RADARS® SYSTEM

Researched Abuse, Diversion and Addiction-Related Surveillance System

QUARTERLY Technical REPORT

Fourth Quarter, 2020

Accounting for Bias from Repeat Respondents in the Survey of Non-Medical Use of Prescription Drugs Program

Key Finding

- Drug use is less common among respondents who return to the survey than first time respondents, which introduces bias
- This bias is mitigated by estimating outcomes using a generalized linear mixed model
- A design-based bootstrap was the most statistically conservative method tested to estimate drug use outcomes

Introduction

Online survey panels are an emerging method for estimating health outcomes. Since errors between online panels and traditional surveys compared to errors between different traditional surveys have similar magnitudes[1], validity can be achieved using an online survey panel. Validity, however, depends on limitations being addressed[2]. The Researched Abuse, Diversion and Addiction-Related Surveillance (RADARS®) System Survey of Non-Medical Use of Prescription Drugs (NMURx) Program already corrects for major bias sources by applying careless response exclusions and calibration weights that use health-related variables[3]. Yet estimates can be further improved by addressing other sources of bias. One such source is receiving multiple surveys from a single respondent[4]. The NMURx Program already prevents respondents from providing multiple surveys in a single calendar year. However, respondents may return across multiple calendar years, which potentially introduces a bias when conducting multiyear analyses. This technical report quantifies the bias due to these repeat respondents and demonstrates a mixed effect regression methodology for correcting for the bias.

Methods

Data Sources

The NMURx Program deploys repeated, cross-sectional general population drug use surveys to an online panel twice per calendar year. Full methodological details are provided in previously published work and concurrent validity demonstrated using national probability surveys[3]. The study period for this analysis includes the second half of 2018 (n=29,841), 2019 (n=59,601), and 2020 (n=59,532).Three outcomes were selected to demonstrate the effect different methodologies have on estimates: use in the last 12 months of prescription pain relievers, prescription stimulants, and cannabis. Estimates of use in the last 12 months calculated multiple survey years can be interpreted as the average estimate of last 12 month use across the full study period.

Statistical Analysis

The percentages of respondents in each calendar year that participated in other calendar years were calculated. To quantify bias due to repeat respondents, estimates from 2019 were stratified by whether respondents participated in only 2019 versus respondents who also participated in 2018 or 2020. Three different methods were used to calculate estimates, each with different approaches. First, estimates were calculated using weighted logistic regression with Wald confidence intervals. This method assumes asymptotic normality of the parameters and does not integrate the design into the standard error. Second, estimates were calculated using weighted frequencies with design-based confidence intervals. Standard errors were calculated using a bootstrap technique where the responses were clustered by individual. This method integrates variance between weighting strata into the overall standard error estimate and is a design-based method. Finally, estimates were calculated using a generalized linear mixed model (GLMM) where a random effect was introduced that correlated model residuals from repeated responses from the same individual. An autoregressive covariance structure was specified, which assumes that correlation between responses decay exponentially the further away they are in time. Confidence intervals were calculated using the classical sandwich estimator. This method accounts for variance between weighting strata, adjusts estimates for correlated responses, and provides a robust estimator of standard error.

Results

Prevalence of Repeat Respondents

A majority of respondents took the survey only a single time across the three year study period (Figure 1). A total of 81.4% in 2018, 76.6% in 2019, and 77.0% in 2020 participated in only a single wave. Because only one wave was included in 2018, the overall sample size is smaller, which accounts for the higher percentage that participated in only a single wave.

In 2019, those who took more than one survey wave had lower drug use estimates than those who only took the survey once (Table 1). Estimates among repeat respondents were 7.0 percentage points (pps) smaller for prescription pain reliever use, 4.1 pps smaller for prescription stimulant use, and 10.3 pps smaller for cannabis use.

Figure 1: Percentage of Repeat Respondents. Percentage of respondents who participated in multiple years of data collection. Respondents can only participate once per calendar year.

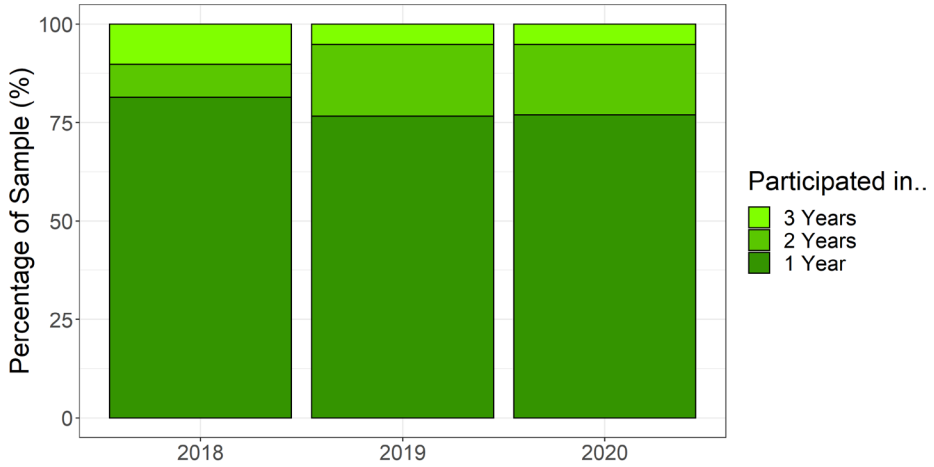


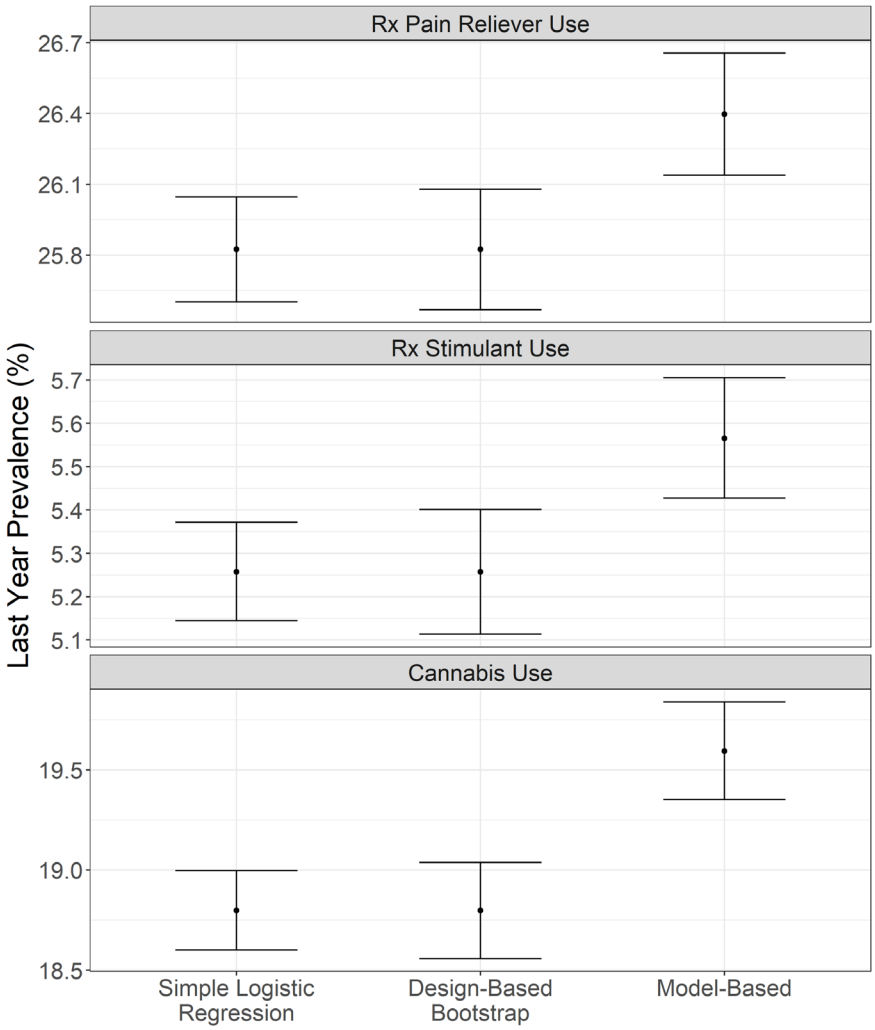
Table 1: 2019 Last 12 Month Drug Use Stratified by Repeat Respondent Status

Respondent Status	Prescription Pain Reliever Use % (95% CI)	Prescription Stimulant Use % (95% CI)	Cannabis Use % (95% CI)
Only participated in 2019	27.7 (27.2, 28.1)	6.0 (5.7, 6.2)	20.8 (20.3, 21.2)
Participated in 2019 and one or both of 2018 and 2020	20.7 (20.0, 21.5)	1.9 (1.7, 2.2)	10.5 (9.9, 11.0)

Effect of Estimation Method

Confidence intervals from simple, weighted logistic regression were the smallest of the three methods, as expected based on the assumptions (Figure 2). Using the design-based bootstrap method, confidence intervals were wider and more conservative than other methods. The estimate itself was identical between weighted logistic regression and design-based estimation, and estimates were larger with the GLMM method.

Figure 2: Prevalence of Drug Use Obtained from Different Estimation Methods. Estimates of past-year prevalence across three calendar years of data collection. Confidence intervals are widest for the design-based bootstrap method. GLMM: generalized linear mixed models.



Conclusions

Different methods to estimate prevalence produced slightly different results and statistical confidences. The model-based method accounted for bias from respondents returning to the survey, which is best suited for estimates of trends in prevalence. Because differences were generally small, design-based approaches should be considered when possible and are likely sufficient for annual estimates used in on-going surveillance. Targeted hypothesis testing utilizing multiple years of data collection might require the more customized model-based approach if overlap is substantial. Changes to estimates and confidence intervals observed here were small, but this is likely a consequence of the large sample size for national prevalence estimates. Domain analyses could be more affected.

Though the model-based approach accounted for many confounding elements of the study, limitations do exist for this approach. Model convergence is not guaranteed, and multiple models may need to be specified for a study with many outcomes. One key limitation of this study is that an analysis such as this would change based on the number of years included. Since the NMURx Program has only been collecting data for 3 years, the effect of repeat respondents on longer time scales cannot be quantified yet.

Suggested Citation

Black JC, Rockhill K, Amioka E, Forber A, Iwanicki J, Dart RC (2020): Accounting for bias from repeat respondents in the Survey of Non-Medical Use of Prescription Drugs Program. RADARS® System Technical Report, 2020-Q4.

References

[1]Unangst J, Amaya AE, Sanders HL, Howard J, Ferrell A, Karon S, Dever JA. "A process for decomposing total survey error in probability and nonprobability surveys: A case study comparing health statistics in US Internet panels." Journal of Survey Statistics and Methodology. 2020; 8(1):62-88.

[2]Cornesse C, Blom AG, Dutwin D, Krosnick JA, De Leeuw ED, Legleye S, Pasek J, Pennay D, Phillips B, Sakshaug JW, Struminskaya B. A review of conceptual approaches and empirical evidence on probability and nonprobability sample survey research. Journal of Survey Statistics and Methodology. 2020; 8(1):4-36.

[3]Black JC, Rockhill K, Forber A, Amioka E, May KP, Haynes CM, Dasgupta N, Dart RC. "An Online Survey for Pharmacoepidemiological Investigation (Survey of Non-Medical Use of Prescription Drugs Program): Validation Study." J Med Internet Res 2019; 21(10):e15830

[4]Hays RD, Liu H, Kapteyn A. Use of Internet panels to conduct surveys. Behavior research methods. 2015; 47(3):685-90.