

Reduction of Survey Length through Split Questionnaire Design, and Implications for Nonresponse and Measurement Error

Andy Peytchev, University of Michigan

WSS Workshop of Survey Nonresponse

April 3, 2017

Acknowledgments and Disclaimer

Collaborators: Emilia Peytcheva and Trivellore Raghunathan

This material is based upon work supported by the National Science Foundation under Grant No. 1259985.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Outline

- Part I: Survey length and respondent burden
- Part II: Split Questionnaire Design
- Part III: Survey length and nonresponse
- Part IV: Survey length and measurement error
- Part V: Modification to the Split Questionnaire Design
- Part VI: Preliminary study
 - Design using simulation on existing data
 - Results
- Part VII: Experimental design and procedures

PART I: SURVEY LENGTH AND RESPONDENT BURDEN

Survey Length

- Survey researchers have to balance the need for variable-rich data and reduction of burden due to survey length
- A number of major federal surveys conduct interviews that take up to an hour to administer, on average, or longer (e.g., NHIS, CPS, CES, NCVS, NSFG, NSDUH, NIS)

Survey Length in Self-Administered Surveys

- Despite efforts to limit length, demand for data is often independent of survey mode.
 - About 30 minutes for the National Postsecondary Student Aid Study, Web instrument
 - About 40 minutes for the American Community Survey, both mail and Web instruments
- Perceived burden related to survey length may be different in self-administered surveys, and may be exhibited in a different manner

PART II: SPLIT QUESTIONNAIRE DESIGN

Split Questionnaire Design (Raghunathan and Grizzle, 1995)

- Main objective: shorten the survey instrument to reduce respondent burden while maintaining a rectangular dataset with all survey variables
- Extension of the multiple matrix sampling design (Shoemaker, 1973 and Munger and Lloyd, 1988)

Split Questionnaire Design

- Divide questionnaire into modules
- Administer a subset to each sampled individual, while observing all possible combinations of variables (i.e., bivariate associations)
- Multiply impute data for omitted module(s)

		Core	Module A	Module B	Module C
Full qnnre	Group 0	■	■	■	■
		Core	Module A	Module B	Module C
Split qnnre	Group 1	■	■	■	□
	Group 2	■	□	■	■
	Group 3	■	■	□	■

Extending Split Questionnaire Design to a Survey Error Paradigm

- Key focus of studies to date has been the application of SQD to reduce respondent burden, with emphasis on application, including:
 - How to split the questionnaire
 - Efficient estimation methods
 - Few studies have looked at response rates
- **What impact can SQD have on reduction of nonresponse bias and measurement error?**

Common Assumptions in SQD with Respect to Survey Errors - I

- Assumption 1: Covariance structure is independent of module subset
 - Main impetus for splitting the questionnaire as opposed to sampling of questions

Common Assumptions in SQD with Respect to Survey Errors - II

Assumption 2: Covariance structure is independent of survey length

- Measurement properties for variables in Modules B and C will be the same in Group 0 and Group 2.

		Core	Module A	Module B	Module C
Full qnnre	Group 0	■	■	■	■
		Core	Module A	Module B	Module C
Split qnnre	Group 1	■	■	■	
	Group 2	■		■	■
	Group 3	■	■		■

PART III: UNIT NONRESPONSE

Effects of Survey Length on Survey Participation

- A key negative impact of burden related to survey length that has been used to argue for shorter instruments is increased unit nonresponse
 - Long questionnaires have higher NR rates - e.g., Heberlein and Baumgartner (1978); Adams and Darwin (1982); Dillman, Sinclair and Clark (1993)
 - Especially problematic for surveys of the whole HH
 - 26% of the second members refused to participate in the Cancer Risk Behavior Survey after finding out how long it was (30-50 min)

Select Examples of the Impact of Survey Length on Nonresponse

- 1990 Decennial Census: 4.5 percentage points higher mail return rate for the short form compared to the long form
- Census experiment: 71% response rate for the short and very short forms, 67% response rate for the booklet form (Dillman, Sinclair, and Clark, 1993)
- Just informing that a Web survey is longer, reduces the number of sample members who start it (Crawford, Couper, and Lamias, 2001)
- **But does survey length impact nonresponse bias?**

Sharp and Frankel (1983)

- Study motivated by concern about respondent burden in the longitudinal component of the Annual Housing Survey
- In-person interviews, sample of 886 addresses in Pennsylvania
- Two-factor experiment:
 - Length (25 minutes vs. 75 minutes)
 - Type of survey demand (recall vs. retrieval of records)
- Second wave attempted on 200 addresses

Sharp and Frankel (1983): Nonresponse

Table 1. Final Field Work Results, by Length of Intended Interview: Phase I

	<i>Long Interview (75 Minutes)</i>		<i>Short Interview (25 Minutes)</i>		<i>Total</i>	
	<i>(N)</i>	<i>%</i>	<i>(N)</i>	<i>%</i>	<i>(N)</i>	<i>%</i>
1. Total number of interviewer assignments	(453)	100	(433)	100	(886)	100
2. Completed interview	(250) ^a	55	(250)	58	(500)	56
3. Refusal before interview length was disclosed	(62)	14	(52)	12	(114)	13
4. Refusal after interview length was disclosed	(37)	8	(32)	7	(69)	8
5. Refusal, time unknown	(9)	2	(10)	2	(19)	2
6. Could not contact after 4 calls	(43)	10	(50)	12	(93)	11
7. All other noncompletions (vacant; not a housing unit; language barrier; eligible respondent away; other)	(52)	11	(39)	9	(91)	10

Sharp and Frankel (1983): Indicators of the Quality of Responding

Table 3. Respondent Burden Indicators, by Interview Length (Phase I) (In Percentages)

<i>Indicators</i>	<i>Length</i>		<i>Sig.</i>
	<i>Short</i>	<i>Long</i>	
Respondent reported by interviewer as "pre-occupied" during interview ^a	5	14	$p < .01^b$
Respondent unwilling to be reinterviewed next year	13	27	$p < .01$
Interview was "too long"	13	47	$p < .0001$
Use of shorter questionnaires "would be an improvement"	34	57	$p < .001$
Overall, time and effort put into answering questions was "not very well spent"	5	14	$p < .05^b$
Overall interview was "not very interesting" or "not at all interesting"	13	13	ns
Survey was "not important"	10	15	ns
On the whole, answering the questions was "hard" or "very hard"	4	6	ns
Average item nonresponse rate	2.7	3.0	ns

PART IV: MEASUREMENT ERROR

Survey Length and Measurement Error

- Preceding questions can change the functioning of questions that follow—e.g., Schwarz, Strack, and Mai (1991)
- Responses can change as a function of survey length—Peytchev (2007)

Recall Assumption 1

- Covariance structure is independent of module subset
 - Main impetus for split questionnaire design as opposed to sampling of questions

National Survey of Drug Use and Health (NSDUH)

- National samples since 1971, sampling design also for state-level estimates since 1999, sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA)
- In-person interviews of the civilian, noninstitutional population, age 12 and older
- About 67,500 respondents per year
- 2002: survey name change from the National Household Survey of Drug Abuse (NHSDA), introduction of \$30 respondent incentive
- 2003: instrument changes

Trying Marijuana through 2002 in the National Survey of Drug Use and Health

How do you feel about **adults smoking one or more packs of cigarettes per day?**

1. Neither approve nor disapprove
2. Somewhat disapprove
3. Strongly disapprove

DK/REF

How do you feel about **adults trying marijuana or hashish once or twice?**

1. Neither approve nor disapprove
2. Somewhat disapprove
3. Strongly disapprove

DK/REF

Trying Marijuana from 2003 in NSDUH

Question on smoking was dropped in 2003:

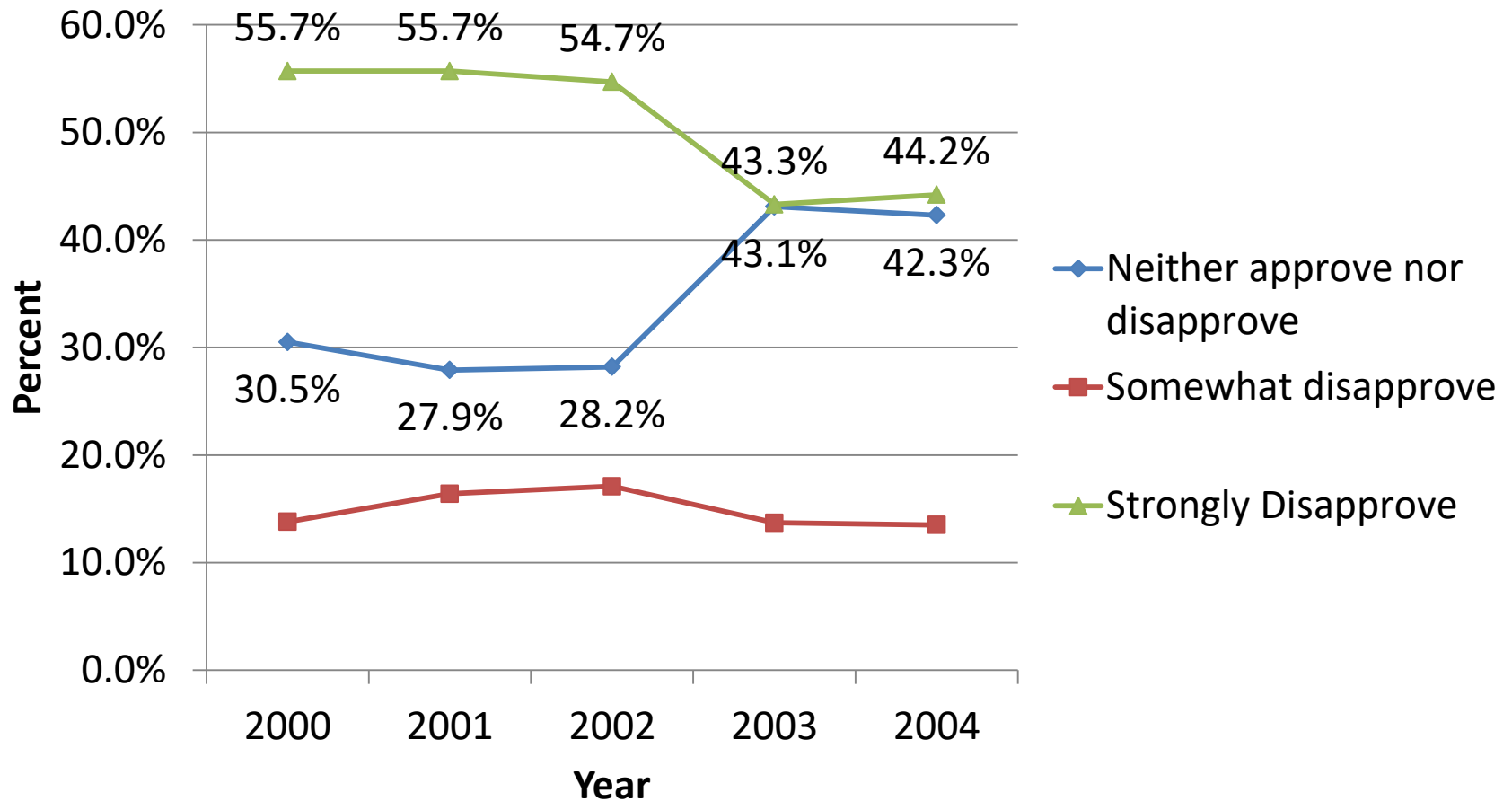
During the past 12 months, how many times have you attacked someone with the intent to seriously injure them?

- | | | |
|-----------------|-----------------|---------------------|
| 1. 0 times | 3. 2 to 5 times | 5. 10 or more times |
| 2. 1 or 2 times | 4. 3 to 9 times | DK/REF |

How do you feel about **adults trying marijuana or hashish** once or twice?

1. Neither approve nor disapprove
 2. Somewhat disapprove
 3. Strongly disapprove
- DK/REF

Attitudes Towards Trying Marijuana, 2000-2004 NHSDA/NSDUH



Data source: Wang, K., R. Baxter, and D. Painter. (2005). Modeling Context Effects in the National Survey of Drug Use and Health (NSDUH). Proceedings of the Joint Statistical Meetings.

Recall Assumption 2

- Covariance structure is independent of survey length
 - Measurement properties for variables in Modules B and C will be the same in Group 0 and Group 2.

		Core	Module A	Module B	Module C
Full qnnre	Group 0	■	■	■	■
		Core	Module A	Module B	Module C
Split qnnre	Group 1	■	■	■	
	Group 2	■		■	■
	Group 3	■	■		■


Experiment Testing the Impact of Survey Length on Measurement Error

- SSI web survey panel members and AOL river sample
- 2,831 started, 2,587 completed
- 18 minutes median time
- 8 questions, 4 on Diet and 4 on Exercise, measuring the two key constructs affecting weight status
 - Programming error with 1 diet and 1 exercise question, thus omitted
- Respondents asked for their height and weight at the end of the survey, used to compute Body Mass Index (BMI)

Experimental Factors

- Location
 - Early in the questionnaire
 - Late
- Layout
 - All on separate pages
 - Same page, but in a list
 - Same page, in a grid
- Instructions (both in an introductory page and as a heading)
 - Related – 8 questions on health
 - Independent – 4 on diet, 4 on exercise
 - No lead-in
- Topic order
 - By topic – diet questions, then exercise
 - Intermixed

Different Pages Layout, Related Instructions



[Frequently Asked Questions](#)
Email us at lif@msisurvey.com
Call toll free 1.866.674.3375

HEALTH

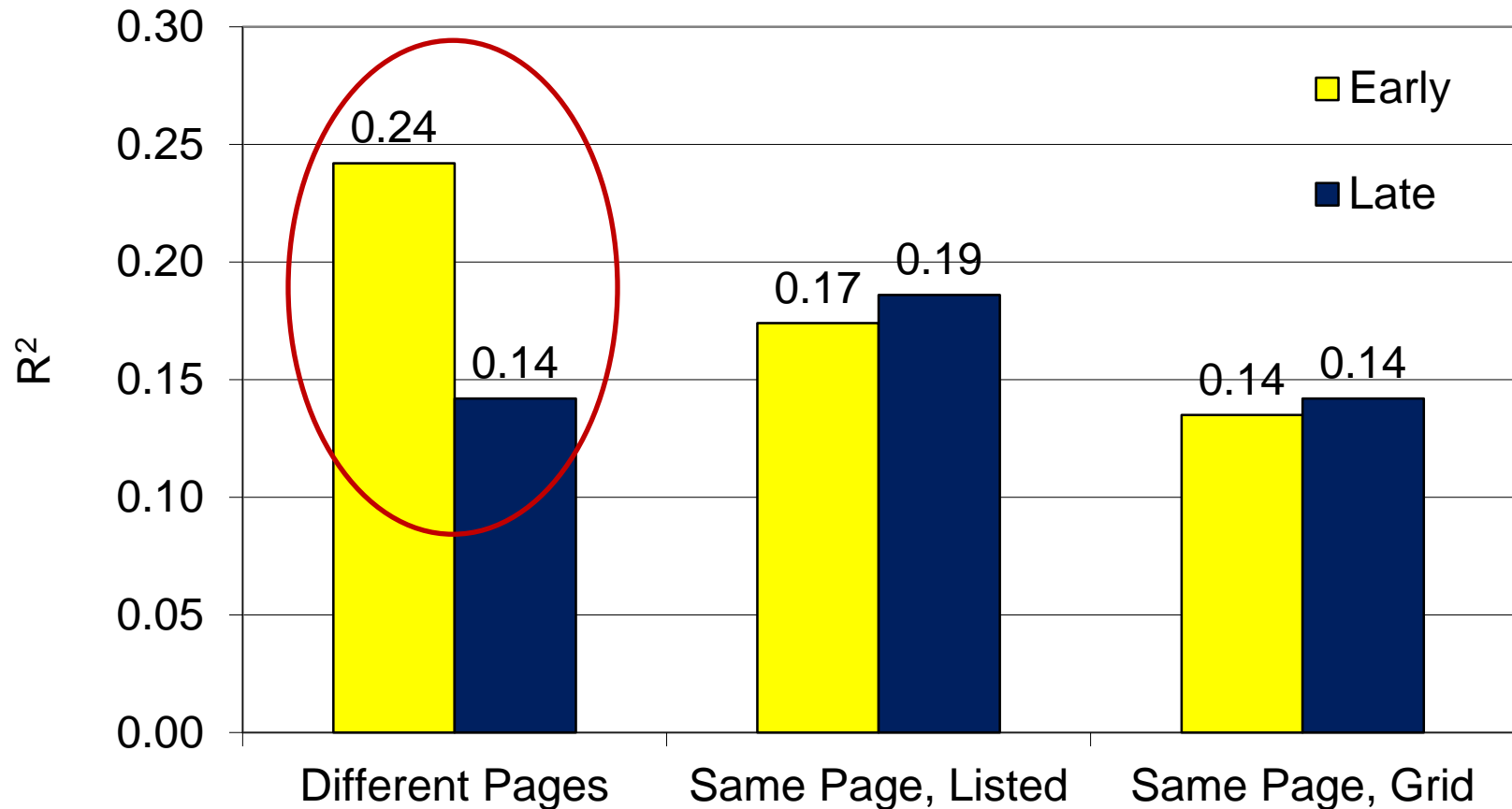
Relative to the average person your age, how often do you eat vegetables?

Far less than average	Less than average	Slightly less than average	Average	Slightly more than average	More than average	Far more than average
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Effect of Survey Length on Measurement Error

- Greater measurement error should attenuate associations with related outcomes
- $BMI = \beta_0 + \beta_1 \text{Diet}_1 + \beta_2 \text{Diet}_2 + \beta_3 \text{Diet}_3 + \beta_4 \text{Exercise}_1 + \beta_5 \text{Exercise}_2 + \beta_6 \text{Exercise}_3 + \varepsilon$
- Fit model separately by Location in the survey instrument (early vs. late) and by Layout
- Compare the proportion of the variability in BMI being explained by the Diet and Exercise questions

Ability of the Diet and Exercise Questions to Explain Variability in BMI



Source: Peytchev, A. (2007). Participation Decisions and Measurement Error in Web Surveys. Doctoral Dissertation, University of Michigan.

PART V: IMPLICATIONS FOR SPLIT QUESTIONNAIRE DESIGN

Implications for Split Questionnaire Design

- Covariance structure can be affected by survey length, thus:
 - Mixing data from a full instrument with reduced instrument without additional adjustments may be undesirable
 - Potential for measurement error provides additional motivation for reduction of survey length
 - Split questionnaire designs may need to not only randomly assign *which* modules, but also the *order* of the modules

PART VI: PRELIMINARY STUDY

Acknowledgment

- Special appreciation is given to Roger Tourangeau, Mick Couper, Fred Conrad, and Reg Baker for allowing these data to be used in the presented analyses.
- Collection of these data was supported in part by grants from the National Science Foundation (Grant SES-0106222) and the National Institutes of Health (Grant R01 HD041386-01A1) awarded to the above.

Data and Methods

- SSI web survey panel members and AOL river sample
- 2,831 started, 2,587 completed
- 18 minutes median time

- 8 questions, 4 on Diet and 4 on Exercise, measuring the two key constructs affecting weight status
 - Programming error with 1 diet and 1 exercise question, thus omitted
- Several manipulations, two of which:
 - Location of questions in the questionnaire (earlier vs. later)
 - Layout (one per page, four per page listed, four per page grid)
 - Main effects and interactions detected, thus using only one per page
- Respondents asked for their height and weight at the end of the survey, used to compute Body Mass Index (BMI)

Split Questionnaire Design Evaluation

Full Questionnaire, n=369

		Survey Questions					Number of Modules
		Set 1	Set 2	Set 3	Set 4	Set 5	
Respondents	1	Gray	Gray	Gray	Black	Gray	5
	...	Gray	Gray	Gray	Black	Gray	5
	n	Gray	Gray	Gray	Black	Gray	5

Split Questionnaire Design, n=363

		Set 1	Set 4	Set 3	Set 2	Set 5	Number of Modules
Respondents	1	Gray	Checkered	Gray	White	Gray	
	...	Gray	Black	Gray	White	Gray	4
	n	Gray	Black	Gray	White	Gray	4

Legend:

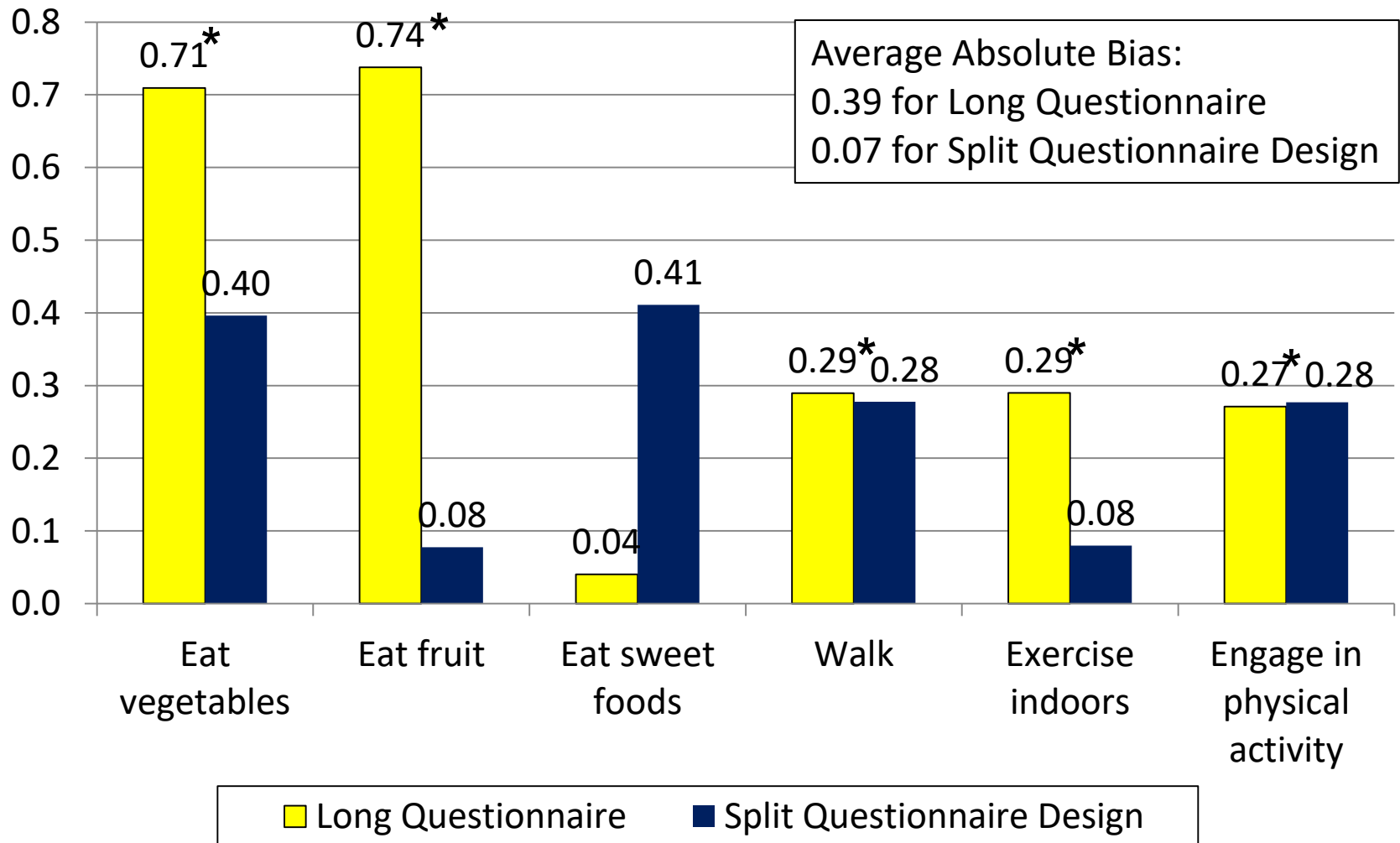
Gray	Survey questions
Black	Experimentally manipulated questions on diet and exercise
White	Deleted data to create the split questionnaire design
Checkered	Deleted and imputed data

Source: Peytchev, A. and E. Peytcheva (under review). "Reduction of Measurement Error due to Survey Length: Evaluation of the Split Questionnaire Design Approach."

Evaluation Criteria

- Bias and Variance
 - No hypothesis for bias for these items, but expect to be lower in the split questionnaire design
 - Possibly larger variance for split questionnaire design
 - MSE
- Criterion validity
 - BMI as a function of the six diet and exercise questions
 - Increased measurement error should lead to lower R^2
- Key research question: Can the split questionnaire design improve measurement properties of data relative to asking all respondents (later in the instrument)?

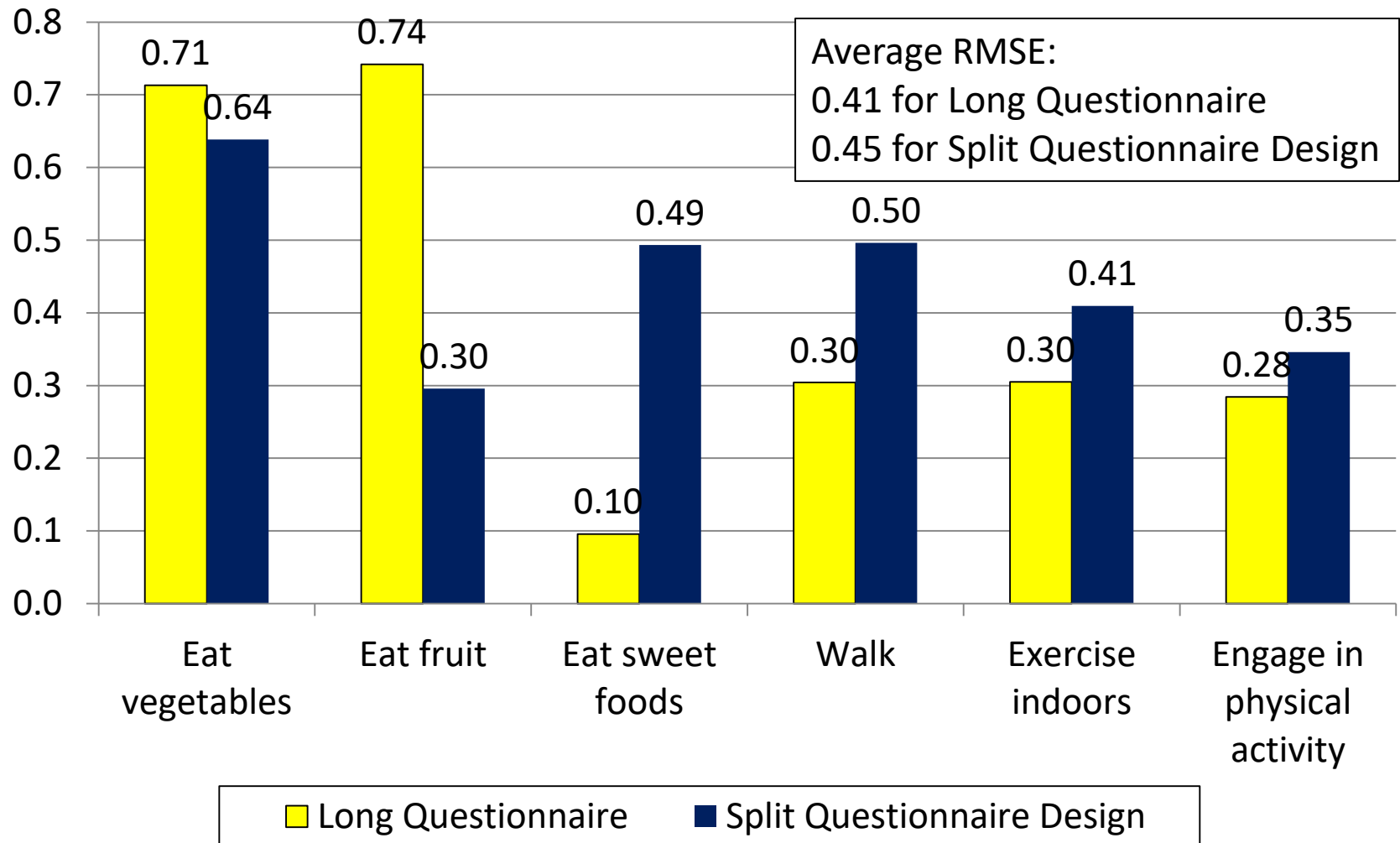
Results: Absolute Bias



* Bias significant at $p < .05$;

Measurement on a 7-point scale where 1=much less than I should and 7=much more than I should.

Results: Total Error (RMSE)



* Bias significant at $p < .05$;

Measurement on a 7-point scale where 1=much less than I should and 7=much more than I should.

Results: Strength of Expected Associations

	Gold Standard (CI)	Long Questionnaire (CI)	Split Quest. Design (CI)	Split Quest. Design [Multinomial Logistic Regr.] (CI)
Correlation of Eating Vegetables with Eating Sweet Foods	-0.29 (-0.38, -0.20)	-0.15 (-0.25, -0.05)	-0.22 (-0.33, -0.10)	-0.25 (-0.64, 0.24)
Square Multiple Correlation of the Six Questions with BMI (R²)	0.25	0.17	0.18	0.22

Source: Peytchev, A. and E. Peytcheva (under review). "Reduction of Measurement Error due to Survey Length: Evaluation of the Split Questionnaire Design Approach."

Conclusions

- Using a split questionnaire design to reduce measurement error due to survey length can:
 - Yield lower biases relative to a long questionnaire
 - Produce similar RMSE (despite the relatively small sample sizes)
- Split questionnaire design produced data with less measurement error (i.e., higher criterion validity) than using a full length survey design
- Shows promise in reducing
 - respondent burden
 - survey error
 - survey cost (depending on data collection design)

PART VII: DESIGN OF THE CURRENT STUDY

Study Objectives

1. Identify whether measurement error is increased as a function of survey length.
2. Explore the impact of survey length on nonresponse rates and on nonresponse bias in these measures, separating nonresponse bias from measurement error.
3. Evaluate the reduction of nonresponse and measurement error bias and impact on mean square error from using split questionnaire design, after multiply imputing the full data for all respondents
4. Extend the approach to include semi-parametric and nonparametric imputation methods and evaluate the effect on preserving variable distributions.

Study Design

- Address-based national sample
- Mail invitations to a Web survey
- Instrument based on NHIS, GSS, and ANES
- Two-phase data collection
- Four experimental conditions, manipulating
 - Survey length
 - Order of modules
- Embedding a variety of measures to evaluate differences in responding across conditions

Experimental Design

Exper. Group	Survey Instrument, Minutes					Expected Length
	5	10	10	10	5	
1	Core health questions	First module	Second module	Third module	Core demog. & biometric qns	40 minutes
2	Core health questions	Third module	Second module	First module	Core demog. & biometric qns	40 minutes
3	Core health questions	First module	Omitted	Omitted	Core demog. & biometric qns	20 minutes
4	Core health questions	Third module	Omitted	Omitted	Core demog. & biometric qns	20 minutes

Current Activities

- Designed the questionnaire modules
- Currently in programming, data collection to start in May
- Expect to have results for the European Survey Research Association conference in July

Thank You

Contact information:

Andy Peytchev

Research Assistant Professor

Survey Methodology Program

Institute for Social Research

University of Michigan

andrey@umich.edu

(734) 647-4619