
Average	18.9	29.3	12.66*
Large	22.4	23.7	.13
Economic wellness (EW7)			
Small	46.2	27.6	4.60*
Average	17.6	22.2	2.57
Large	18.6	23.5	1.80

* ($p<0.05$)