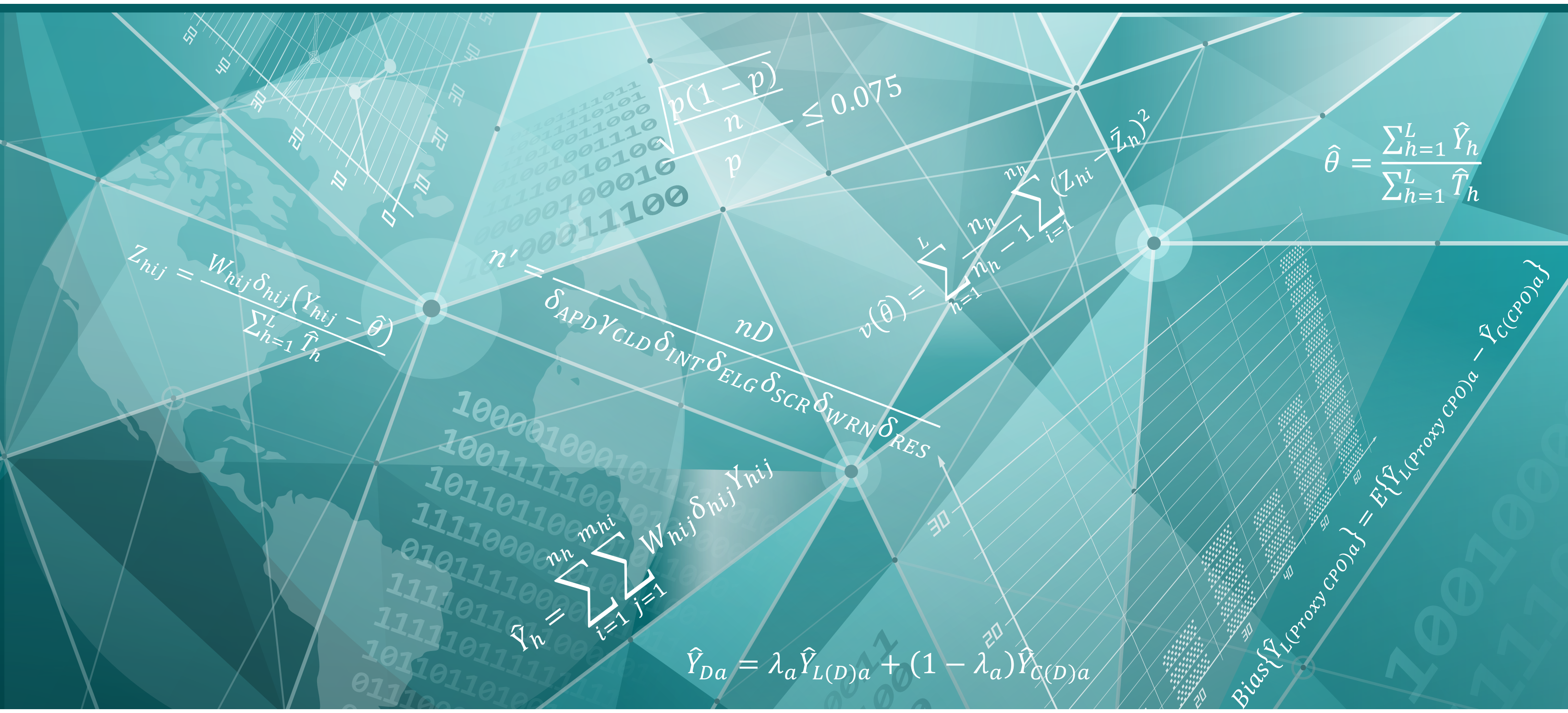
NATIONAL CENTER FOR HEALTH STATISTICS

Vital and Health Statistics

Series 2, Number 199 March 2023



Calibration Weighting Methods for the National Center for Health Statistics Research and Development Survey

Data Evaluation and Methods Research

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

Suggested citation

Irimata KE, He Y, Parsons VL, Shin H-C, Zhang G. Calibration weighting methods for the National Center for Health Statistics Research and Development Survey. National Center for Health Statistics. Vital Health Stat 2(199). 2023. DOI: https://dx.doi.org/10.15620/cdc:123463.

For sale by the U.S. Government Publishing Office
Superintendent of Documents
Mail Stop: SSOP
Washington, DC 20401–0001
Printed on acid-free paper.

NATIONAL CENTER FOR HEALTH STATISTICS

Vital and Health Statistics

Series 2, Number 199 March 2023

Calibration Weighting Methods for the National Center for Health Statistics Research and Development Survey

Data Evaluation and Methods Research

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

Hyattsville, Maryland
March 2023

Contents

Appendix Figure

Appendix Tables

Calibration Weighting Methods for the National Center for Health Statistics Research and Development Survey

by Katherine E. Irimata, Ph.D., Yulei He, Ph.D., Van L. Parsons, Ph.D., Hee-Choon Shin, Ph.D., and Guangyu Zhang, Ph.D.

Abstract

Background

The Research and Development Survey (RANDS) is a series of web-based, commercial panel surveys that have been conducted by the National Center for Health Statistics (NCHS) since 2015. RANDS was designed for methodological research purposes, including supplementing NCHS' evaluation of surveys and questionnaires to detect measurement error, and exploring methods to integrate data from commercial survey panels with high-quality data collections to improve survey estimation. The latter goal of improving survey estimation is in response to limitations of web surveys, including coverage and nonresponse bias. To address the potential bias in estimates from RANDS, NCHS has investigated various calibration weighting methods to adjust the RANDS panel weights using one of NCHS' national household surveys, the National Health Interview Survey. This report describes calibration weighting methods and the approaches used to calibrate weights in web-based panel surveys at NCHS.

Keywords: web panel survey • raking• propensity score weighting • weight adjustment

Introduction

The Research and Development Survey (RANDS) is a series of primarily probability-sampled, commercial panel surveys conducted at the National Center for Health Statistics (NCHS). Unlike other data systems at NCHS, including population health surveys, provider surveys, and vital statistics, RANDS is primarily web-based and was designed initially for methodological research purposes. Since the RANDS program started in 2015, data from RANDS have been used to evaluate questionnaire designs and for estimation research, including the study of possible estimation bias in web-based panel surveys. However, during the COVID-19 pandemic, the special series RANDS during COVID-19 was used to publicly release experimental estimates to rapidly report on the impact of the COVID-19 pandemic (1). RANDS questionnaires focus on health outcomes and have included a variety of questions, including ones about general health, mental health, health insurance, chronic conditions, opioids, disability, and, more recently, COVID-19. More information on the completed rounds of RANDS can be found on the NCHS web page: https://www.cdc.gov/nchs/rands/index.htm.

Although RANDS is an NCHS survey, it is collected through commercial panels that are maintained by external organizations. For example, RANDS 1 and RANDS 2 were conducted using the Gallup Panel (https://www.gallup.com/topic/gallup_panel.aspx) maintained by Gallup, Inc., while RANDS 3 and later rounds, including the three rounds of RANDS during COVID-19, were conducted using the AmeriSpeak Panel (https://amerispeak.norc.org/) maintained by NORC at the University of Chicago (NORC). These panels are multimode and probability-based. While opt-in, or nonprobability surveys, do not have known sampling weights, probability-based panels involve probability sampling to select panelists, and sampling weights are calculated according to the sampling procedure. The final weights are typically adjusted to population totals. However, although the respondents have known probabilities of selection, there is potential bias compared with NCHS' traditional household surveys. For example, web-based panel surveys may have additional nonsampling errors due to larger nonresponse and potential coverage bias compared with traditional household surveys. Because web-based panel surveys may exclude certain members of the target population, such as people without internet access, the sampling frame may not be representative of the target population. Additionally, because recruited panels must be maintained and refreshed, RANDS has smaller sample sizes compared with NCHS' household surveys, which can impact the precision of estimates.

While organizations that maintain commercial panels, such as Gallup or NORC, will often provide final survey weights that have been adjusted to selected population totals, additional calibration of the panel survey weights may be of interest to adjust for additional bias in the panel survey estimates.

Calibration weighting further adjusts the panel weights to align with auxiliary information (2–7), such as the weights of a reference survey on specified covariates. In the processing and evaluation of complex survey data, calibration weighting is a standard procedure because national household surveys typically use calibration weighting methods like poststratification or raking to adjust for nonresponse and to reflect the target U.S. population through selected reference covariates. In practice, applying calibration to survey panel weights can be challenging because many potential approaches and different considerations exist, including selecting the reference survey, selecting the calibration variables, and implementing checks to evaluate the performance of the weighting approach.

For RANDS surveys, an additional adjustment step that incorporates information from NCHS surveys, such as the National Health Interview Survey (NHIS), is used to reduce bias in the estimates. The NHIS-calibrated weights incorporate additional covariates such as health variables, which often results in reduced bias in RANDS estimates when compared with estimates from NHIS. Early rounds of RANDS were used for researching calibration weighting methods, and this approach was implemented officially beginning with RANDS during COVID-19.

This report describes the RANDS survey design and approaches for calibrating RANDS panel weights. Although the specific calibration weighting procedures may vary between rounds, the Appendix presents an example of applying calibration weighting to the RANDS weights and assessing the impact on the sample weights and resulting estimates. The Appendix describes the calibration weighting procedures for the three rounds of RANDS during COVID-19; additional information on these rounds is available from the NCHS website (https://www.cdc.gov/nchs/covid19/rands.htm) and in the technical documentation (8–10).

RANDS Survey Design

Sample Design

The survey design and development of RANDS panel weights varies by round, although generally the target population consists of the U.S. civilian noninstitutionalized adult population aged 18 and over. The published technical documentation files provide more details on the procedures for specific rounds of RANDS (https://www.cdc.gov/nchs/rands/data.htm). The general procedure used for rounds conducted by NORC is described in the following sections.

NORC's AmeriSpeak Panel is formed by sampling strata and primary sampling units (PSUs) at geographic levels. From this probability-based panel, the RANDS sample is selected using a stratified sampling design where the demographic variables, including features such as age group, sex, race and Hispanic origin, and education level, are used to form strata for study-specific samples. Within strata, PSUs are selected using simple random sampling. The sampling design for each round of RANDS can be found in the respective technical documentation (https://www.cdc.gov/nchs/rands/data.htm).

Sample Weights

Before any adjustment by NCHS, the RANDS final panel sampling weights are obtained through several steps and account for the sample design of the panel (each panel member has a panel base sampling weight) as well as selection of a sampled panel member into RANDS (each member selected has a study-specific base sampling weight). The panel base sampling weights are computed as the inverse probability of selection from the national frame and account for inclusion in the commercial panel (for example, the Gallup Panel or the AmeriSpeak Panel). The panel base sampling weights are adjusted typically for nonresponse and undercoverage at the household-level or the person-level and may include nonresponse follow-up.

The final panel sampling weights are adjusted to external population totals, such as from the decennial census or the American Community Survey. The study-specific base sampling weight uses the final panel weight and is multiplied by the inverse probability of selection of panel member into the RANDS study. The probability of selection of a panelist within a stratum is the ratio of the number of panelists sampled to the total number of panelists available in that stratum. To decrease potential nonresponse bias, there is an adjustment for survey nonrespondents, and the final study-specific weights are adjusted to general population totals on selected sociodemographic characteristics. In the final stage of weighting, extreme weights may be trimmed (criterion of minimizing mean squared error) and re-raked to population totals. Weights may be proportionally adjusted to sum to the total number of survey respondents.

RANDS Sampling Weights Calibration

Overview of Calibration Methods

Calibration weighting generally refers to the method of aligning a survey of interest, often referred to as a target survey, to a reference survey by adjusting the target survey weights (6,7). Traditional survey calibration weighting approaches include raking and poststratification and are not limited to applications with alternative data sources such as web-based panel surveys. For example, using these approaches, survey weights for population health surveys like NHIS are typically adjusted to external population totals such as counts from the U.S. Census Bureau, including the decennial census and census surveys such as the American Community Survey. These methods also can be applied to panel surveys, although the primary difference is the selection of the reference survey, which may be a population health survey rather than census data (see “Reference Survey Selection”). For a general reference on survey design and weighting approaches, see Särndal, Swensson, and Wretman (11).

More recently, alternative survey weight adjustment techniques such as weighting or matching based on propensity scores (12) have been proposed to generate pseudoweights. These methods have been considered primarily for developing weights for opt-in or nonprobability surveys (13), although the methodology can also be applied to probability samples where the base weights are the weights developed based on the sampling approach.

Although there are several methods for adjusting weights, two approaches—raking and inverse propensity score weighting (IPSW)—are described below. These methods have been investigated by NCHS for calibrating the RANDS weights and are discussed in the context of this application.

Raking, also known as iterative proportional fitting, is a method of adjusting the sample weights to reflect the true population distribution (14). The method iterates to adjust the marginal totals for a set of specified variables (for example, sex and education) from the target survey (for example, RANDS) to external population counts for a specified set of variables. This process is repeated until convergence is achieved, which is typically defined as the marginal total of the raked weights being within a specified tolerance of the external population total. In practice, a maximum number of iterations is specified as a stopping point in case convergence cannot be achieved for the specified variables and specified tolerance. Raking can incorporate interactions between the specified variables but may not achieve convergence within the specified number of iterations if there are too many terms. Note that poststratification, an alternative calibration weighting approach, is a special case of raking with one set of marginals. Missing values can be included or excluded in the raking procedure. Raking assumes that the provided distribution of the counts for the reference sample is the true population distribution, so the counts of missing values from the reference survey are used to adjust the counts of missing values in the target survey. Alternatively, the missing values can be ignored or the weights for the nonmissing values can be adjusted before raking to account for the proportion of missing values. Because raking is a standard procedure for adjusting survey weights, it is available in many survey software packages such as SUDAAN and SAS, using the %RAKING macro (15).

An alternative weighting adjustment method compared with traditional calibration techniques is IPSW, a propensity score-based method (16–18). Propensity score-based methods were developed as a post-hoc approach to reduce confounding in observational studies (12). The approach adjusts for confounding effects by balancing the covariates between the comparison groups. In the case of survey calibration, the propensity score is the estimated probability of inclusion in the target survey (for example, RANDS). IPSW, along with other propensity score-based methods, requires individual-level records to estimate the propensity scores. This differs from raking, which only requires the marginal totals from the reference data set. In IPSW, the target data and reference data are combined, and the probabilities are estimated using propensity score modeling. This approach assumes that the probability is estimable from the combined sample and that, given a set of covariate values, there is a nonzero probability of being in the target survey. Logistic regression is commonly used to estimate propensity scores as a function of specified covariates, although other models can also be used to estimate the expected probability. The usual assumptions for logistic regression imply the logit must be linear through the model variables. If the variables included in the logistic-regression model do not result in a linear relationship, then the model would have a lack of fit (19).

However, propensity score-based methods have the benefit of being highly flexible, as interactions and higher-order terms can easily be included. These models can also be overspecified, particularly when using categorical variables that include main effect terms for the number of categories minus one, although modern software programs can handle overspecified models. Records with missing values will be excluded from typical logistic regression. If it is reasonable to balance the percentage of missing values in both data sets, a unique value for categorical variables could be used to signify missing data, and records with missing values would not be excluded from the model fit.

At NCHS, raking and IPSW have been evaluated through RANDS and compared for a set of specified calibration variables. In general, the weights produced from both approaches are similar (results not shown) and both have been used for various research purposes. The three rounds of RANDS during COVID-19 were the only rounds with national and subnational estimates, termed experimental estimates, that were publicly released. For RANDS during COVID-19, raking was used for producing the calibrated weights. Raking was used because it can be directly implemented using several computer software packages and has the benefit that the raked weights produce estimates for the calibration variables that align with the marginal distributions from the reference data set.

Reference Survey Selection

Although the selection of the reference survey is flexible for calibrating panel surveys, it is assumed that the selected reference data reflect a reasonable benchmark for the true distribution of the selected calibration variables.

For the completed rounds of RANDS, NHIS has been used as the reference data set for calibration. NHIS is a national household survey used to measure a variety of health outcomes. The sample adult file, which includes information on U.S. adults aged 18 and over, covers the same target population. More information on NHIS can be found on its website: https://www.cdc.gov/nchs/nhis/index.htm.

NHIS is a high-quality population health survey and has been used as the benchmark for reporting on many health outcomes in the United States. For example, Healthy People 2030 uses NHIS to measure selected health objectives and assess changes (https://health.gov/healthypeople). Because RANDS is a health survey, NHIS has been a useful reference data source to adjust for survey errors in the RANDS weights associated with measuring health outcomes and to reduce the bias of mean estimates. Additionally, many questions included on the RANDS questionnaires are from NHIS, so calibrating to NHIS has been beneficial for studying and evaluating calibrated estimates. However, the reference data set could vary in future rounds of RANDS based on the target population or changing questionnaire topics.

In practice, NCHS uses the most recent data file available from the reference survey for comparison. For example, for the release of RANDS during COVID-19 Round 1 and Round 2 experimental estimates, the 2018 NHIS was the most recent year of available data and was used as the reference data set. However, for the RANDS during COVID-19 Round 3 experimental estimates, the 2019 NHIS was available and was used for calibrating Round 3 panel weights. Yearly data have been used for calibration, although Irimata et al. (18) evaluated the time frame and reference survey size for calibration and found that RANDS estimates were robust to using quarterly or yearly NHIS data.

Variable Selection

After identifying a reference data set, the calibration variables must be selected. Generally for propensity score-based methods, it is recommended to include as many variables as possible that may be associated with the treatment assignment (for example, participation in one of the two surveys) or the outcome for calibration (20). Although provided panel weights may have been adjusted previously on demographic variables as NORC does with the RANDS panel weights, including demographic variables may be important for several reasons. First, if the additional calibration weighting step does not include demographic variables, the previous alignment of the demographic variables to population totals may not be maintained. Second, the reference year for national benchmarks may vary, as well as the specified categorizations or groupings for the demographic variables. Third, including additional demographic variables, particularly for web-based panel surveys, may account for additional differences in the population of a web-based panel compared with the target population.

At NCHS, studies have found that incorporating health variables into the calibration has improved estimates of selected health outcomes by reducing the bias relative to NHIS (17). Additionally, variables that account for mode differences or other features of the survey may be considered. For example, Round 3 of RANDS during COVID-19 featured a telephone oversample, and the calibration incorporated two additional variables (metropolitan status and phone service) to account for potential differences when compared with the previous two rounds of RANDS during COVID-19.

Covariate selection research in the context of propensity, score-based, and calibration weighting approaches for survey weight adjustment has shown that variables related to the data source and the outcome of interest are the most important for reducing bias (21). Two methods to screen potential calibration variables include: 1) evaluating correlation and 2) evaluating the balance of the covariate distribution. Assessing the correlation between a covariate and the outcome can be used to identify covariates that may significantly improve estimation of a particular outcome. In addition, covariates that have different distributions or poor covariate balance in the two data sources are good candidates for calibration.

As an example of selected covariates for calibration in the RANDS data, the following covariates were included in raking RANDS during COVID-19 Round 1 and Round 2 weights: age group (18–34, 35–49, 50–64, 65 and over), sex (male, female), race and Hispanic origin (Hispanic, non-Hispanic White, non-Hispanic Black, non-Hispanic Other), education (high school diploma or less, some college, bachelor's degree or higher), household income ($0–$49,999, $50,000–$99,999, $100,000 or higher), census region (Northeast, Midwest, South, West), marital status (married, widowed, divorced, separated, never married, living with partner), ever diagnosed with high cholesterol (yes, no), ever diagnosed with asthma (yes, no), ever diagnosed with hypertension (yes, no), and ever diagnosed with diabetes (yes, no). Note that for race and Hispanic origin, the category non-Hispanic Other includes non-Hispanic Asian, non-Hispanic American Indian or Alaska Native, non-Hispanic Native Hawaiian or Other Pacific Islander, and non-Hispanic people of two or more races. Race and Hispanic-origin groups were selected based on categories with reliable data, as reported in the RANDS during COVID-19 public-use files.

Calibration Weighting Procedures

After selecting the reference survey, calibration variables, and calibration method, the calibration approach can be applied to obtain the calibrated weights. In the case of raking, the calibrated weights are the final weights produced through iteration that meet the convergence criteria. For example, the %RAKING macro developed for SAS (15) uses starting weights and provides marginal totals from a reference data set, and it iterates until the marginal totals of the raked weights are within a specified tolerance of the provided controls (default tolerance of 1 used) or until it reaches the maximum number of iterations (default maximum of 50 iterations used). The final raked weights are used as the calibrated weights for the target survey (for example, RANDS).

For IPSW, the fitted model is used to obtain predicted propensity scores (p). The calibrated weights (w_c) are calculated using the original weights (w) and the inverse odds for the predicted propensity score as:

$$w_c = w \bullet \frac{(1-p)}{p}$$

For the RANDS data, the final calibrated weights (using either approach) are adjusted proportionally to sum to the total number of RANDS respondents for each round. After the final calibrated weights have been produced and checked (see "Validating Calibrated Weights"), the adjusted weights can be used to calculate selected mean estimates using the Horvitz–Thompson estimator (22). The weights must be used to produce mean estimates with meaningful population representativeness, and strata, clusters, and weights are needed for variance estimation.

Validating Calibrated Weights

This section provides some practical guidelines for evaluating the calibration approach and the calibrated weights before estimation.

For each calibration method, certain checks can be performed to evaluate the performance of the approach. For raking, convergence can be assessed to determine if the final target weights achieved the specified tolerance relative to the reference data set. The number of iterations can also be used to evaluate convergence (for example, if the number of iterations is less than a user-specified maximum number of iterations). For propensity score weighting, checks can be performed to evaluate the estimation of the propensity scores. If logistic regression is used to estimate the propensity scores, then the fit of the logistic-regression model can be evaluated including an assessment of any outliers or extreme values. In addition, for models that account for the survey design (including strata and PSUs), the fit of the model and the appropriateness of the selected calibration variables can be evaluated by examining the significance of the covariates in the propensity score model. The distribution of the estimated propensity scores or the distribution of the inverse odds of the estimated propensity scores may also be examined to identify extreme values. Note that the uncalibrated and calibrated panel weights are correlated since both sets of weights share the same base weights. For statistical comparisons between the two sets of weights, comparisons need to account for the correlation or need to be evaluated using an alternative approach such as only evaluating the adjustment factor; for example, testing the inverse odds of estimated propensity scores for IPSW.

After the final calibrated weights are obtained, standard checks for survey weights can be used to evaluate the weights. Reporting descriptive statistics for the weights, including the minimum, mean, maximum, and quantiles (first quartile, median, third quartile) is useful for reviewing the distribution of the weights. The coefficient of variation, defined as the standard deviation of the weights divided by the mean of the weights multiplied by 100 to convert to a percentage, can be used to evaluate the variability of the weights. Weights identified as outliers may be trimmed or may not require further adjustment if the percent of outliers is within an acceptable bound. Various thresholds have been proposed for identifying outlier survey weights, including percentiles, compound weight pooling (23), and the median weight plus a multiple of the interquartile range of the weights (typically, 4, 5, or 6; 24,25). For the RANDS data, a cutoff of three standard deviations within the median has been used for identifying outliers. Calibration of the RANDS weights has resulted in less than 3% of weights being identified as outliers using this bound.

After the calibrated weights are evaluated and determined to be appropriate for use, the outcome estimates calculated using the calibrated weights can be compared with the outcome estimates calculated using the original panel weights (precalibration) to determine the impact of the calibration. In the case where the estimate is available in a reference data set for benchmarking purposes, measures such as the relative bias or standardized bias can be used to compare the estimate derived using calibrated weights compared with the estimate from the reference data set. For example, for the estimated prevalence of a condition in RANDS ($\hat{p}_{RANDS}$) and the estimated prevalence of the same condition in NHIS ($\hat{p}_{NHIS}$), the relative bias and standardized bias are calculated as:

$$\text{Relative Bias} = \frac{\hat{p}_{RANDS} - \hat{p}_{NHIS}}{\hat{p}_{NHIS}}$$

$$\text{Standardized Bias} = \frac{\left|\hat{p}_{RANDS} - \hat{p}_{NHIS}\right|}{\sqrt{\hat{p}_{NHIS} \bullet (1-\hat{p}_{NHIS})}}$$

Note that the standardized bias is a form of Cohen's w, the effect size for categorical data. Standardized bias cutoffs of 0.10, 0.30, and 0.50 can be used to identify estimates that have low, medium, and high bias, respectively, compared with the reference estimate (26). This measure can also be used for assessing the impact of the calibration by comparing the bias of the panel estimates using the original, uncalibrated panel weights and the calibrated weights.

The standard error of the estimate can also be used for evaluating the impact of the calibration. If available, the target survey's first stage strata and PSUs are used, along with the assumption that the calibrated weights can be treated as sampling weights, for variance estimation. If the target survey has no survey design information, then the target design is treated as sampling with replacement with size proportional to the unit's calibration weight.

Discussion

The availability of probability-sampled, commercial survey panels has expanded opportunities for alternative modes of data collection in the federal government. For example, through research conducted using RANDS and the public release of COVID-19-related experimental estimates using RANDS during COVID-19, NCHS has demonstrated the value of using these data sources to collect timely information on emerging topics to supplement data collected from traditional household surveys. However, in light of known limitations of web-based panel data, NCHS has investigated calibration weighting to adjust the panel weights and reduce bias in health outcome estimates from RANDS. This report describes the RANDS calibration weighting procedures used at NCHS and general guidelines for calibrating weights for web-based, commercial panel surveys based on previous research and experience using the RANDS platform. A detailed example of the RANDS calibration weighting procedure applied to the three rounds of RANDS during COVID-19 is provided in the Appendix.

Calibration involves selecting the calibration procedure, reference data set, and calibration covariates. For the RANDS program, raking has commonly been used to calibrate the panel weights to the corresponding year of data from NHIS on selected demographic and health variables. The RANDS research program has demonstrated that web-based panel surveys can be used to improve timeliness of data collection and that calibration generally improves estimates relative to benchmark data sets such as NHIS. Calibration for new rounds of RANDS may vary depending on the most recent reference data set available or on the survey features and variables available for calibration.

References

1. Irimata KE, Scanlon PJ. The Research and Development Survey (RANDS) during COVID-19. Stat J IAOS 38(1):13–21. 2022.
2. Deville J-C, Särndal C-E. Calibration estimators in survey sampling. J Am Stat Assoc 87(418):376–82. 1992.
3. Särndal C-E. The calibration approach in survey theory and practice. Surv Method 33(2):99–119. 2007.
4. Kim JK, Park M. Calibration estimation in survey sampling. Int Stat Rev 78(1):21–39. 2010. DOI: https://dx.doi.org/10.1111/j.1751-5823.2010.00099.x
5. Wu C, Lu WW. Calibration weighting methods for complex surveys. Int Stat Rev 84(1):79–98. 2016. DOI: https://dx.doi.org/10.1111/insr.12097.
6. Kott PS. Calibration weighting in survey sampling. Wiley Interdiscip Rev Comput Stat 8(1):39–53. 2016.
7. Wu C, Thompson ME. Calibration weighting and estimation. In: Chen J, Chen D-G, editors. Sampling theory and practice. Switzerland: Springer Cham. 115–37. 2020.
8. National Center for Health Statistics. RANDS during COVID-19 Round 1 probability sample technical documentation. 2022. Available from: https://www.cdc.gov/nchs/rands/files/RANDS_COVID_1_technical_documentation.pdf.
9. National Center for Health Statistics. RANDS during COVID-19 Round 2 probability sample technical documentation. 2022. Available from: https://www.cdc.gov/nchs/rands/files/RANDS_COVID_2_technical_documentation.pdf.
10. National Center for Health Statistics. RANDS during COVID-19 Round 3 technical documentation. 2022. Available from: https://www.cdc.gov/nchs/rands/files/RANDS_COVID_3_technical_documentation.pdf.
11. Särndal C-E, Swensson B, Wretman J. Model assisted survey sampling. New York: Springer. 1992.
12. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 70(1):41–55. 1983. DOI: https://dx.doi.org/10.1093/biomet/70.1.41.
13. Elliott MR, Valliant R. Inference for nonprobability samples. Stat Sci 32(2):249–64. 2017.
14. Deming WE, Stephan FF. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Ann Math Stat 11(4):427–44. 1940.
15. Izrael D, Hoaglin DC, Battaglia MP. A SAS macro for balancing a weighted sample. 2000. Available from: https://support.sas.com/resources/papers/proceedings/proceedings/sugi25/25/st/25p258.pdf.
16. Valliant R, Dever JA. Estimating propensity adjustments for volunteer web surveys. Sociol Methods Res 40(1):105–37. 2011.
17. Parker J, Miller K, He Y, Scanlon P, Cai B, Shin H-C, et al. Overview and initial results of the National Center for Health Statistics' Research and Development Survey. Stat J IAOS 36(4):1199–1211. 2020. DOI: https://dx.doi.org/10.3233/SJI-200678.
18. Irimata KE, He Y, Cai B, Shin H-C, Parsons VL, Parker JD. Comparison of quarterly and yearly calibration data for propensity score adjusted web survey estimates. In: Survey methods: Insights from the field, special issue: Advancements in online and mobile survey methods. 2020. Available from: https://surveyinsights.org/?p=13426.

19. Toth D, Phipps P. Regression tree models for analyzing survey response. In: Proceedings of the Joint Statistical Meetings, Government Statistics Section. Alexandria, VA: American Statistical Association. 2014.

20. Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci 25(1):1–21. 2010. DOI: https://dx.doi.org/10.1214/09-STS313.

21. Li Y, Irimata KE, He Y, Parker J. Variable inclusion strategies through directed acyclic graphs to adjust health surveys subject to selection bias for producing national estimates. J Off Stat 38(3):875–900. 2022.

22. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 47(260):663–85. 1952.

23. Pedlow S, Wang Y, Scheib E, Shin H-C. More outlier weight issues in REACH 2010. In: Proceedings of the Joint Statistical Meetings, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 2005. Available from: http://www.asasrms.org/Proceedings/y2005/files/JSM2005-000926.pdf.

24. Chowdhury S, Khare M, Wolter K. Weight trimming in the National Immunization Survey. In: Proceedings of the Joint Statistical Meetings, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 2007. Available from: http://www.asasrms.org/Proceedings/y2007/Files/JSM2007-000077.pdf.

25. Potter F, Zheng Y. Methods and issues in trimming extreme weights in sample surveys. In: Proceedings of the Joint Statistical Meetings, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 2015. Available from: http://www.asasrms.org/Proceedings/y2015/files/234115.pdf.

26. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates. 1988.

Appendix. Example: Research and Development Survey During COVID-19 Sampling Weights Calibration

To provide an example of applying calibration weighting and evaluating the checks described in this report, the three rounds of the Research and Development Survey (RANDS) during COVID-19 are considered below. This Appendix demonstrates the calibration weighting procedures that were used for RANDS during COVID-19 Rounds 1–3 to produce the publicly released experimental estimates on COVID-19-related outcomes (available from: https://www.cdc.gov/nchs/covid19/rands.htm). More information on the sample weighting and design of RANDS during COVID-19 can be found on the RANDS web page: https://www.cdc.gov/nchs/rands/data.htm.

All three rounds of RANDS during COVID-19 were conducted by NORC at the University of Chicago (NORC). RANDS during COVID-19 Rounds 1 and 2, collected in summer 2020 (June 9, 2020, to July 6, 2020, and August 3, 2020, to August 20, 2020, respectively), were calibrated to the 2018 National Health Interview Survey (NHIS), as these were the most recent NHIS data publicly available at the time. Round 3, collected in spring 2021 (May 17, 2021, to June 30, 2021), was calibrated to the 2019 NHIS because these were the most recent data available to the public (released September 2020) when RANDS during COVID-19 Round 3 calibration weighting was performed.

RANDS during COVID-19 Rounds 1 and 2 included 11 calibration variables, including 7 sociodemographic variables and 4 health variables. These variables included: age group, sex, race and Hispanic origin, education, household income, census region, marital status, ever diagnosed with high cholesterol, ever diagnosed with asthma, ever diagnosed with hypertension, and ever diagnosed with diabetes. RANDS during COVID-19 Round 3 used 13 calibration variables, including the 11 calibration variables from the previous two rounds in addition to metropolitan status and phone service. For each round, the distributions of the selected calibration variables in RANDS during COVID-19 and the corresponding reference year of NHIS were compared (Tables I–III).

Tables I–III show that the distributions of the selected calibration variables in RANDS (see column "NORC weighted percent" in Table I, for example) appear to differ from the distributions of the variables in NHIS (see column "Weighted percent" in Table I, for example) before calibrating the weights.

Calibration weighting for all three rounds was performed using raking in SAS, through the %RAKING SAS macro (15). The NORC-provided weights were used as the starting weights and were adjusted iteratively to align the distributions of the calibration variables to the distributions of the variables in the respective year of NHIS (2018 for Rounds 1 and 2, 2019 for Round 3). For covariates that were missing values in only NHIS, the raking procedure was adjusted to ignore the missing values in NHIS and align the distribution of the RANDS covariates to the distribution of the remaining nonmissing values in NHIS. The raking procedure converged in six iterations for the calibration of each round of RANDS during COVID-19 data. A sample call to the %RAKING SAS macro for RANDS during COVID-19 Round 1 is shown below:

```
%RAKING(inds=rc1_recode,
        outds=rc1_raked,
        inwt=weight,
        freqlist=,
        outwt=WEIGHT_CALIBRATED,
        byvar=,
        varlist=AGEGRP GENDER RACETH EDUC3
        INCGRP REGION4 MARITAL CHLEV ASEV
        HYPEV DIBEV,
        numvar=11,
        cntotal=6800,
        trmprec=1,
        numiter=50;
```

where:

- *inds* is the name of the input data set
- *outds* is the name of the data set output from the raking procedure
- *inwt* is the name of the weight variable (available in the input data set) to be raked
- *freqlist* is the list of data sets with marginal frequencies or marginal control totals (must contain name of raking

variable and either column PERCENT or MRGTOTAL)

- *outwt* is the name of the resulting raked weight
- *byvar* is the variable over which a user can set up raking (by default, raking is done over the whole input data set)
- *varlist* is the list of variables to be raked on
- *numvar* is the number of raking variables
- *cntotal* is the general control total
- *trmprec* is the tolerance (default is 1)
- *numiter* is the number of iterations (default is 50)

Note that the default for *freqlist* is a list of data sets with the names of the raking variables.

In this example, *rc1_recode* is an input data set that has already been adjusted for missing values, and *weight* is the variable in this data set that contains the RANDS weights. There are 11 calibration variables (*AGEGRP*, *GENDER*, *RACETH*, *EDUC3*, *INCGRP*, *REGION4*, *MARITAL*, *CHLEV*, *ASEV*, *HYPEV*, *DIBEV*) that are specified in the raking and data sets, named according to the raking variables containing the column PERCENT with the control totals from NHIS having already been created before this %RAKING macro call. The output data set *rc1_raked* contains the raked weights in the column *WEIGHT_CALIBRATED*.

After raking, the distributions of the calibration variables in each round of RANDS during COVID-19 closely match the distributions of the variables in the corresponding round of NHIS (see columns "NHIS-calibrated weighted percent" and "Weighted, adjusted percent" in Tables I–III).

The calibrated weights were evaluated using summary statistics and checks for extreme weights before estimation. Table IV reports summary statistics for both the original NORC-provided weights and the NHIS-calibrated weights for

Table I. Unweighted sample size and weighted percent distributions of calibration variables in the Research and Development Survey during COVID-19 Round 1 and the 2018 National Health Interview Survey

Calibration variable	RANDS during COVID-19 Round 1 (number = 6,800)			2018 NHIS (number = 25,417)		
	Sample size	NORC weighted percent	NHIS-calibrated weighted percent	Sample size	Weighted percent	Weighted, adjusted percent[1]
Age group (years)						
18–34	1,470	30.24	29.64	5,762	29.64	29.64
35–49	1,624	23.40	24.53	5,766	24.53	24.53
50–64	1,900	24.87	25.21	6,592	25.21	25.21
65 and over	1,806	21.49	20.61	7,297	20.61	20.61
Missing	–	–	–	–	–	–
Sex						
Male	2,969	48.30	48.28	11,550	48.28	48.28
Female	3,831	51.70	51.72	13,867	51.72	51.72
Missing	–	–	–	–	–	–
Race and Hispanic origin						
Hispanic	943	16.66	16.33	3,179	16.34	16.34
Non-Hispanic White	4,515	62.79	63.8	17,569	63.80	63.80
Non-Hispanic Black	813	11.93	12.31	2,978	12.31	12.31
Non-Hispanic Other[2]	529	8.62	7.55	1,691	7.55	7.55
Missing	–	–	–	–	–	–
Education						
High school diploma or less	1,296	38.01	35.95	8,996	35.75	35.95
Some college	2,557	27.73	30.55	7,658	30.39	30.55
Bachelor's degree or higher	2,947	34.26	33.50	8,657	33.32	33.50
Missing	–	–	–	106	0.54	…
Household income (dollars)						
0–49,999	2,601	42.99	37.66	10,734	34.78	37.66
50,000–99,999	2,344	32.74	30.55	7,096	28.21	30.55
100,000 or more	1,855	24.27	31.79	5,754	29.36	31.79
Missing	–	–	–	1,833	7.66	…
Region						
Northeast	1,023	17.44	17.34	4,143	17.34	17.34
Midwest	1,837	20.73	21.98	5,949	21.98	21.98
South	2,325	38.00	36.9	9,312	36.9	36.9
West	1,615	23.83	23.78	6,013	23.78	23.78
Missing	–	–	–	–	–	–

See footnotes at end of table.

Table I. Unweighted sample size and weighted percent distributions of calibration variables in the Research and Development Survey during COVID-19 Round 1 and the 2018 National Health Interview Survey—Con.

Calibration variable	RANDS during COVID-19 Round 1 (number = 6,800)			2018 NHIS (number = 25,417)		
	Sample size	NORC weighted percent	NHIS-calibrated weighted percent	Sample size	Weighted percent	Weighted, adjusted percent[1]
Marital status						
Married	3,538	46.91	52.41	11,458	52.32	52.41
Widowed	347	4.77	5.73	2,532	5.72	5.73
Divorced	861	11.15	9.07	3,691	9.06	9.07
Separated	104	1.60	1.81	604	1.81	1.81
Never married	1,447	27.01	23.41	5,594	23.37	23.41
Living with partner	503	8.56	7.57	1,485	7.56	7.57
Missing	–	–	–	53	0.17	…
Ever diagnosed with high cholesterol						
Yes	2,289	31.42	27.38	7,922	27.38	27.38
No	4,477	68.20	72.30	17,410	72.30	72.30
Missing	34	0.37	0.32	85	0.32	0.32
Ever diagnosed with asthma						
Yes	1,014	14.87	13.39	3,445	13.39	13.39
No	5,731	84.38	86.51	21,942	86.51	86.51
Missing	55	0.75	0.10	30	0.10	0.10
Ever diagnosed with hypertension						
Yes	2,428	33.47	31.44	9,217	31.44	31.44
No	4,350	66.27	68.39	16,153	68.39	68.39
Missing	22	0.26	0.17	47	0.17	0.17
Ever diagnosed with diabetes[3]						
Yes	769	10.85	12.72	3,677	12.72	12.72
No	5,990	88.61	87.22	21,720	87.22	87.22
Missing	41	0.54	0.06	20	0.06	0.06

– Quantity zero.
… Category not applicable.
[1]Missing values for education, household income, and marital status in the 2018 NHIS were removed for the calibration.
[2]Non-Hispanic Other people includes non-Hispanic Asian, non-Hispanic American Indian or Alaska Native, non-Hispanic Native Hawaiian or Other Pacific Islander, and non-Hispanic people of two or more races.
[3]Excludes prediabetes and borderline diabetes.

NOTES: RANDS is Research and Development Survey. NHIS is National Health Interview Survey. NORC is NORC at the University of Chicago. NHIS-calibrated weights were calibrated to the 2018 National Health Interview Survey.

SOURCES: National Center for Health Statistics, Research and Development Survey during COVID-19 Round 1, 2020; National Health Interview Survey, 2018.

Table II. Unweighted sample size and weighted percent distributions of calibration variables in the Research and Development Survey during COVID-19 Round 2 and the 2018 National Health Interview Survey

Calibration variable	RANDS during COVID-19 Round 2 (number = 5,981)			2018 NHIS (number = 25,417)		
	Sample size	NORC weighted percent	NHIS-calibrated weights percent	Sample size	Weighted percent	Weighted, adjusted percent[1]
Age group (years)						
18–34	1,208	30.22	29.64	5,762	29.64	29.64
35–49	1,434	23.42	24.53	5,766	24.53	24.53
50–64	1,657	24.87	25.21	6,592	25.21	25.21
65 and over	1,682	21.49	20.61	7,297	20.61	20.61
Missing	–	–	–	–	–	–
Sex						
Male	2,592	48.3	48.28	11,550	48.28	48.28
Female	3,389	51.7	51.72	13,867	51.72	51.72
Missing	–	–	–	–	–	–

See footnotes at end of table.

Table II. Unweighted sample size and weighted percent distributions of calibration variables in the Research and Development Survey during COVID-19 Round 2 and the 2018 National Health Interview Survey—Con.

Calibration variable	RANDS during COVID-19 Round 2 (number = 5,981)			2018 NHIS (number = 25,417)		
	Sample size	NORC weighted percent	NHIS-calibrated weighted percent	Sample size	Weighted percent	Weighted, adjusted percent[1]
Race and Hispanic origin						
Hispanic	750	16.66	16.33	3,179	16.34	16.34
Non-Hispanic White	4,078	62.79	63.8	17,569	63.8	63.8
Non-Hispanic Black	691	11.93	12.31	2,978	12.31	12.31
Non-Hispanic Other[2]	462	8.62	7.55	1,691	7.55	7.55
Missing	–	–	–	–	–	–
Education						
High school diploma or less	1,104	38.01	35.95	8,996	35.75	35.95
Some college	2,229	27.73	30.55	7,658	30.39	30.55
Bachelor's degree or higher	2,648	34.26	33.5	8,657	33.32	33.5
Missing	–	–	–	106	0.54	…
Household income (dollars)						
0–49,999	2,279	43.31	37.66	10,734	34.78	37.66
50,000–99,999	2,060	32.33	30.55	7,096	28.21	30.55
100,000 or more	1,642	24.36	31.79	5,754	29.36	31.79
Missing	–	–	–	1,833	7.66	…
Region						
Northeast	876	17.44	17.34	4,143	17.34	17.34
Midwest	1,640	20.73	21.98	5,949	21.98	21.98
South	2,029	38	36.9	9,312	36.9	36.9
West	1,436	23.83	23.78	6,013	23.78	23.78
Missing	–	–	–	–	–	–
Marital status						
Married	3,146	46.89	52.41	11,458	52.32	52.41
Widowed	315	4.56	5.73	2,532	5.72	5.73
Divorced	777	11.38	9.07	3,691	9.06	9.07
Separated	84	1.38	1.81	604	1.81	1.81
Never married	1,233	26.94	23.41	5,594	23.37	23.41
Living with partner	426	8.85	7.57	1,485	7.56	7.57
Missing	–	–	–	53	0.17	…
Ever diagnosed with high cholesterol						
Yes	2,081	31.88	27.38	7,922	27.38	27.38
No	3,868	67.51	72.30	17,410	72.30	72.30
Missing	32	0.61	0.32	85	0.32	0.32
Ever diagnosed with asthma						
Yes	916	15.09	13.39	3,445	13.39	13.39
No	5,020	84.17	86.51	21,942	86.51	86.51
Missing	45	0.74	0.10	30	0.10	0.10
Ever diagnosed with hypertension						
Yes	2,192	33.04	31.44	9,217	31.44	31.44
No	3,767	66.56	68.39	16,153	68.39	68.39
Missing	22	0.40	0.17	47	0.17	0.17
Ever diagnosed with diabetes[3]						
Yes	691	11.02	12.72	3,677	12.72	12.72
No	5,253	88.47	87.22	21,720	87.22	87.22
Missing	37	0.51	0.06	20	0.06	0.06

– Quantity zero.
… Category not applicable.
[1]Missing values for education, household income, and marital status in the 2018 NHIS were removed for the calibration.
[2]Non-Hispanic Other people includes non-Hispanic Asian, non-Hispanic American Indian or Alaska Native, non-Hispanic Native Hawaiian or Other Pacific Islander, and non-Hispanic people of two or more races.
[3]Excludes prediabetes and borderline diabetes.

NOTES: RANDS is Research and Development Survey. NHIS is National Health Interview Survey. NORC is NORC at the University of Chicago. NHIS-calibrated weights were calibrated to the 2018 National Health Interview Survey.

SOURCES: National Center for Health Statistics, Research and Development Survey during COVID-19 Round 2, 2020; National Health Interview Survey, 2018.

Table III. Unweighted sample size and weighted percent distributions of calibration variables in the Research and Development Survey during COVID-19 Round 3 and the 2019 National Health Interview Survey

Calibration variable	RANDS during COVID-19 Round 3 (number = 5,458)			2019 NHIS (number = 31,997)		
	Sample size	NORC weighted percent	NHIS-calibrated weighted percent	Sample size	Weighted percent	Weighted, adjusted percent[1]
Age group (years)						
18–34	1,249	31.08	29.67	7,058	29.60	29.67
35–49	1,062	22.20	24.34	7,250	24.29	24.34
50–64	1,402	24.70	24.93	8,313	24.88	24.93
65 and over	1,745	22.01	21.06	9,295	21.00	21.05
Missing	–	–	–	81	0.23	…
Sex						
Male	2,464	48.24	48.29	14,733	48.29	48.29
Female	2,994	51.76	51.71	17,261	51.71	51.71
Missing	–	–	–	3	0.01	…
Race and Hispanic origin						
Hispanic	821	16.89	16.54	4,152	16.54	16.54
Non-Hispanic White	3,481	62.41	63.23	21,915	63.23	63.23
Non-Hispanic Black	763	11.98	11.75	3,483	11.75	11.75
Non-Hispanic Other[2]	393	8.72	8.48	2,447	8.48	8.48
Missing	–	–	–	–	–	–
Education						
High school diploma or less	1,282	37.93	39.90	11,155	39.60	39.90
Some college	2,582	26.95	31.10	9,386	30.87	31.10
Bachelor's degree or higher	1,594	35.12	29.00	11,277	28.78	29.00
Missing	–	–	–	179	0.75	…
Household income (dollars)						
0–49,999	2,184	41.47	37.66	12,352	34.38	37.66
50,000–99,999	1,863	32.25	32.00	8,976	29.21	32.00
100,000 or more	1,411	26.28	30.34	8,047	27.69	30.34
Missing	–	–	–	2,622	8.72	…
Region						
Northeast	800	17.27	17.76	5,410	17.76	17.76
Midwest	1,475	20.66	21.03	7,104	21.03	21.03
South	1,833	38.21	37.67	11,676	37.68	37.68
West	1,350	23.86	23.53	7,807	23.53	23.53
Missing	–	–	–	–	–	–
Marital status						
Married	2,754	46.70	52.36	14,759	50.84	52.36
Widowed	384	5.48	5.98	3,115	5.80	5.98
Divorced	675	10.75	9.02	4,317	8.76	9.02
Separated	99	2.18	1.17	456	1.13	1.17
Never married	1,185	28.02	22.54	6,368	21.88	22.54
Living with partner	361	6.88	8.94	2,136	8.68	8.94
Missing	–	–	–	846	2.91	…
Ever diagnosed with high cholesterol						
Yes	2,010	33.09	24.76	9,179	24.76	24.76
No	3,427	66.49	74.87	22,697	74.87	74.87
Missing	21	0.42	0.37	121	0.37	0.37
Ever diagnosed with asthma						
Yes	893	16.80	13.44	4,229	13.44	13.44
No	4,523	82.38	86.42	27,718	86.42	86.42
Missing	42	0.82	0.14	50	0.14	0.14
Ever diagnosed with hypertension						
Yes	2,106	33.82	31.59	11,480	31.59	31.59
No	3,330	65.76	68.22	20,458	68.22	68.22
Missing	22	0.41	0.18	59	0.18	0.18

See footnotes at end of table.

Table III. Unweighted sample size and weighted percent distributions of calibration variables in the Research and Development Survey during COVID-19 Round 3 and the 2019 National Health Interview Survey—Con.

Calibration variable	RANDS during COVID-19 Round 3 (number = 5,458)			2019 NHIS (number = 31,997)		
	Sample size	NORC weighted percent	NHIS-calibrated weighted percent	Sample size	Weighted percent	Weighted, adjusted percent[1]
Ever diagnosed with diabetes[3]						
Yes	742	12.23	9.33	3,355	9.33	9.33
No	4,684	87.25	90.52	28,594	90.52	90.52
Missing	32	0.52	0.14	48	0.14	0.14
Metropolitan status[4]						
Metropolitan	4,624	83.69	85.66	26,916	85.66	85.66
Nonmetropolitan	834	16.31	14.34	5,081	14.34	14.34
Phone service						
Landline telephone only	405	5.50	2.35	1,127	2.26	2.35
Landline and cellular phone	2,146	38.24	36.10	11,367	34.85	36.1
Cellular phone only	2,876	55.60	60.60	18,210	58.50	60.6
No telephone service	31	0.65	0.95	346	0.92	0.95
Missing	–	–	–	947	3.47	…

– Quantity zero.
… Category not applicable.
[1]Missing values for age group, sex, education, household income, marital status, and phone service in the 2019 NHIS were removed for the calibration.
[2]Non-Hispanic Other includes non-Hispanic Asian, non-Hispanic American Indian or Alaska Native, non-Hispanic Native Hawaiian or Other Pacific Islander, and non-Hispanic people of two or more races.
[3]Excludes prediabetes and borderline diabetes.
[4]Metropolitan status assigned by ZIP code, where metropolitan includes metropolitan and micropolitan areas, and nonmetropolitan includes all other designations.

NOTES: RANDS is Research and Development Survey. NHIS is National Health Interview Survey. NORC is NORC at the University of Chicago. NHIS-calibrated weights were calibrated to the 2019 National Health Interview Survey.

SOURCES: National Center for Health Statistics, Research and Development Survey during COVID-19 Round 3, 2021; National Health Interview Survey, 2019.

each round. A review of the statistics does not reveal any concerning results. The minimum, first quartile, median, and third quartile of the NHIS-calibrated weights were generally consistent with the NORC-provided weights across the three rounds. The maximum weights decreased after calibration for Rounds 1 and 2, although the maximum calibrated weight increased compared with the maximum NORC-provided weight for Round 3. As expected, the standard deviation and coefficient of variation of the weights increased after calibration compared with the original NORC-provided weights. Approximately 2% of weights were identified to be out of range, although this is within an acceptable bound for RANDS and is consistent before and after calibration. Further checks reveal that there were singleton primary sampling units in Round 3 (minimum of one sample by the design variables, strata, and primary sampling units). Additionally, using the survey design structure of the three rounds, pseudovariables with specified mean proportions of 0.1 and 0.5 were randomly generated, and the design effects for the estimated means were evaluated. The range of design effects for the pseudovariables did not identify any extreme values. Because the calibrated weights meet all the checks, these weights can be used for producing weighted estimates.

In the public release of experimental estimates from RANDS during COVID-19, estimates were published on the topics of loss of work due to illness with COVID-19, telemedicine, and reduced access to care, which were not available in NHIS early in the pandemic. To compare estimates from RANDS during COVID-19 with NHIS for this example, four outcomes collected in both surveys were evaluated. For the evaluation, the selected outcomes were 1) available in both RANDS during COVID-19 and NHIS for comparison, 2) did not include calibration variables, and 3) had the same question text in the corresponding questionnaires. Note that because many questions in the RANDS during COVID-19 series were developed to test new COVID-19-related questions or were used for calibration, there is a limited set for evaluating bias.

These outcomes include 1) the prevalence of adults who currently smoke, 2) the percentage of adults who rate their health positively, 3) the prevalence of adults ever diagnosed with cancer, and 4) the prevalence of adults ever diagnosed with coronary heart disease. The prevalence of adults who currently smoke is determined using two questions on cigarette smoking, “Have you smoked at least 100 cigarettes in your entire life?” (yes or no) and “How often do you now smoke cigarettes?” (every day, some days, or not at all). Current smoking status is defined as people who have smoked at least 100 cigarettes in their lifetime and currently smoke either every day or some days. The percentage of adults with positively rated health is determined as the percentage of adults who respond “excellent,” “very good,” or “good” to the question, “Would you say your health in general is excellent, very good, good, fair, or poor?” The

Table IV. Summary statistics for NORC-provided weights and National Health Interview Survey-calibrated weights, by round of Research and Development Survey during COVID-19

Summary statistic	Round 1 NORC-provided weights	Round 1 NHIS-calibrated weights	Round 2 NORC-provided weights	Round 2 NHIS-calibrated weights	Round 3 NORC-provided weights	Round 3 NHIS-calibrated weights
Descriptive statistics						
Sample size	6,800	6,800	5,981	5,981	5,458	5,458
Minimum	0.09	0.01	0.09	0.01	0.02	0.01
First quartile	0.46	0.45	0.46	0.43	0.42	0.37
Median	0.73	0.72	0.72	0.71	0.71	0.69
Third quartile	1.18	1.21	1.18	1.21	1.25	1.24
Maximum	16.75	13.87	18.87	16.49	12.94	17.65
Mean	1.00	1.00	1.00	1.00	1.00	1.00
Standard deviation	0.98	0.99	1.01	1.03	0.98	1.07
Coefficient of variation	98.27	99.16	101.47	103.10	97.68	107.20
Percent of out-of-range[1] weights						
Low percent	0.00	0.00	0.00	0.00	0.00	0.00
High percent	2.19	2.16	2.24	2.21	2.16	2.24
Summary statistics by design variables (strata and primary sampling unit)						
Minimum number of samples	26	26	20	20	1	1
Maximum number of samples	305	305	272	272	194	194
Minimum sum of weights	11.98	11.28	11.04	9.56	0.22	0.09
Maximum sum of weights	217.75	219.49	204.03	205.94	166.69	154.87
Range of design effects for 10 pseudovariables ($p = 0.5$)						
Minimum	1.38	1.23	1.16	1.15	1.26	1.65
Maximum	2.79	2.41	2.38	2.44	2.89	3.20
Range of design effects for 10 pseudovariables ($p = 0.1$)						
Minimum	1.30	1.31	1.44	1.47	1.31	1.62
Maximum	2.67	3.03	2.66	2.36	2.61	2.97

[1]Out-of-range weights identified using a threshold of three standard deviations beyond the median.

NOTES: NORC is NORC at the University of Chicago. NHIS is National Health Interview Survey. RANDS (Research and Development Survey) during COVID-19 Rounds 1 and 2 NHIS-calibrated weights were calibrated to the 2018 National Health Interview Survey. RANDS during COVID-19 Round 3 NHIS-calibrated weights were calibrated to the 2019 National Health Interview Survey.

SOURCES: National Center for Health Statistics, Research and Development Survey during COVID-19 Rounds 1–3, 2020–2021; National Health Interview Survey, 2018–2019.

prevalence of adults who have ever been diagnosed with cancer is determined based on a response of "yes" to the question, "Have you ever been told by a doctor or other health professional that you had cancer or a malignancy?" (yes or no). The prevalence of adults who have ever been diagnosed with coronary heart disease is determined based on a response of "yes" to the question, "Have you ever been told by a doctor or other health professional that you had coronary heart disease?" (yes or no).

Table V reports the estimates and standard errors of the four selected outcomes from RANDS during COVID-19 before and after calibration, as well as from the corresponding reference year of NHIS. The relative bias and standardized bias of the estimates compared with NHIS are also reported. Overall, the RANDS during COVID-19 estimates using the NORC-provided weights tend to overestimate current smoking, ever diagnosed cancer, and ever diagnosed coronary heart disease, and underestimate the percentage of adults who rate their health as excellent, very good, or good compared with NHIS across all three rounds. After performing calibration weighting, the estimates were generally closer to the NHIS estimates compared with the original estimates using the NORC-provided weights, except for the estimate of current smoking in Round 3.

To visualize the impact of the calibration weighting on the bias, the standardized biases of the four outcomes relative to NHIS before and after calibration weighting are displayed in the Figure. Reference lines are included at standardized biases of 0.10 and 0.30 as indicators for low and midlevels of standardized bias.

The Figure shows the change in standardized bias between each of the RANDS during COVID-19 estimates using the NORC-provided weight (light blue) and the NHIS-calibrated weight (green). As seen in Table V, the standardized bias relative to NHIS is reduced for nearly all estimates when using the NHIS-calibrated weights, except for the current smoking

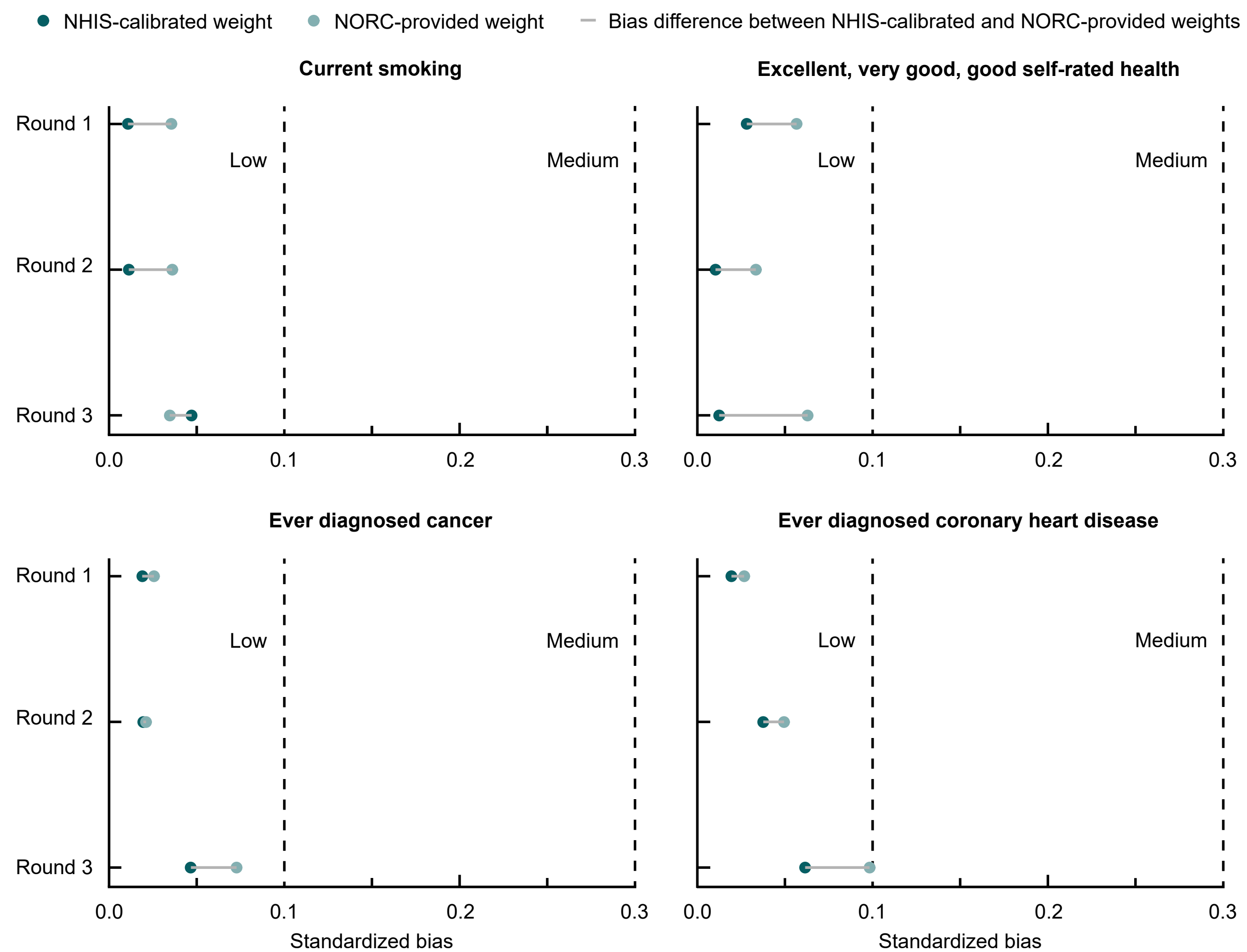
Table V. Estimates and bias of Research and Development Survey during COVID-19 and National Health Interview Survey estimates for selected health outcomes

RANDS during COVID-19 round, NHIS year, and health outcome	NHIS		RANDS during COVID-19, NORC-provided weight				RANDS during COVID-19, NHIS-calibrated weight			
	Percent	Standard error	Percent	Standard error	Relative bias	Standardized bias	Percent	Standard error	Relative bias	Standardized bias
Round 1 and 2018 NHIS										
Current smoking	13.73	0.31	14.96	0.52	0.09	0.04	14.10	0.51	0.03	0.01
Excellent, very good, good self-rated health	87.31	0.27	85.42	0.60	-0.02	0.06	86.37	0.61	-0.01	0.03
Ever diagnosed cancer	9.35	0.21	10.10	0.40	0.08	0.03	9.90	0.42	0.06	0.02
Ever diagnosed coronary heart disease	4.60	0.15	5.15	0.31	0.12	0.03	5.00	0.31	0.09	0.02
Round 2 and 2018 NHIS										
Current smoking	13.73	0.31	14.97	0.80	0.09	0.04	14.12	0.77	0.03	0.01
Excellent, very good, good self-rated health	87.31	0.27	86.20	0.67	-0.01	0.03	86.96	0.68	0.00	0.01
Ever diagnosed cancer	9.35	0.21	9.97	0.43	0.07	0.02	9.92	0.46	0.06	0.02
Ever diagnosed coronary heart disease	4.60	0.15	5.63	0.29	0.22	0.05	5.38	0.28	0.17	0.04
Round 3 and 2019 NHIS										
Current smoking	13.98	0.27	15.18	0.80	0.09	0.03	15.61	0.97	0.12	0.05
Excellent, very good, good self-rated health	84.73	0.30	82.47	0.73	-0.03	0.06	84.29	0.73	-0.01	0.01
Ever diagnosed cancer	9.49	0.21	11.62	0.62	0.22	0.07	10.85	0.61	0.14	0.05
Ever diagnosed coronary heart disease	4.60	0.13	6.66	0.37	0.45	0.10	5.89	0.35	0.28	0.06

NOTES: NHIS is National Health Interview Survey. RANDS is Research and Development Survey. NORC is NORC at the University of Chicago. RANDS during COVID-19 Rounds 1 and 2 NHIS-calibrated weights were calibrated to the 2018 National Health Interview Survey. RANDS during COVID-19 Round 3 NHIS-calibrated weights were calibrated to the 2019 National Health Interview Survey.

SOURCES: National Center for Health Statistics, Research and Development Survey during COVID-19 Rounds 1–3, 2020–2021; National Health Interview Survey, 2018–2019.

Figure. Standardized bias of estimates in the Research and Development Survey during COVID-19, by round using National Health Interview Survey-calibrated and NORC-provided weights



NOTES: RANDS is Research and Development Survey. NHIS is National Health Interview Survey. NORC is NORC at the University of Chicago. RANDS during COVID-19 Rounds 1 and 2, NHIS-calibrated weights were calibrated to the 2018 NHIS. RANDS during COVID-19 Round 3, NHIS-calibrated weights were calibrated to the 2019 NHIS.
SOURCES: National Center for Health Statistics, Research and Development Survey during COVID-19 Rounds 1–3, 2020–2021; National Health Interview Survey, 2018–2019.

estimate in Round 3. While calibration weighting generally reduced the standardized bias, all estimates (before and after calibration weighting) for this example were at or below the 0.10 cutoff for low standardized bias. The estimate for adults with diagnosed coronary heart disease in Round 3 was the closest to this cutoff using the NORC-provided weights (standardized bias of 0.10), although calibration weighting reduced the standardized bias to 0.06. The calibration weighting procedure used for these three rounds generally resulted in larger improvements in the standardized bias for current smoking and excellent, very good, and good self-rated health, and smaller improvements for ever diagnosed cancer and ever diagnosed coronary heart disease.

This example demonstrates the use of calibration weighting and the impact on health estimates in the RANDS data. As seen in the four outcomes evaluated, calibration weighting generally improves estimates from RANDS relative to the benchmark data set, although this reduction in bias may not be seen for all variables when using a general calibration weighting approach that is not specific to a certain outcome (such as for current smoking in RANDS during COVID-19 Round 3). While the four outcomes in this example had relatively low standardized biases compared with NHIS, calibration weighting is useful for reducing bias and can be particularly important for adjusting estimates of outcomes with high bias.

Vital and Health Statistics Series Descriptions

Active Series

Series 1. Programs and Collection Procedures
Reports describe the programs and data systems of the National Center for Health Statistics, and the data collection and survey methods used. Series 1 reports also include definitions, survey design, estimation, and other material necessary for understanding and analyzing the data.

Series 2. Data Evaluation and Methods Research
Reports present new statistical methodology including experimental tests of new survey methods, studies of vital and health statistics collection methods, new analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory. Reports also include comparison of U.S. methodology with those of other countries.

Series 3. Analytical and Epidemiological Studies
Reports present data analyses, epidemiological studies, and descriptive statistics based on national surveys and data systems. As of 2015, Series 3 includes reports that would have previously been published in Series 5, 10–15, and 20–23.

Discontinued Series

Series 4. Documents and Committee Reports
Reports contain findings of major committees concerned with vital and health statistics and documents. The last Series 4 report was published in 2002; these are now included in Series 2 or another appropriate series.

Series 5. International Vital and Health Statistics Reports
Reports present analytical and descriptive comparisons of U.S. vital and health statistics with those of other countries. The last Series 5 report was published in 2003; these are now included in Series 3 or another appropriate series.

Series 6. Cognition and Survey Measurement
Reports use methods of cognitive science to design, evaluate, and test survey instruments. The last Series 6 report was published in 1999; these are now included in Series 2.

Series 10. Data From the National Health Interview Survey
Reports present statistics on illness; accidental injuries; disability; use of hospital, medical, dental, and other services; and other health-related topics. As of 2015, these are included in Series 3.

Series 11. Data From the National Health Examination Survey, the National Health and Nutrition Examination Surveys, and the Hispanic Health and Nutrition Examination Survey
Reports present 1) estimates of the medically defined prevalence of specific diseases in the United States and the distribution of the population with respect to physical, physiological, and psychological characteristics and 2) analysis of relationships among the various measurements. As of 2015, these are included in Series 3.

Series 12. Data From the Institutionalized Population Surveys
The last Series 12 report was published in 1974; these reports were included in Series 13, and as of 2015 are in Series 3.

Series 13. Data From the National Health Care Survey
Reports present statistics on health resources and use of health care resources based on data collected from health care providers and provider records. As of 2015, these reports are included in Series 3.

Series 14. Data on Health Resources: Manpower and Facilities
The last Series 14 report was published in 1989; these reports were included in Series 13, and are now included in Series 3.

Series 15. Data From Special Surveys
Reports contain statistics on health and health-related topics from surveys that are not a part of the continuing data systems of the National Center for Health Statistics. The last Series 15 report was published in 2002; these reports are now included in Series 3.

Series 16. Compilations of Advance Data From Vital and Health Statistics
The last Series 16 report was published in 1996. All reports are available online; compilations are no longer needed.

Series 20. Data on Mortality
Reports include analyses by cause of death and demographic variables, and geographic and trend analyses. The last Series 20 report was published in 2007; these reports are now included in Series 3.

Series 21. Data on Natality, Marriage, and Divorce
Reports include analyses by health and demographic variables, and geographic and trend analyses. The last Series 21 report was published in 2006; these reports are now included in Series 3.

Series 22. Data From the National Mortality and Natality Surveys
The last Series 22 report was published in 1973. Reports from sample surveys of vital records were included in Series 20 or 21, and are now included in Series 3.

Series 23. Data From the National Survey of Family Growth
Reports contain statistics on factors that affect birth rates, factors affecting the formation and dissolution of families, and behavior related to the risk of HIV and other sexually transmitted diseases. The last Series 23 report was published in 2011; these reports are now included in Series 3.

Series 24. Compilations of Data on Natality, Mortality, Marriage, and Divorce
The last Series 24 report was published in 1996. All reports are available online; compilations are no longer needed.

For answers to questions about this report or for a list of reports published in these series, contact:

Information Dissemination Staff
National Center for Health Statistics
Centers for Disease Control and Prevention
3311 Toledo Road, Room 4551, MS P08
Hyattsville, MD 20782

Tel: 1–800–CDC–INFO (1–800–232–4636)
TTY: 1–888–232–6348
Internet: https://www.cdc.gov/nchs
Online request form: https://www.cdc.gov/info
For e-mail updates on NCHS publication releases, subscribe online at: https://www.cdc.gov/nchs/email-updates.htm.

U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES

Centers for Disease Control and Prevention
National Center for Health Statistics
3311 Toledo Road, Room 4551, MS P08
Hyattsville, MD 20782–2064

For more NCHS Series Reports, visit:
https://www.cdc.gov/nchs/products/series.htm

Series 2, No. 199
CS336699