# Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA)
## *2023 National Health Interview Survey*

### Overview

In the *Journal of Survey Statistics and Methodology*, Seidenberg, Moser, and West (2023) proposed an itemized checklist to guide researchers publishing analyses using complex sample survey data.  This checklist—the Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA)—is intended to help eliminate analytic and reporting errors and increase transparency and reproducibility.  NCHS is providing the following information to support researchers publishing analyses using 2023 National Health Interview Survey (NHIS) data.

Additional information about data collection methods, year-specific content, and other useful resources for the 2023 NHIS is available in the 2023 *Survey Description* document, available at:
https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2023/srvydesc-508.pdf

### PRICSSA-related Content, in Brief

- Name and wave of survey: 2023 National Health Interview Survey
- Data collection mode: Computer-assisted personal interviewing (CAPI), usually conducted in respondent's homes, but follow-ups to complete interviews may be conducted over the telephone
- Dates of data collection: January 1, 2023-December 31, 2023
- Target population: Civilian noninstitutionalized population residing in the 50 US states and Washington, DC
- Populations excluded: Active-duty military, civilians living on military bases, persons in long-term care and correctional facilities, persons with no fixed household address
- Design: Stratified cluster sample
- Variance estimation: Taylor Series Linearization
- Weight and design variables
  - Weight: WTFA_A for sample adults and WTFA_C for sample children
  - PSU: PPSU
  - Stratum: PSTRAT
- Unweighted sample size: 29,552 sample adults and 7,692 sample children in 30,670 households
- Response rate: 47.0% for sample adults and 44.9% for sample children, based on the American Association of Public Opinion Research (AAPOR) Response Rate Definition # 2, or AAPOR RR2 (AAPOR, 2023)

## PRICSSA-related Content, in Detail

| 2023 National Health Interview Survey (NHIS) | |
|---|---|
| **1.1**<br>**Data collection dates** | The 2023 NHIS interviews were conducted between January 1, 2023–December 31, 2023. |
| **1.2**<br>**Data collection mode(s)** | The NHIS is typically conducted via face-to-face interviews in respondents' homes, but telephone interviews may also be conducted when the respondent requests a telephone interview, or when road conditions or travel distances would make it difficult to schedule a personal visit before the required completion date. In 2023, 54.5% of Sample Adult and 54.5% of Sample Child interviews were completed at least partially by telephone. |
| **1.3**<br>**Target population** | The target population for the NHIS is the civilian noninstitutionalized population residing within the 50 states and the District of Columbia at the time of the interview. Active-duty military personnel, civilians living on military bases, persons in long-term care and correctional facilities, and persons with no fixed household address are excluded. |
| **1.4**<br>**Sample design** | NHIS is a cross-sectional household interview survey and uses geographically clustered sampling techniques to select the sample of dwelling units for the NHIS. Commercial address lists are used as the main source of addresses, supplemented by field listing in selected areas. One adult (the sample adult) is randomly selected from each household and responds for themselves (unless physically or cognitively unable). One child (the sample child) is randomly selected in households with children, and an adult knowledgeable and responsible for the child's health responds for them. |
| **1.5**<br>**Survey response rate(s)** | Response rates are calculated using the RR2 formula of the American Association of Public Opinion Research (AAPOR, 2023). For the 2023 NHIS, the final Sample Child response rate was 44.9%, and the final Sample Adult response rate was 47.0%. |
| **2.1**<br>**Missingness rates** | For most measures in the NHIS, percentages with unknown values are typically small (less than 5% and often less than 1%). However, item nonresponse for family income is higher and not random, which may result in biased analyses if observations with missing income are excluded. Thus, missing family income was imputed using multiple imputation methodology (10 values). |
| **2.2**<br>**Observation deletion** | A limited number of fully complete and sufficiently complete cases were removed from the Sample Adult and Sample Child files because of concerns related to data quality (e.g., too many "don't know" or "refused" responses; see *2023 Paradata File Description*). |
| **2.3**<br>**Sample sizes** | Sample sizes for the 2023 NHIS are as follows: 29,522 Sample Adults and 7,692 Sample Children in 30,670 households. |

| | |
|---|---|
| **2.4**<br>**Confidence intervals and standard errors** | NCHS recommends using two-sided 95% confidence intervals calculated using the Clopper-Pearson method adapted for complex surveys by Korn and Graubard. Standard errors used in this calculation should be obtained using statistical software that takes into account the complex sampling design of NHIS. NCHS uses the Taylor series linearization method for variance estimation (see Parker JD, Talih M, Malec DJ, et al., 2017: *National Center for Health Statistics Data Presentation Standards for Proportions*). |
| **2.5**<br>**Weighting** | The Final Annual Weight should be used to generate national estimates based on the NHIS. This weight accounts for sampling probabilities and nonresponse. Weight variables are WTFA_A for Sample Adults and WTFA_C for Sample Children. |
| **2.6**<br>**Variance estimation** | Because of the complex nature of the NHIS sampling design, key nesting variables were created to capture explicit stratification and to identify clustering for a more accurate estimation of the sampling error. PSTRAT (stratum) and PPSU (primary sampling unit) are versions of the sample design variables created for the public-use data files in order to protect the identity of survey respondents. Analysts who instead apply simple random sampling techniques to NHIS data generally will produce standard error estimates that are, on average, too small and are likely to produce results that are subject to excessive Type I error.<br><br>Variance estimates in NCHS reports are typically based on stratum and PSU definitions in NHIS restricted-use data files. Analysts using PSTRAT and PPSU from the public-use data files may not be able to perfectly match the variance estimates published by NCHS. |
| **2.7**<br>**Subpopulation analysis** | To compute accurate standard errors, NCHS recommends that subpopulation analyses be carried out using the full data file instead of subsetting the data file to the subpopulation of interest. This can be done with the SUBPOPN statement in SUDAAN or an equivalent procedure with another software package that supports complex design variance estimation. Sample code for Stata, SAS, SPSS and R is available in the *2023 Survey Description*. |
| **2.8**<br>**Suppression rules** | NCHS recommends suppressing percentages that do not meet the criteria specified in *National Center for Health Statistics Data Presentation Standards for Proportions*. These standards include thresholds for sample size, absolute and relative confidence interval width, and degrees of freedom. |
| **2.9**<br>**Software and code** | To appropriately analyze NHIS data, it is necessary to utilize weights and variance estimation variables and software that can appropriately analyze complex samples. This includes, but is not limited to, SUDAAN, Stata, SPSS, SAS, and R software packages. Sample code for several of these software packages is available in the *2023 Survey Description*. |
| **2.10**<br>**Singleton problem** | Taylor Series Linearization requires at least two PSUs per stratum for variance estimation. NCHS uses the MISSUNIT option in SUDAAN to account for the presence of strata with only one PSU. |

| 2.11 **Public/restricted data** | For data users and researchers throughout the world, public-use data files are freely available on the internet. Analysts interested in working with data that were suppressed or edited to protect confidentiality may apply to access unmodified restricted-use data files through the NCHS Research Data Center (RDC). Codebooks listing variables available only on restricted-use files are available on the NHIS website. |
|---|---|
| 2.12 **Embedded experiments** | 2023 NHIS data collection did not include any embedded experiments. |

## Suggested Citation

All information in this document has been summarized from the *2023 Survey Description*:

National Center for Health Statistics. National Health Interview Survey, 2023 survey description. 2024. Available from: https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2023/srvydesc-508.pdf

Therefore, we recommend that users cite the *2023 Survey Description* when reporting any of the information contained in this document.

## References

American Association for Public Opinion Research (AAPOR). 2023. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys (10th edition). AAPOR.

National Center for Health Statistics. Paradata File Description, National Health Interview Survey, 2023. 2024. Available from: https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2023/srvydesc-paradata-508.pdf.

Parker JD, Talih M, Malec DJ, et al. National Center for Health Statistics data presentation standards for proportions. National Center for Health Statistics. Vital Health Stat 2(175). 2017. Available from: https://www.cdc.gov/nchs/data/series/sr_02/sr02_175.pdf.

Seidenberg AB, Moser RP, West BT. Preferred Reporting Items for Complex Sample Survey Analysis (PRICSSA). J Surv Stat Methodol 11(4). 2023. Available from: https://doi.org/10.1093/jssam/smac040.