Comparison of Estimates From An Address-Based Mail Survey And A RDD Telephone Survey

David Cantor
Brett McBride

Westat

Presentation for the International Total Survey Error Workshop, Stowe, VT., June 13 – 16, 2010

Mode of Interview Matters

- There is a significant amount of research on differences between different modes of interviewing (e.g., de Leeuw, 2005).

- How can we apply this research when making practical decisions when choosing between modes?
  - Is it enough to say that using different modes lead to different results?
  - How much do they differ?
  - What are the implications of these differences for users of data?

How Big are the Differences Between a Mail and Telephone Survey?

- How does the magnitude of the difference vary by different question types?
  - Social desirability
  - Ordinal scales
  - Open vs closed questions
  - Questions asking about knowledge of particular topics
- How does the difference compare to sampling variability?

Health Information National Trends Survey 2007

Mail Component

- Frame: USPS addresses
- Response rate: 31%
- n of completes = 3582
- Coverage – civilian non-institutional

Telephone

- Frame: Random Digit Dial
- Response Rate: 24%
- n of completes = 4081
- Coverage – civilian non-institutional with a landline telephone

Analytic Strategy

- Computed Root Mean Square Error (RMSE) assuming one of the modes is the “best”. For example, if it assumed mail is best:

$$RMSE = \sqrt{MSE}$$

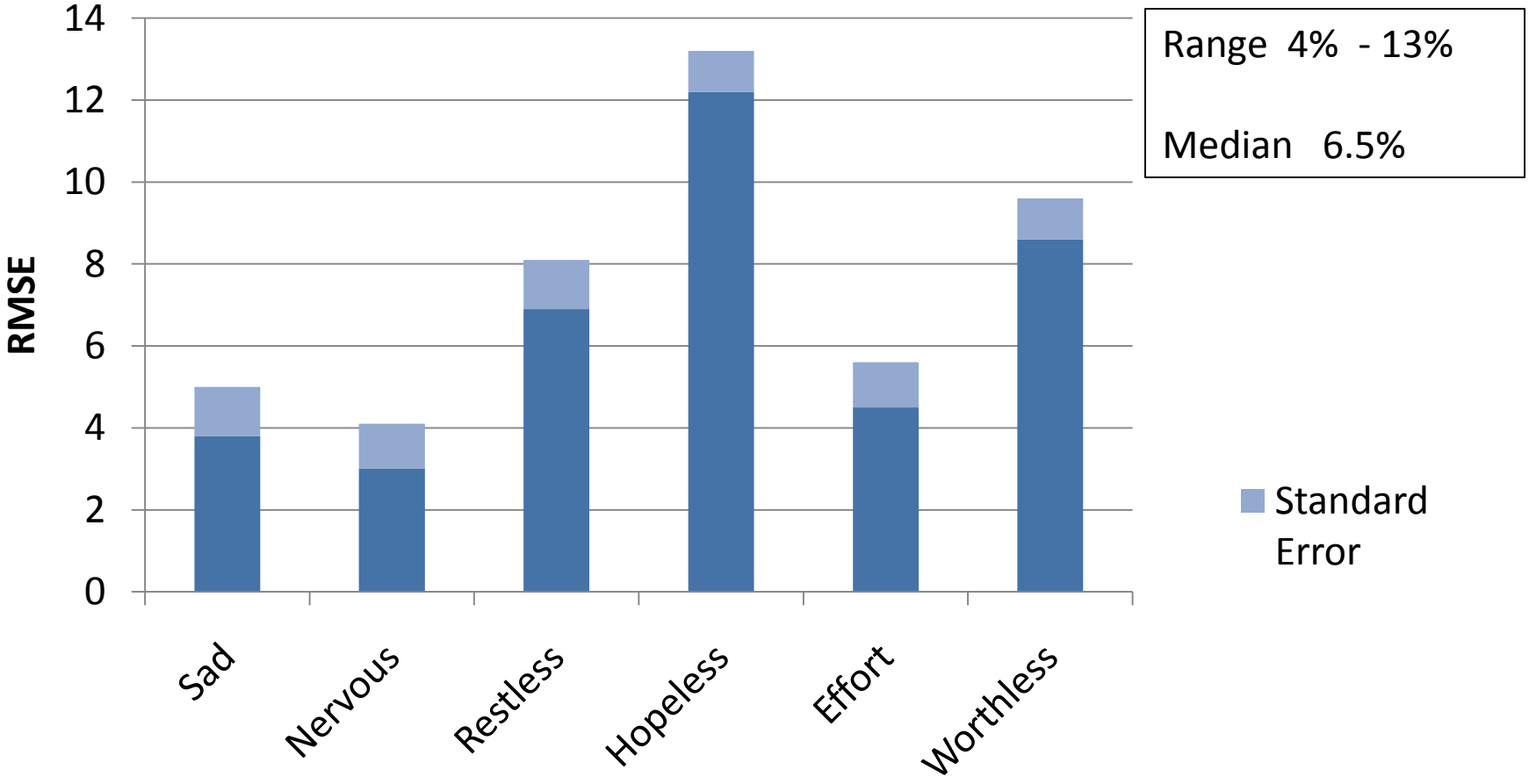
$$MSE = V_T + B_T^2$$

$$B_T = x_T - x_m$$

Items Related to Social Desirability

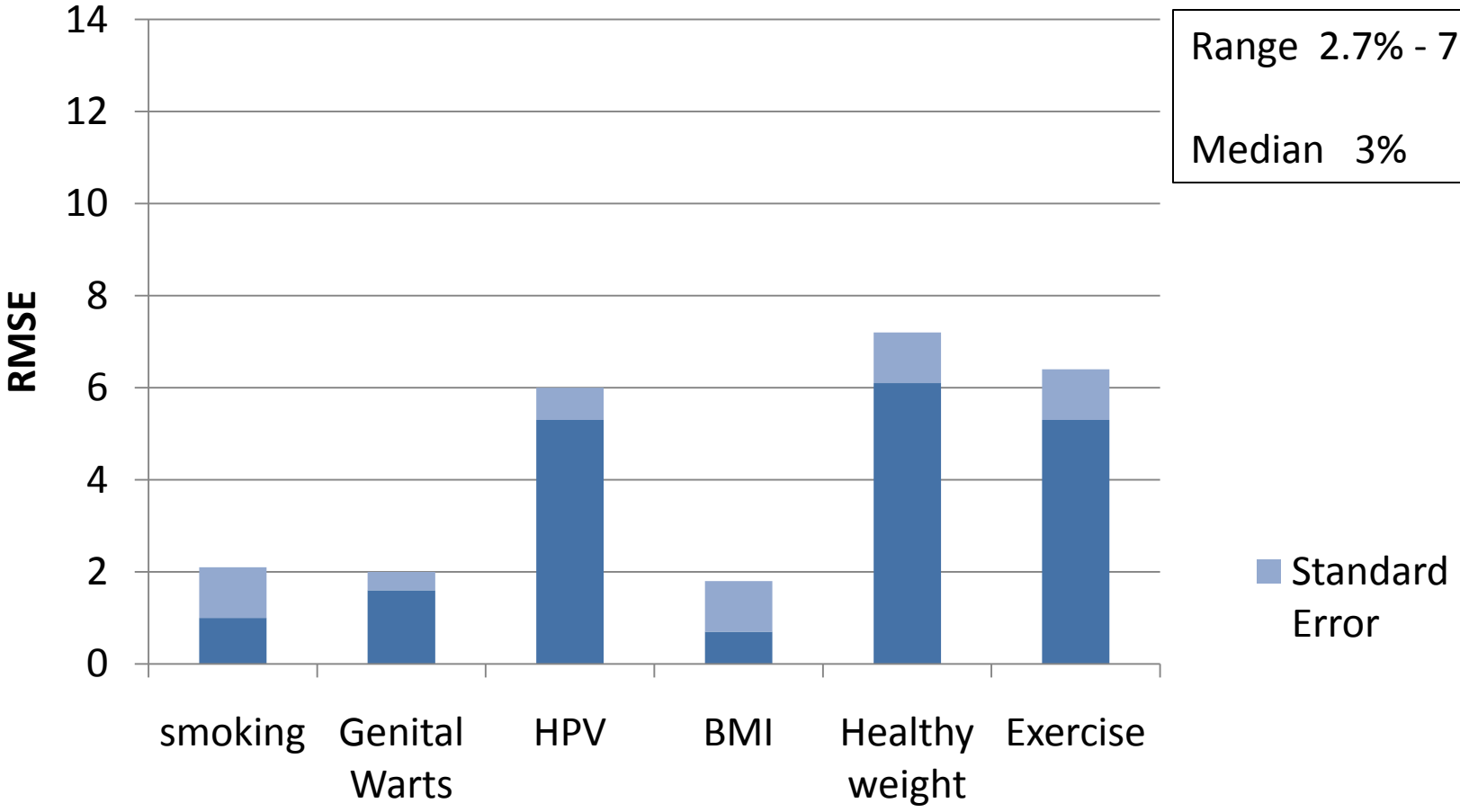
- Serious Mental Illness (sad, nervous, restless..)
  - “none of the time”
- Current smoking status
- Treated for genital warts
- Ever had HPV
- Weight + Height = BMI score of those Obese
- Participate in physical activities or exercise
- Weight perception = “just about right”

RMSE for Serious Mental Illness: "None of the Time"



Difference between modes for all items are statistically significant at p<.02

RMSE for Other Sensitive Items



Difference between modes are statistically significant at p<.02 for all items except smoking and BMI

52 Ordinal Scales Examined

- 19 using 4 point “agree to disagree”
- 11 using 4 point “a lot to none”
- 6 using 4 point “always to never”
- 6 using 5 point “All of the time to none of the time”
- 10 miscellaneous scales with 3, 4 and 5 points

Measure For Multiple Response Categories

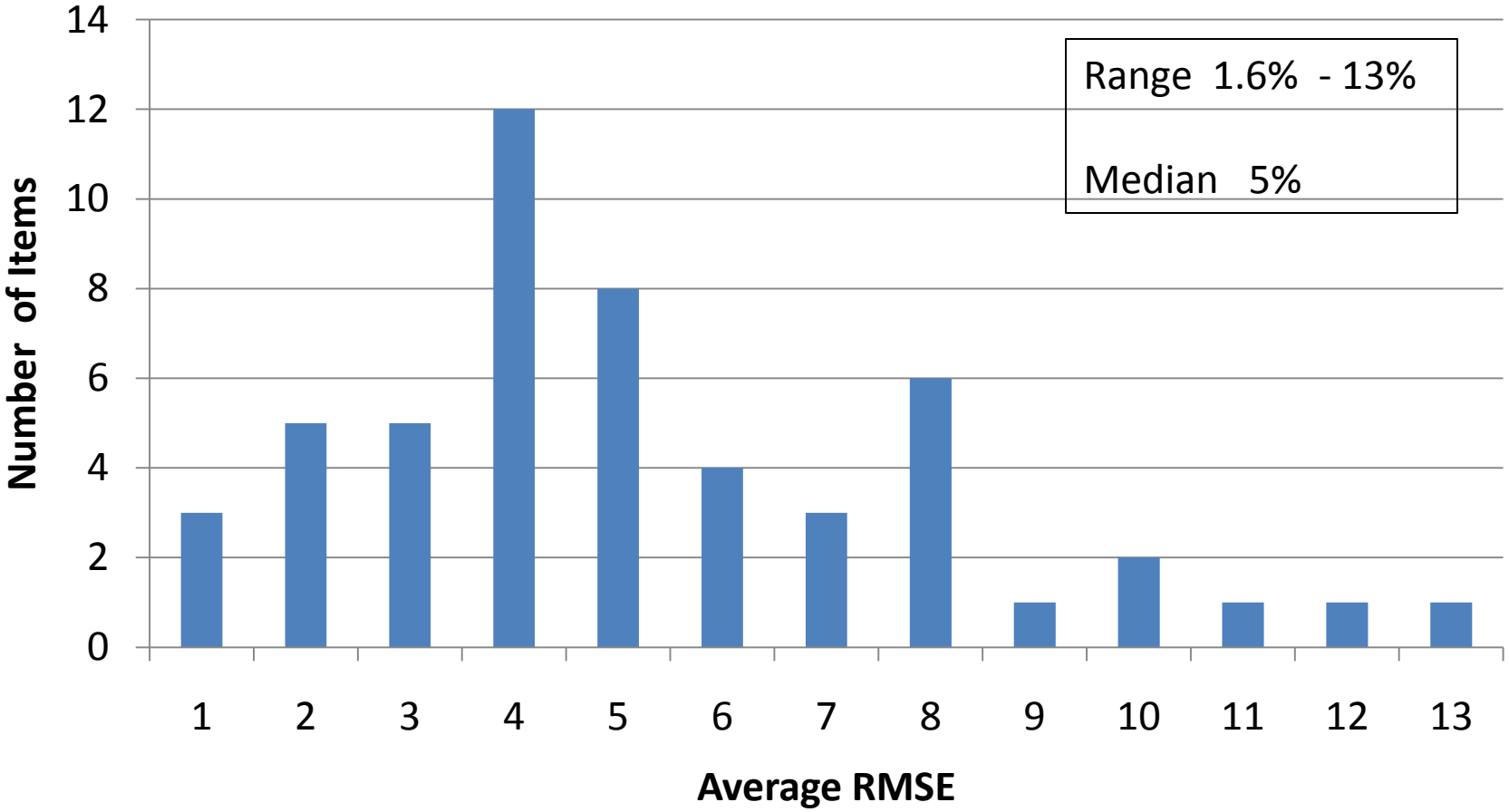
- Measure examined averaged over all response categories. Weighted by the sample size in that category:

$$\text{Average } RMSE = \sum_{i=1}^{n} p_i \; RMSE_i$$

p_i = proportion of sample in ith response category

- Those categories with the largest sample sizes will have the most influence on the result

Average RMSE for 52 Ordinal Scales



47 of 52 items differ by mode at p<.02

Partially Closed Questions on the Telephone with Qualitative Response Categories

- Open-ended for respondent on the telephone. Interviewer has a list of categories to code.
  - Expect respondent to provide the most salient response set
  - Scope and purpose of question may not be fully understood
  - May be difficult to fit response into a pre-coded category

- Mail survey is closed ended with list displayed to respondent
  - Response categories serve as cues to define and promote recall
  - Categories may restrict memory search and/or lead to over-reporting
  - More likely to mark more categories (for “mark all that apply”)

- Long lists are difficult to use “unimodal” approach

High Variability Across 7 Items

Question	Average RMSE	# Cat Differ at p<.02+
Where did you look for information about health?	7	5 of 13
What type of information were you looking for?	21	15 of 15
Where did you use the internet?	10	6 of 7
Where did you look for Cancer information?	2	1 of 13
How did you access the internet?	2	1 of 5
Why did you get a Pap test?	3	3 of 6
Why would you not have daughter get HPV test?	5	4 of 11

+ Number of response categories that differ by mode at p<.02

Partially Closed Questions on the Telephone with Quantitative Response Categories

- Open-ended questions asking for quantitative information
  - When did you last see doctor?
  - How many times have you seen the doctor?

- On the telephone the interviewer has a list of ordinal categories that are not read to respondent
  - Eg. 0 times; 1-2 times; 3-10 times, 11-24 times, 25+
  - Possible estimation problems for respondent

- Mail survey is closed ended
  - Response categories may define “normal” distribution
  - May be a tendency for respondent to choose middle of distribution

Partially Closed with Quantitative Response Categories: Results

	Avg RMSE	P<.02+
How many times did you go to the doctor?	2.1	Y
In a typical week, how many days do you exercise?	1.9	
How many times have you used a tanning bed or booth?	3.7	Y
How many times have you used tanning products?	4.2	Y
When did you have your last Pap test?	6.8	Y
When did doctor last discuss getting a check for colon cancer?	3.3	Y
When was most recent stool blood test using a home kit?	5.2	Y
When was most recent colonoscopy?	2.2	
When was most recent sigmoidoscopy?	5.8	

+ Y = difference between modes is significant at p<.02

Knowledge Questions

- Items asking about knowledge on:
  - Different causes of cancer
  - Ways to prevent cancer
  - Effectiveness of specific tests for cancer
- Total of 10 items
- Did not provide “Don’t Know” category in either mode

Knowledge Questions: Results

- Many more Don’t Know responses on the telephone
  - 8 of 10 items had higher than average percentages reporting DK
  - Range: 5% to 28%
  - Median: 15%
- The two items that did not have any DKs had 4 category responses (a lot, some, a little, Not at all)

Summary of RMSE by Question Type

Question Type	# items	Median RMSE	Range (A)RMSE
Serious Mental Illness	6	6.5	4 – 13
Other Social Desirability	6	3.0	2.7 - 7
Ordinal	52	5.0	1.6 – 13
Partially Closed – Qualitative	7	5.0	2 – 21
Partially Closed – Quantitative	8	3.7	2.1 – 6.8

Conclusions: Magnitude of Differences

- Biggest differences:
  - Open ended questions and ordinal scales.
  - Serious Mental Illness “none of the time”

- Magnitude of differences for analysis
  - Analysis of levels: Depends on precision requirements. Percent with Serious Mental Illness vs. Percent believe “right weight”.
  - Analysis of relationships – not addressed here

- Contribution of sampling error to total error is quite small for these estimates.

Conclusions: Direction of Differences

- Mail yields fewer socially desirable answers
- Telephone displays “extreme” responses
- Partially closed items
  - qualitative data: high variability
  - quantitative data: small differences
- Knowledge questions: Telephone respondents much more likely to use “don’t know” response

Limitations

- This is a case-study
  - Magnitude estimates are specific to the types of questions included on the HINTS questionnaire
  - Some non-comparability between surveys due to differential coverage and non-response
  - Does not explore collapsing categories in different ways (e.g., collapsing ordinal categories)
- The estimates do not fully account for sampling error for mode differences

Discussion Questions

- How useful is this type of analysis? Is it too specific to this survey?

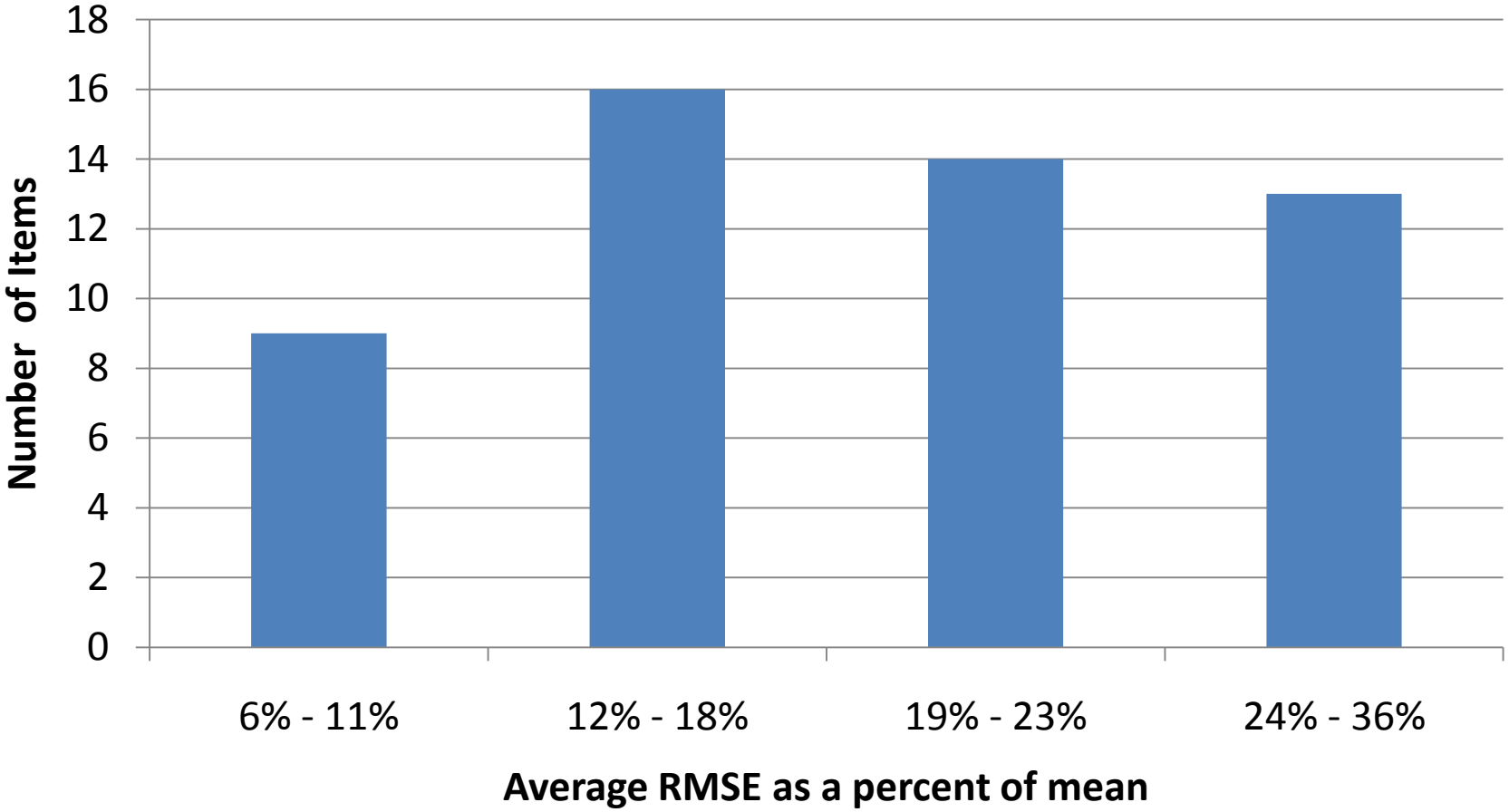
- Are there other question characteristics that are important to capture when comparing mail and telephone?

- Are there other analyses that should be done to solidify estimates of magnitude?

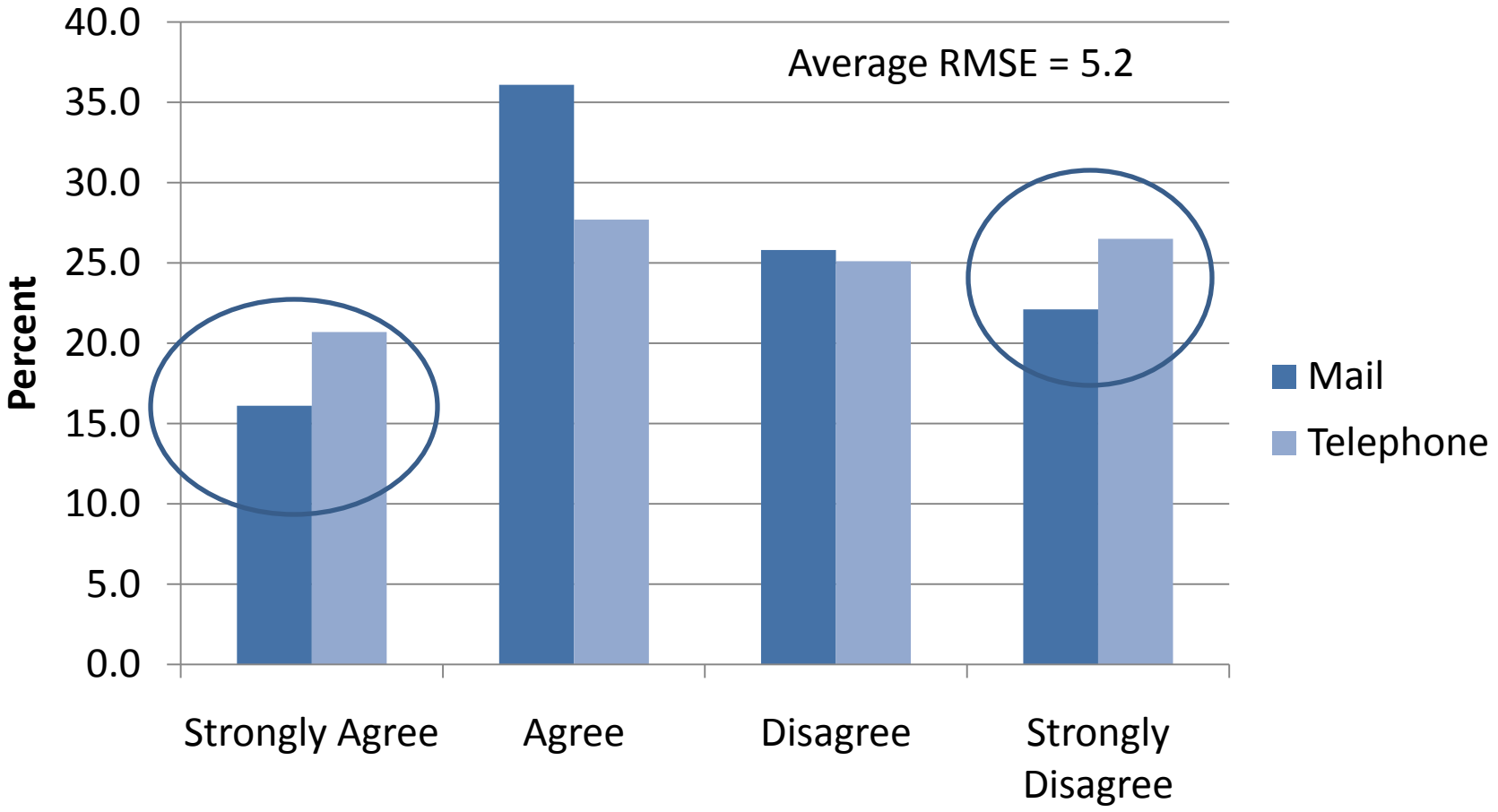
Social Desirability: Results

- Serious Mental Health items:
  - Range: 4% to 13%
  - Median: 6.5 %
- Other items:
  - Range: 2% to 7%
  - Median: 3%
- Contribution of variance is relatively small. The larger errors are driven by non-sampling error
- Does this make a difference?

Average RMSE as a Percent of Mean for 52 Ordinal Scales

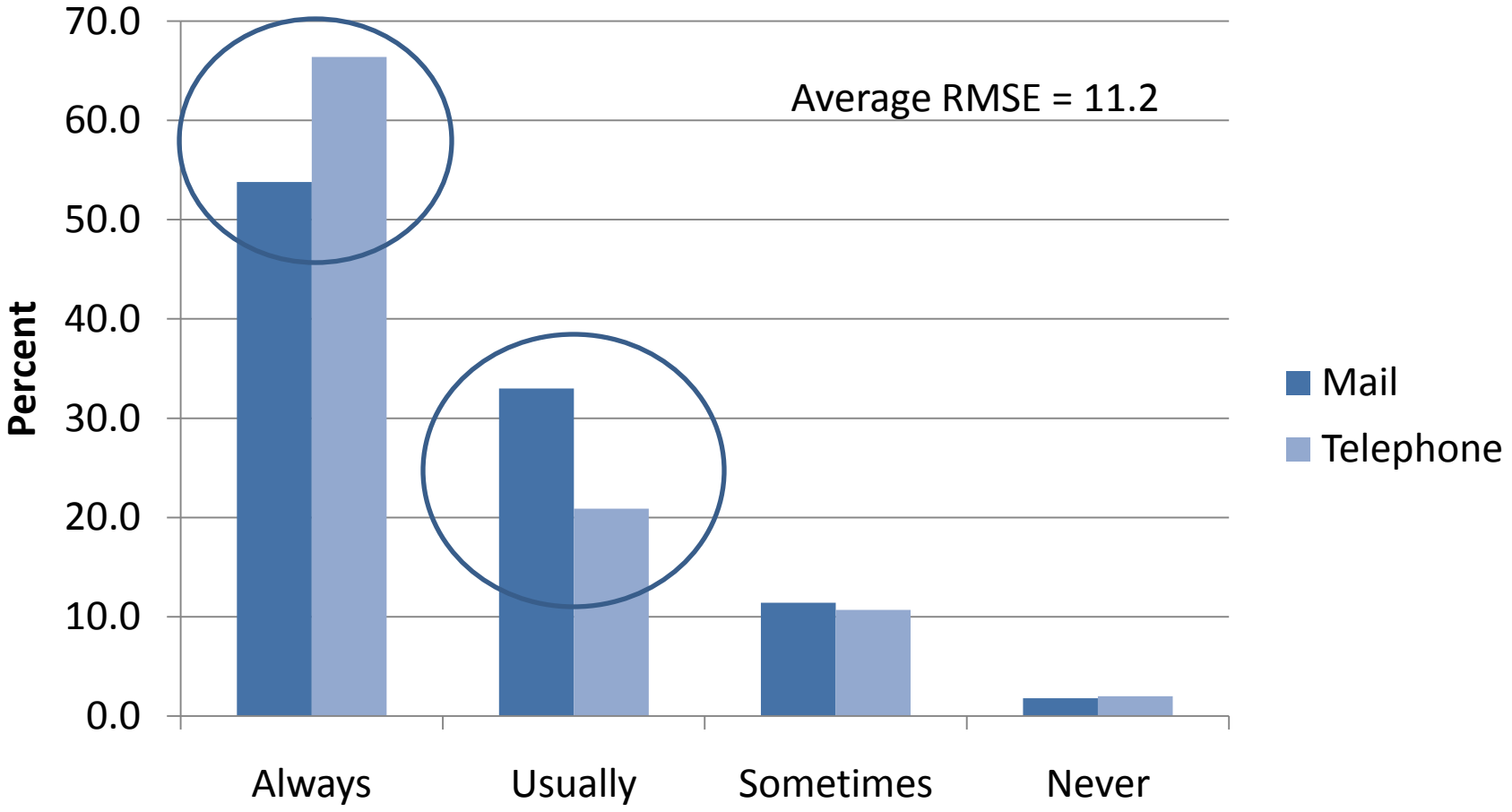


Based on most recent search about health or medical topics: You were concerned with the quality of the information



Difference between modes is significant at p<.02

How often did they make sure you understood the things you needed to do to take care of your health?



Difference between modes is significant at p<.02

Average RMSE for Partially Closed Questions on the Telephone Asking for Quantitative Information