

REVISITING SAMPLE FRAME AND MODE EFFECTS: A COMPARISON OF POINT ESTIMATES

Nancy A. Mathiowetz, University of Wisconsin-Milwaukee

J. Michael Brick, Westat

Sarah Cho and **Jon Cohen**, SurveyMonkey

Ruth Igielnik, **Scott Keeter**, **Andrew Mercer**, and **Kyley McGeeney**,
Pew Research Center

PewResearchCenter



Research questions of interest

- Pew Research Center: Project to examine the effects of sample and the impact of adjustment methods
- Today's paper: To what extent do commonly used designs lead to divergent or similar estimates?
- Are findings from the past literature still evident?

Previous research.....

- Yeager, Krosnick, Chang, Javits, Levendusky, Simpser, and Wang (2011)
 - Estimates from non-probability sample always less accurate
- Ansolabehere and Schaffner (2014)
 - Opt-in panels *can* produce data that looks similar to landline/cell telephone survey
 - Correlational structures similar to benchmarks, regardless of sample and mode
- Callegaro, Villar, Yeager, and Krosnick (2014)
 - Review of point estimates, relationships across variables, and reliability points to issues in data quality with non-probability online polls
 - Exception: pre-election polls

Pew Research Center Design

- American Trends Panel
 - Probability-based panel (RDD and Cell phones)
 - Data for this study come from the July 2014 mode comparison
 - Web panelists randomly assigned to either Web or telephone mode
 - Field work completed by Abt/SRBI
 - N = 1509 (from 2345 request), 64% RR1
 - Cumulative response rate = 3.7%
 - Weighting
 - Weighted to national parameters, despite lacking coverage of the non-web population

SurveyMonkey

- SurveyMonkey Audience
 - Recruited to panel from 2 million daily responses to user-created SurveyMonkey surveys in the U.S.
 - Base = 4 million individuals
 - Incentive: \$0.50 donated to a charity of choice
- Conducted September-October, 2014
- 5301 responses
- Basic weight
 - Weighted to national parameters, despite lacking coverage of the non-web population

Westat ABS Mail Survey Design

- Addressed based sample, mail questionnaire
- 2,668 total completes
 - AAPOR RR3=30.8% (assumes 88% eligible, consistent with Census vacancy rates); RR1= 28.2%
- Weighting
 - Base weights adjusted for nonresponse using variables from sampling frame
 - Adjustment for number of adults in HH
 - Final raking to population totals: presence of child, crossed with race and home ownership; marital status; presence of child crossed with age and sex of adult; race crossed with education and sex; home ownership crossed with education; race crossed with CPO household status; rased crossed with region

Questionnaire Items

- Omnibus questionnaire
 - Questionnaires are not identical, but many of the questions are identical
 - Item selection: expected differences across samples; benchmark availability; of particular interest to one or more of the collaborators
 - Items
 - Volunteering
 - Internet use
 - Health measures –Dr. Visits, Health status, Every smoked, Can't afford Dr.
 - Tenure at address
 - Food insecurity

Benchmarks

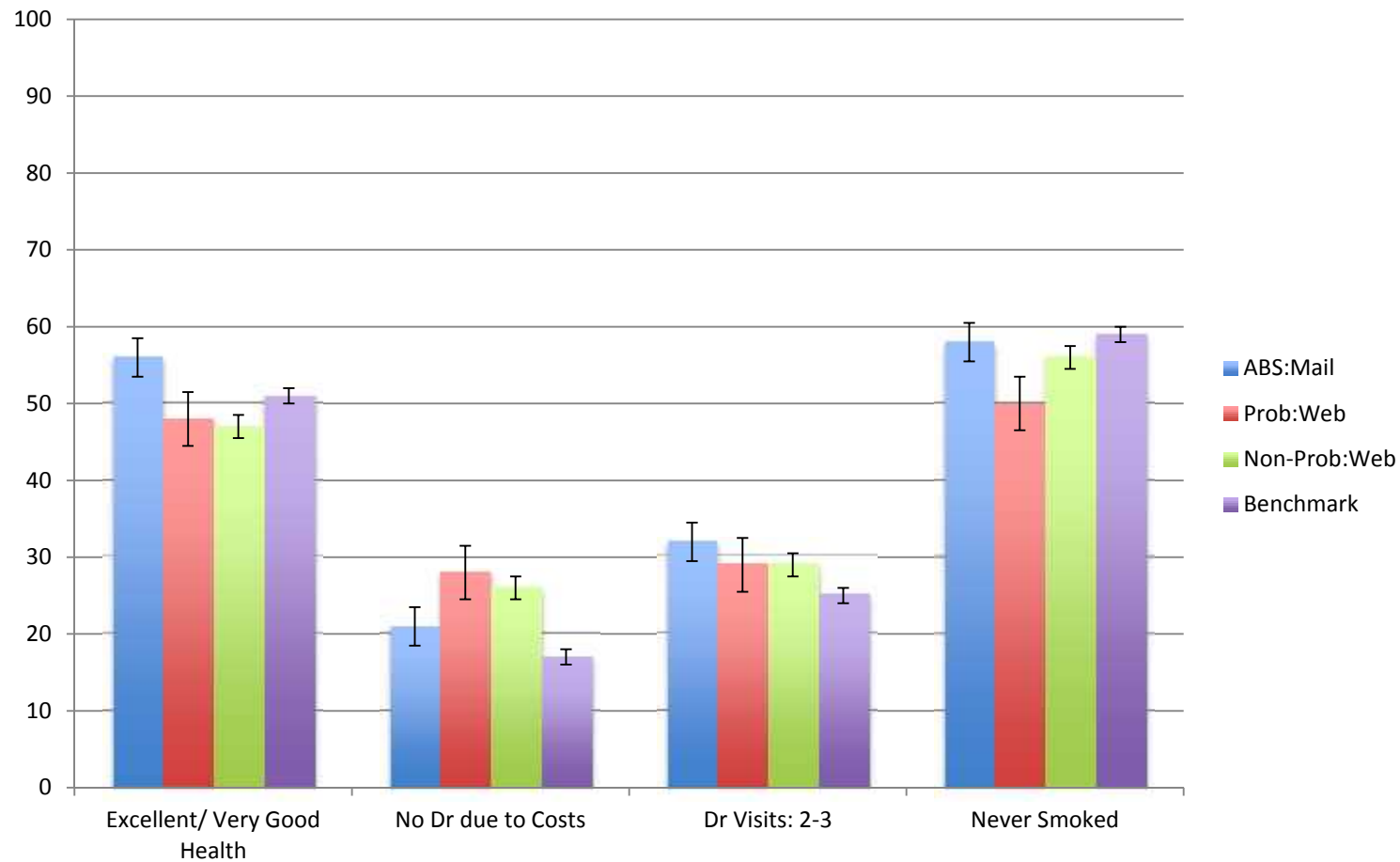
- American Community Survey
- Current Population Survey
 - Internet Supplement
 - Supplement on Volunteering
- National Health Interview Survey

PewResearchCenter

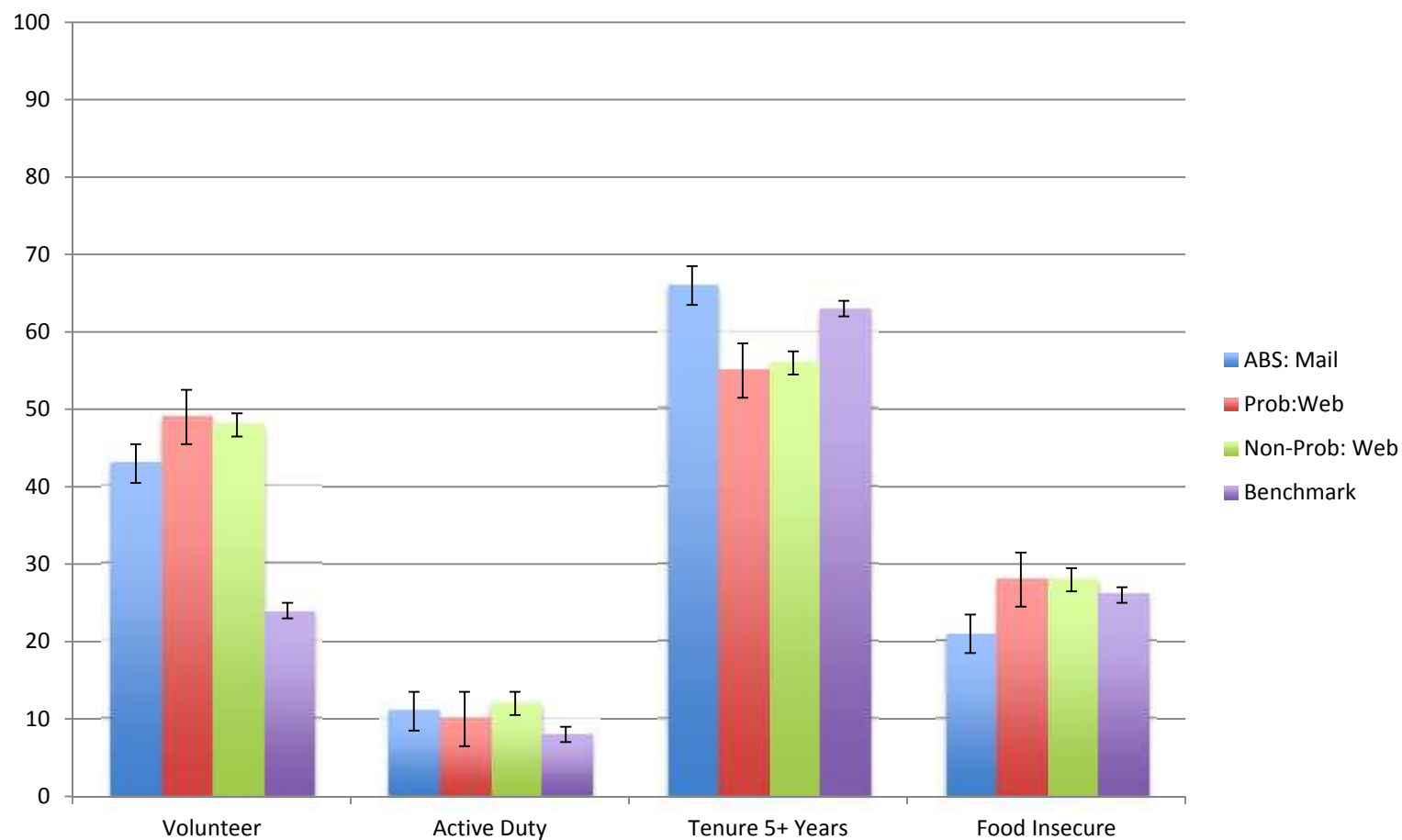

SurveyMonkey

 Westat®

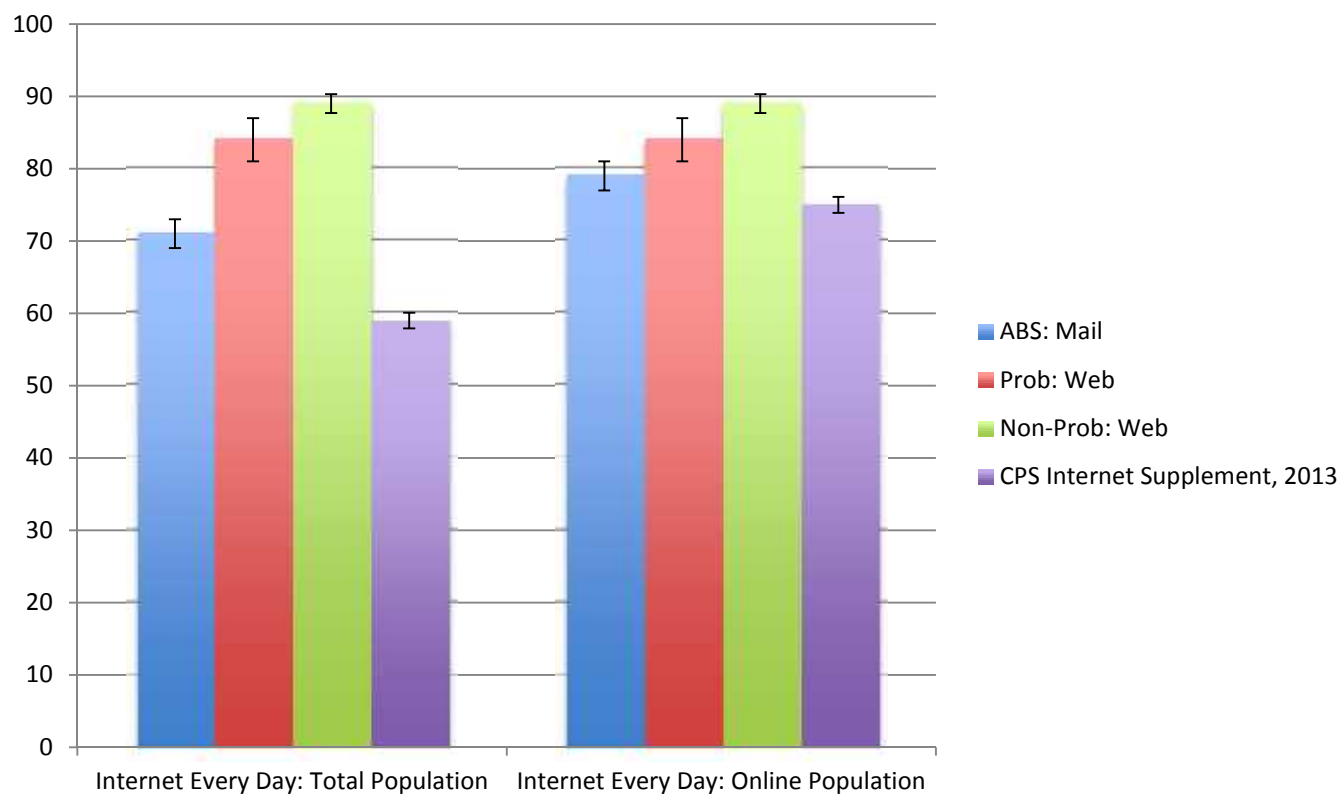
Estimates and Benchmarks: Health Items



Estimates and Benchmarks: Volunteer, Active Duty, Home Tenure, and Food Insecurity



Internet Use Every Day: Expected Differences; Differences not eliminated by weighting



Summary: Point Estimates

- Median Difference between sample and benchmark
 - ABS: 5.4 percentage points
 - Probability-based Web: 9.1 percentage points
 - Non-probability based Web: 6.5 percentage points

Do the samples lead to different predicted probabilities?

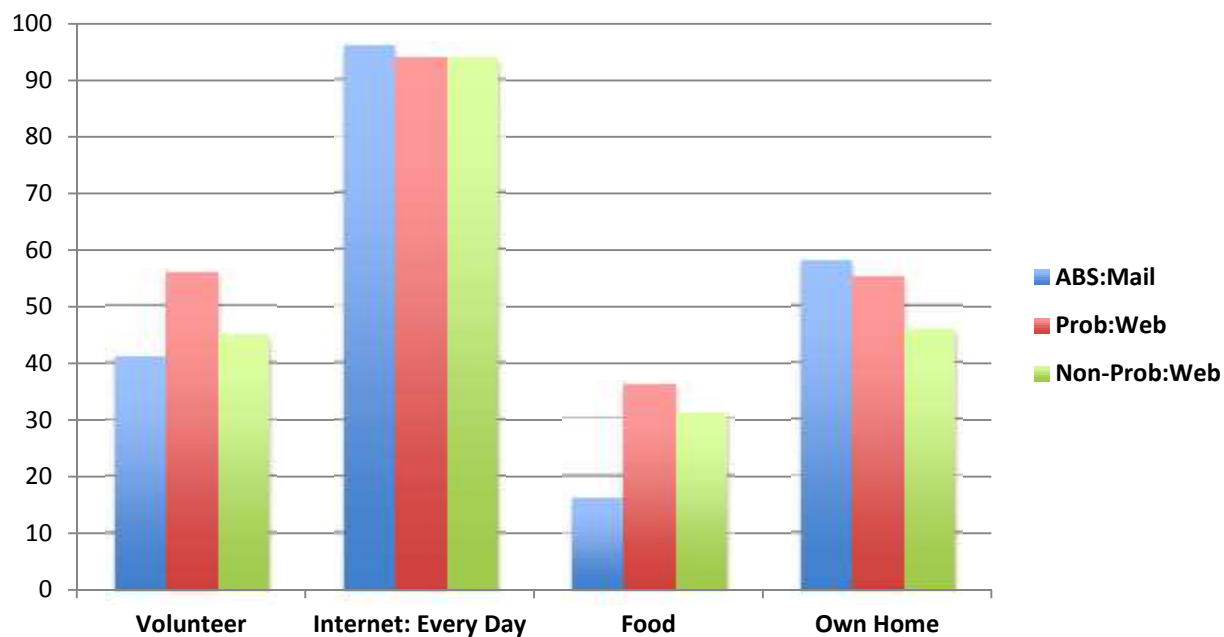
- Logistic regression models for each of the samples
- Independent Variables
 - Gender, age, Race, Education, Income, Currently Employed
 - Not necessarily great models, but consistent!

PewResearchCenter

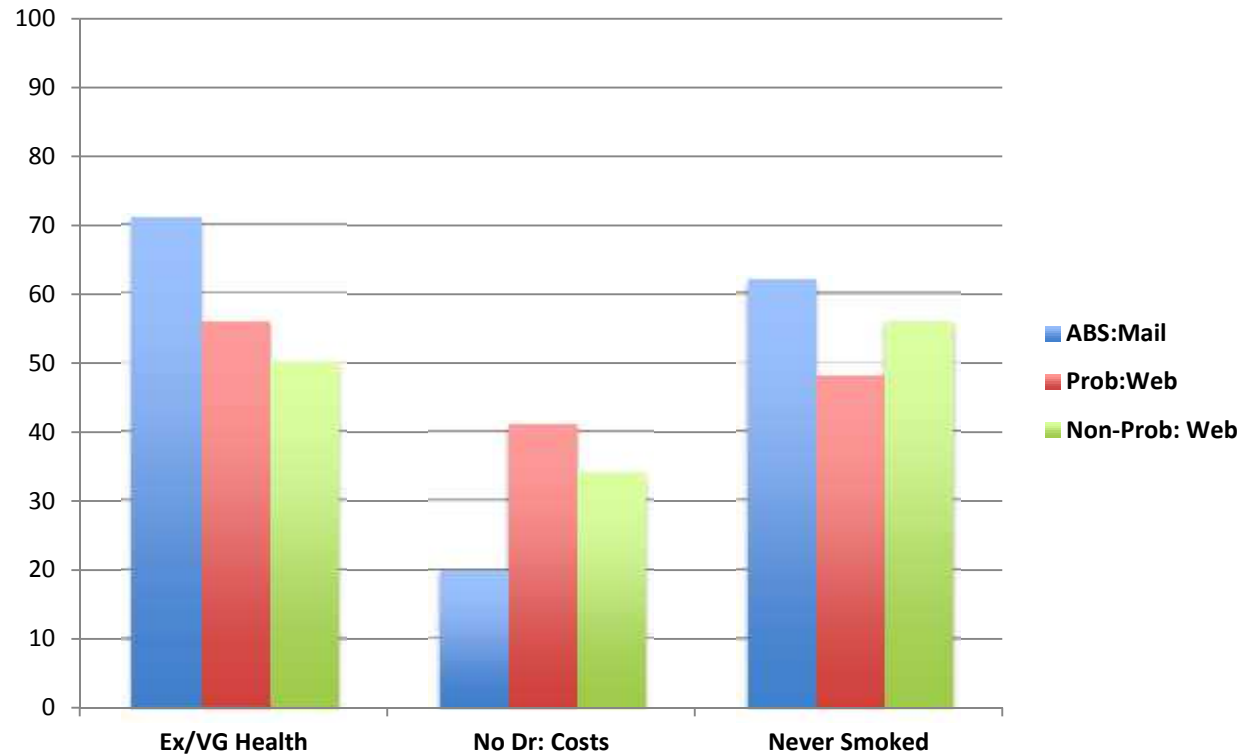

SurveyMonkey

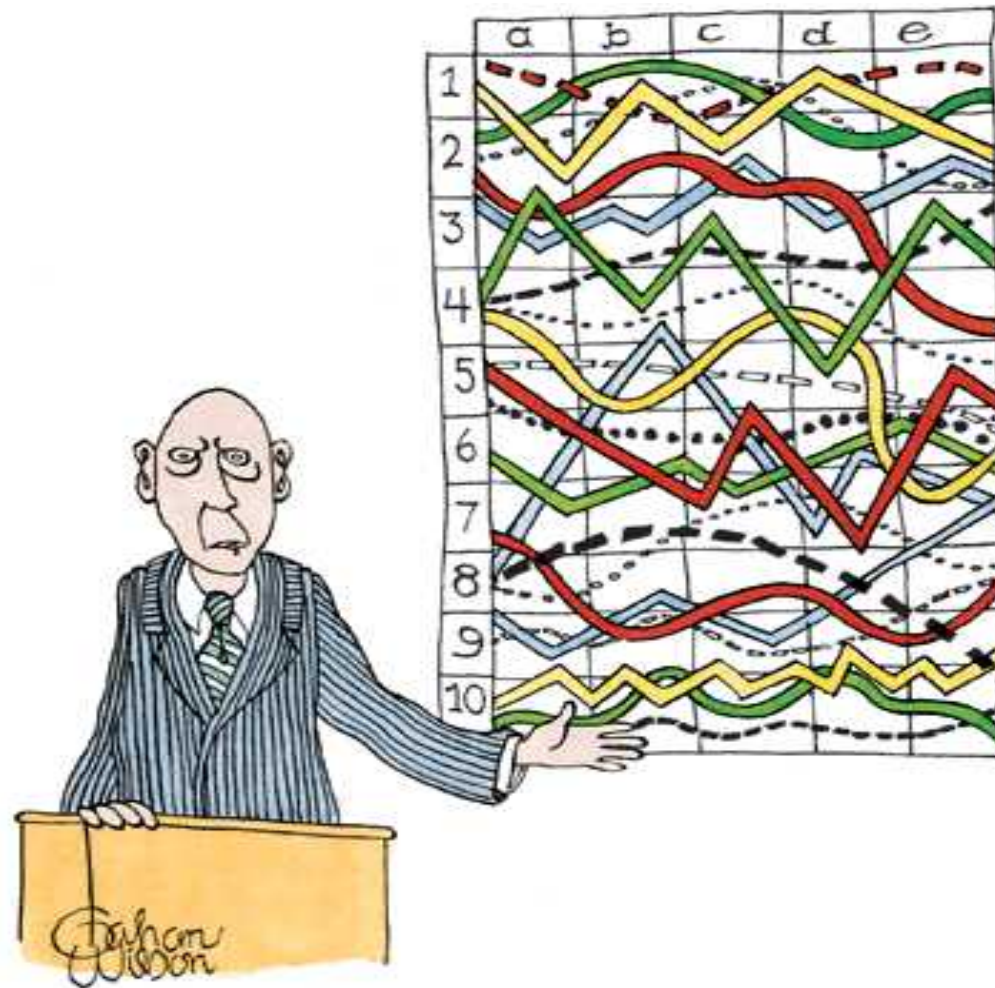
 Westat®

Do the samples lead to different predicted probabilities? One illustrative case.....



Predicted Probabilities—health items





"I'll pause for a moment so you can let this information sink in."

Conclusions

- Point estimates:
 - No one design consistently superior when compared to benchmarks
 - Variability we would expect
 - The findings suggest that estimates from the ABS differs from both the non-probability and the probability web surveys
 - Coverage and mode confounded
- Multivariate models (not shown) show surprisingly few differences with respect to direction or significance
 - Although the size of the coefficients in these models were substantively different—evident in the predicted probabilities

What next?

- First view of the data
 - Additional point estimates
 - Additional models
 - Benchmark models
- Only one representation of each type of sample
 - Additional work ongoing to examine variation across implementations, across non-probability sample vendors and reliability of SurveyMonkey sample
- Saturday sessions
 - Scott Keeter—Mode Effects for the Pew Mode Study
 - Mike Brick—Effects of Alternative Weighting Approaches for Non-Probability Samples