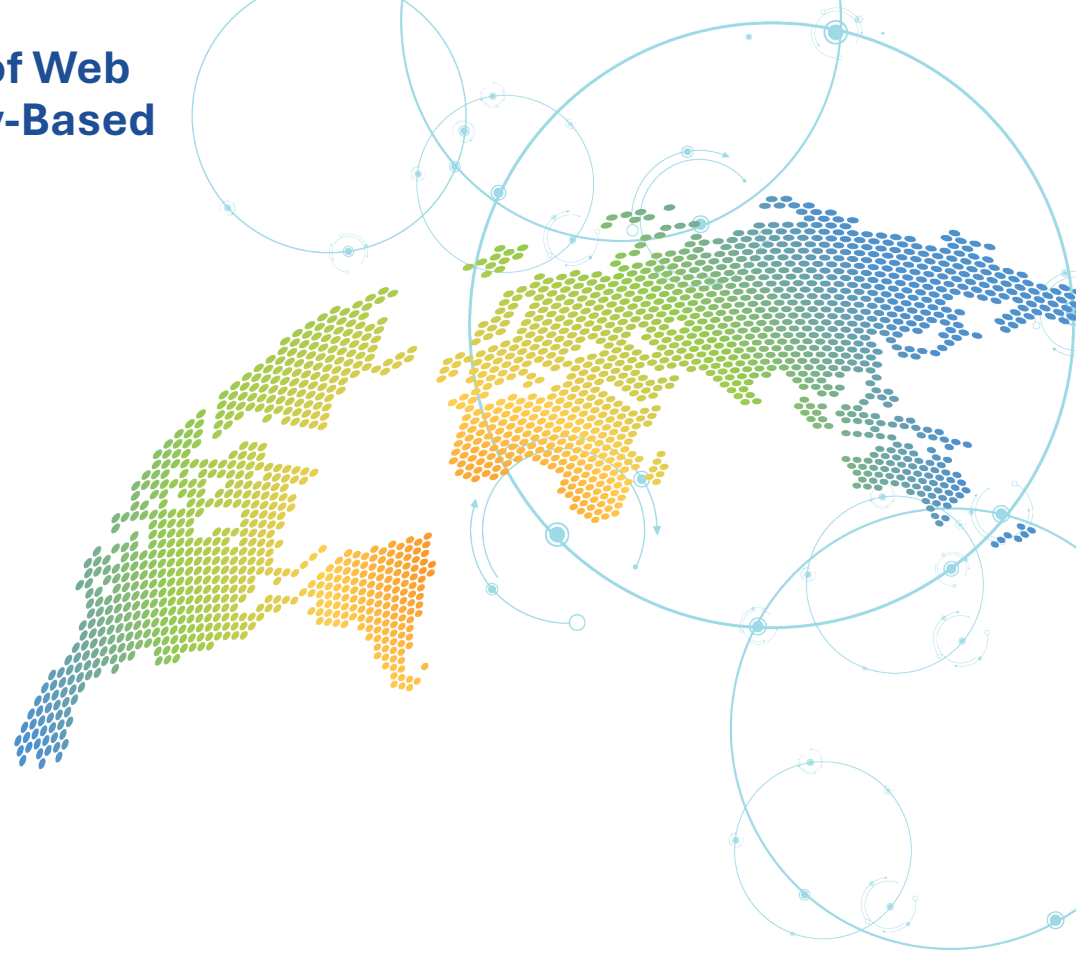


Evaluating Error from the Addition of Web Panel Data to a National Probability-Based Interviewer-Administered Survey

Andy Peytchev, PhD

CIPHER, February 27, 2025
Washington, D.C.



Disclaimer and Acknowledgment

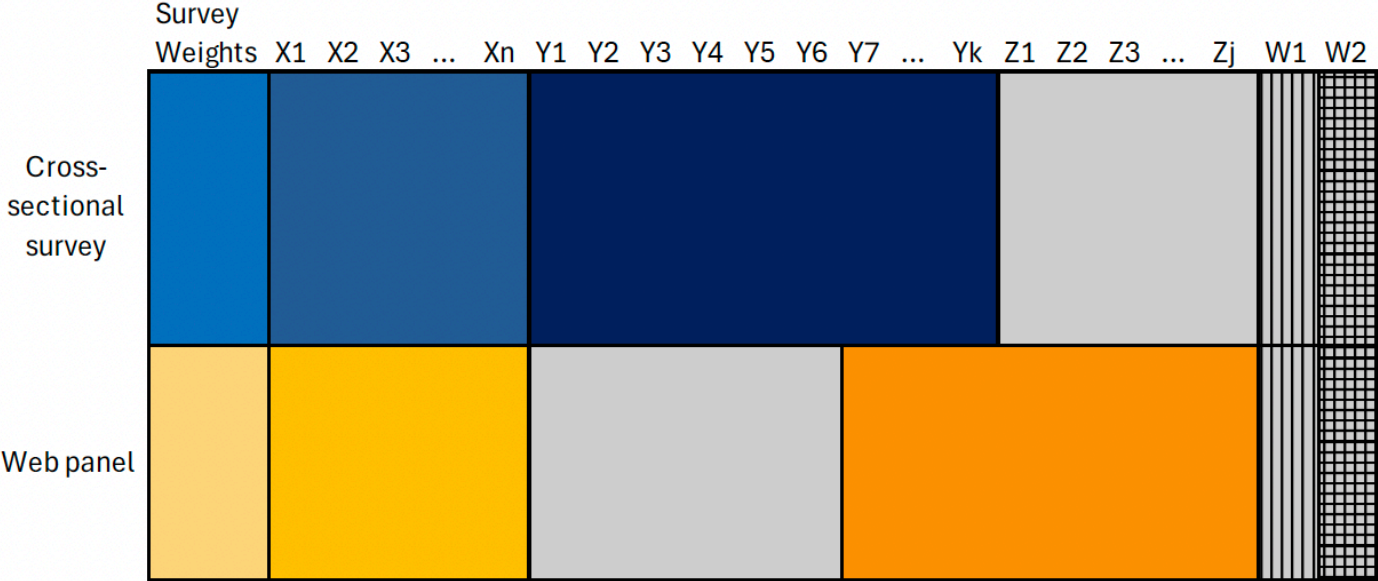
- The author has an advisory role on a contract for National Center for Health Statistics' (NCHS) Rapid Survey System (RSS). The independent research in this presentation was not supported by this contract, used publicly-available data, and does not express the views of the Government.
- The author is grateful to NCHS for making these valuable datasets and accompanying documentation available to researchers, making investigations like this possible and aiding future data collection designs.



A New Paradigm

- Probability-based web panels offer an opportunity to collect data at lower cost (and faster) compared to common cross-sectional household survey designs
 - Collect data on new and emerging topics
 - Track changes
- Can also be used to augment a more costly cross-sectional survey:
 - Adding records (longer dataset)
 - Adding variables (wider dataset)

Data Integration Framework



Xn	Demographic variables
Yk	Substantive variables from core NHIS modules
Zj	Substantive variables not part of NHIS
	Multiple imputation
W1	Combined data survey weights, postratified to socio-demographic controls (X)
W2	Combined data survey weights, postratified to socio-demographic (X) and substantive controls (Y)

Weighting Paradigm: The Model-Assisted Probability-Based Survey Inference

- Socio-demographic characteristics
 - Necessary, but not sufficient
- Voting and volunteering
 - Strong predictors of survey participation (e.g., Peytchev, Presser, and Zhang, 2018)
- Survey topic
 - From another survey (e.g., Westat, 2007)
- Allows to achieve unbiased point estimates relative to benchmark
- Replication-based variance estimation to account for survey-based control totals (e.g., Kott, 2019; Peytchev, 2020)
- *No control over associations that can be affected by nonresponse and measurement differences, yet this is critical to combine data*

NCHS's Rapid Survey System (RSS)

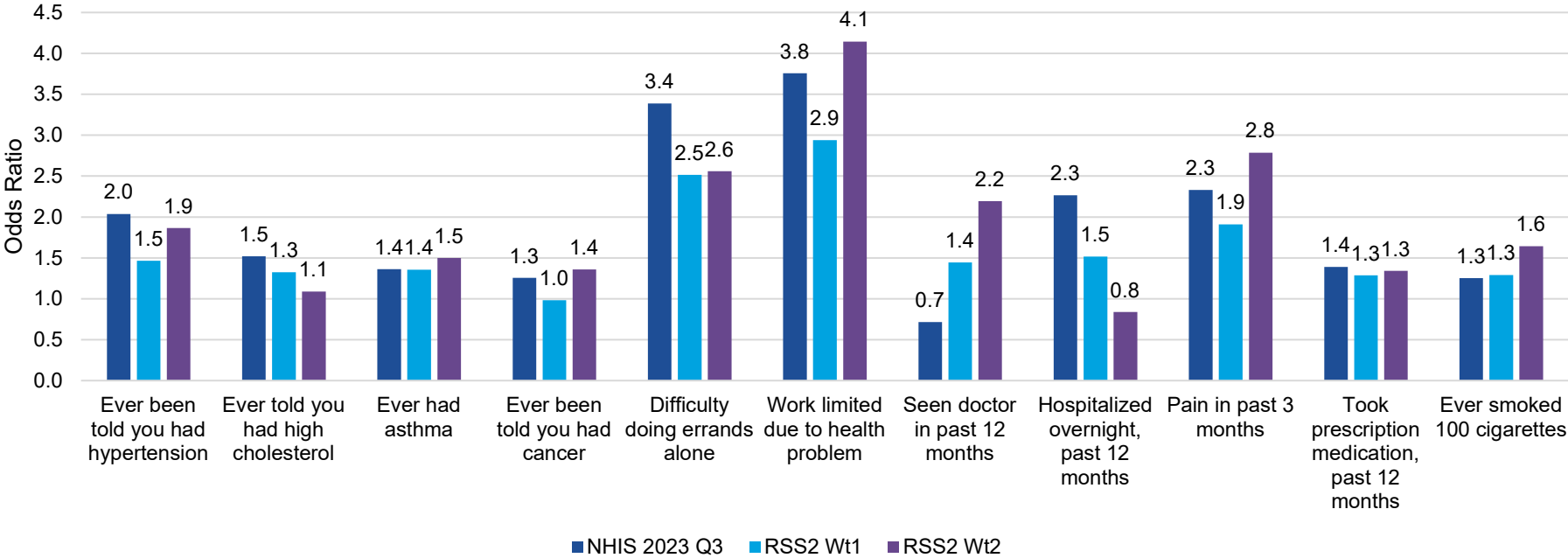
- The instruments are designed to include questions from the National Health Interview Survey (NHIS)
 - Calibration
 - Evaluation
- Two probability-based panel vendors administering the same instruments
- The second round (RSS2) was administered by web (n=6783) with very few by telephone (263), from October 16 to November 6, 2023
- An overall (cumulative) response rate of 4% for both panels

Methods

- Identified 32 variables that are common to NHIS and RSS2
 - Of these, 20 are health-related and 4 are volunteering-related
- For this preliminary analysis, listwise deletion for RSS2 (to be imputed)
- Weight 1. Recalibrated the RSS2 respondent survey weights to NHIS estimates for sex, age (appx. 10-year intervals), race/ethnicity, education, and region
- Weight 2. Added the 24 health- and volunteering-related variables
- Evaluated the impact on bivariate and multivariate associations before and after the more elaborate weighting using topic-related variables
 - Bivariate correlations
 - Logistic regression

Effect on Parameter Estimates in a Multivariate Model

Odds Ratios for Predictors of Self-Rated Poor or Fair Health



Summary and Conclusions

- Probability-based web panels have unusually high reliance on survey weight adjustments for nonresponse
- Widening the net for candidate adjustment variables is necessary despite different measurement error properties for these variables (an assertion)
- By definition, a “wider net” can calibrate panel point estimates of means and proportions but NOT the covariance structure of the data
- Some pairs of variables showed substantial differences in associations in the panel data compared to NHIS, yet many did not
- A logistic regression showed similar results, in a multivariate setting
- Caution should be used if “data stacking” for adding records to a more rigorous survey at a lower cost is considered

Next Steps

- Impute data for all health-related variables to be used in weighting
- Evaluate impact on associations between demographic and health-related variables, as typically used in tabulations
- Incorporate replicate weights
 - Reflects uncertainty in the control totals based on a benchmark survey
 - Reflects the stability of estimates when calibrated to relevant variables
- Repeat the analysis by panel vendor
- Impute RSS variables to the NHIS respondents to create a wide dataset
 - The current preliminary analyses do not support a long (appended) dataset

References

Kott, P. S. (2019). "Integrating the results of a nonresponse follow-up survey into the survey from which its items were selected." Statistical Journal of the IAOS **35**(2): 289-297.

Peytchev, A. (2020). "Split-Sample Design with Parallel Protocols to Reduce Cost and Nonresponse Bias in Surveys." Journal of Survey Statistics and Methodology **8**(4): 748–771.

Peytchev, A., S. Presser and M. Zhang (2018). "Improving Traditional Nonresponse Bias Adjustments: Combining Statistical Properties with Social Theory." Journal of Survey Statistics and Methodology **6**: 491–515.

Westat. (2007). "Health Information National Trends Survey (HINTS) 2007 Final Report." Retrieved from Bethesda, MD: https://hints.cancer.gov/docs/methodologyreports/HINTS_2007_Final_Report.pdf