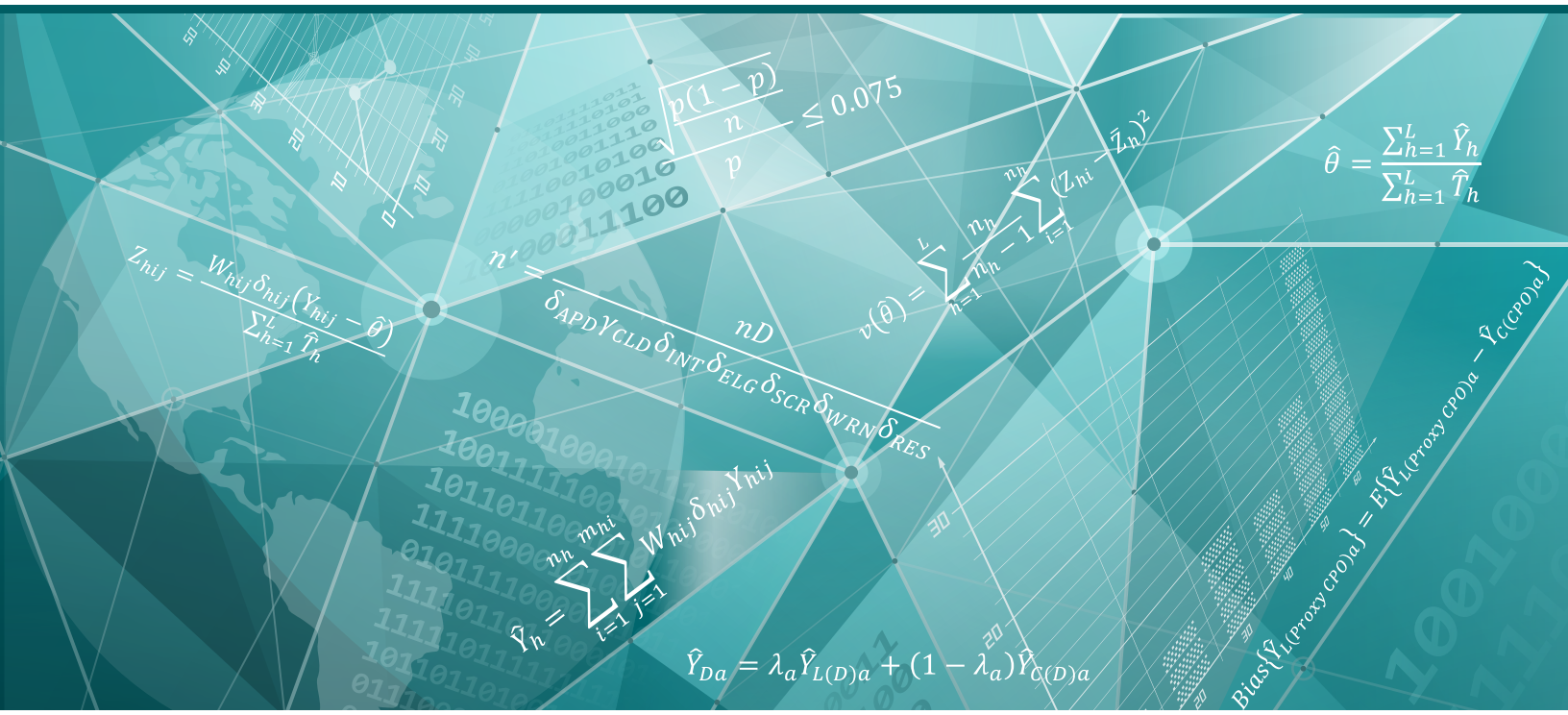
NATIONAL CENTER FOR HEALTH STATISTICS

Vital and Health Statistics

Series 2, Number 186

March 2021



Assessing Linkage Eligibility Bias in the National Health Interview 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

NCHS reports can be downloaded from: https://www.cdc.gov/nchs/products/index.htm.

Suggested citation

Aram J, Zhang C, Golden C, Zelaya CE, Cox CS, Ye Y, Mirel LB. Assessing linkage eligibility bias in the National Health Interview Survey. National Center for Health Statistics. Vital Health Stat 2(186). 2021. DOI: https://dx.doi.org/10.15620/cdc:100468.

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 186 March 2021

Assessing Linkage Eligibility Bias in the National Health Interview 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 2021

Contents

Text Figures

Detailed Tables

Assessing Linkage Eligibility Bias in the National Health Interview Survey

by Jonathan Aram, Cindy Zhang, Cordell Golden, Carla E. Zelaya, Christine S. Cox, Yeats Ye, and Lisa B. Mirel

Abstract

Background

Linking health survey data to administrative records expands the analytic utility of survey participant responses, but also creates the potential for new sources of bias when not all participants are eligible for linkage. Residual differences—bias—can occur between estimates made using the full survey sample and the subset eligible for linkage.

Objective

To assess linkage eligibility bias and provide examples of how bias may be reduced by changes in questionnaire design and adjustment of survey weights for linkage eligibility.

Methods

Linkage eligibility bias was estimated for various sociodemographic groups and health-related variables for the 2000–2013 National Health Interview Surveys.

Conclusions

Analysts using the linked data should consider the potential for linkage eligibility bias when planning their analyses and use approaches to reduce bias, such as survey weight adjustments, when appropriate.

Keywords: linkage consent • data integration • survey weights • National Center for Health Statistics Data Linkage Program

Introduction

Linked survey and administrative data can be used to facilitate richer analyses by supplementing the information collected from the surveys with vital or other administrative data. The National Center for Health Statistics (NCHS) has a data linkage program that is designed to expand the analytic utility of the Center's surveys. To be included in the linkages, participants must meet linkage-eligible (LE) criteria approved by the NCHS Research Ethics Review Board. The determination of linkage eligibility has changed over time and differs across surveys. However, linkage eligibility is based largely on survey participants providing consent and the necessary personally identifiable information (PII) to enable linkage. This report focuses on the linkage eligibility criteria used for the linkage of National Health Interview Survey (NHIS) data to Centers for Medicare & Medicaid Services (CMS) administrative records. The process and criteria used to determine linkage eligibility varies slightly depending on the specific NCHS survey, survey year, and administrative data source included in a linkage. However, the concepts and techniques presented here are applicable to any of the NCHS linked data files (see: https://www.cdc.gov/nchs/data-linkage/index.htm).

The process and criteria for collecting PII for NHIS has changed over time, resulting in changes in linkage eligibility criteria and an increase in the percentage of survey participants who are eligible for linkage. Analysts should be aware that estimates made using the full sample of all NHIS participants, regardless of linkage eligibility, may differ from estimates made using only the LE subset of survey participants. This difference, which is referred to as linkage eligibility bias, may vary by survey year, the statistic being calculated, and the survey weight used in estimation.

For example, in 2005, the estimated prevalence of diabetes among all NHIS sample adults was 7.4%. The estimated prevalence among LE sample adults was 8.1%. The discrepancy between these two estimates is the linkage eligibility bias.

Note that linkage eligibility is distinct from program eligibility. This analysis focuses on those participants who were eligible for linkage, regardless of whether they were actually linked to the administrative data source.

Prior to 2007, consent for data linkage for NHIS participants was based on their willingness to provide key direct identifiers, such as the nine digits of their Social Security Number (SSN9). Specifically, for linkage with CMS administrative data, NHIS participants were considered to have consented to linkage activities if they did not refuse to answer when asked to provide their SSN9 or Medicare health insurance claim number (HICN). But over time, the number of survey participants refusing to provide their SSN9s or HICNs increased (1). To address increasing respondent reluctance to provide their SSN9s or HICNs, and to separate consent determination from the provision of an SSN9 and HICN, NHIS

changed its linkage consent procedures in 2007. In the new procedure, survey participants were asked for only the last four digits of their SSN (SSN4) or the last four digits plus the alphanumeric beneficiary identification code of their HICN (HICN4). Additionally, those who refused, said "don't know," or had a missing value when asked to provide the SSN4 or HICN4 were then asked if they would consent to linkage without the SSN4 or HICN4 (Figure 1). In addition, beginning in 2007, consent was sought for sample adults only, who consent for themselves, and sample children, whose parents grant consent for them (2,3).

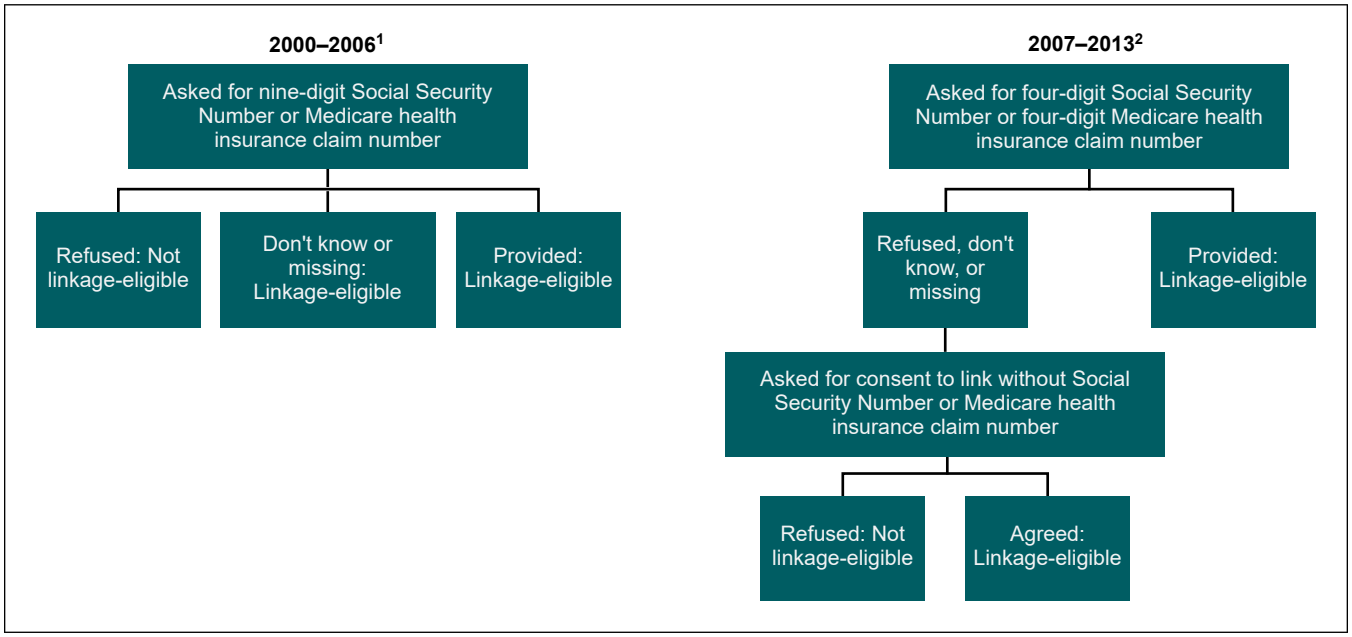
These changes in data collection improved linkage consent rates (Figure 2). However, until now, an evaluation of how this change impacted linkage eligibility bias has not been done. This analysis measures bias in the estimated prevalence of smoking, diabetes, high blood pressure, and other health-related conditions. These measures were chosen as examples of types of health indicators that researchers may consider when analyzing linked data, but they are not comprehensive of all content included in NHIS. Bias also was calculated for the estimated percentages of participants categorized by sex, age, education, and race and ethnicity in the NHIS target population, given their importance in assessing health disparities and social determinants of health.

Methods

NHIS Background

NCHS has administered NHIS, a nationally representative, cross-sectional population health survey, continuously since 1957. NHIS is the principal source of information on the health of the U.S. civilian noninstitutionalized population. Households are selected through a probability sampling frame drawn from each state and the District of Columbia, based on information from the decennial census. The NHIS sample design has been described in more detail elsewhere (4,5). Beginning in 1997, NHIS implemented a questionnaire redesign to obtain more detailed health information for selected individuals within a sampled household. Within each household, families are identified, and a family respondent completes a brief structured interview on family demographics and broad health measures. From each family in NHIS, one adult aged 18 and over (the "sample adult") and, if present, one child (the "sample child") are randomly selected, and information on each is collected from the Sample Adult and the Sample Child interviews. The content of these two interviews differs on some items, but both collect basic information on health status, health care services, and health behaviors. For the Sample Adult interview, the selected individual responds for himself or herself (i.e., no proxy response is allowed, except when the person is unable to respond due to physical or mental condition) (4). This report presents results based on sample adult participants.

Figure 1. National Health Interview Survey linkage eligibility criteria for CMS linkages, 2000–2013



[1]The process for collecting Social Security Numbers (SSN) and Medicare health insurance claim numbers (HICN) varied during this time frame. The questions may have been asked of all National Health Interview Survey participants (2000–2001), only family respondents (2002–2003), or family respondents, sample adults, and sample children (2004–2006).
[2]The process for collecting SSN and HICN and the direct question about consent to link was asked of sample adults and sample children only.
NOTE: CMS is Centers for Medicare & Medicaid Services.
SOURCE: National Center for Health Statistics, "The Linkage of National Center for Health Statistics Surveys to Medicare Enrollment and Claims Data: Methodology and Analytic Considerations."

NCHS Data Linkage Program

The NCHS Data Linkage Program is designed to maximize the scientific value of the Center's population-based surveys by linking data from health surveys with data from vital and other administrative records. More information about the program is available from: https://www.cdc.gov/nchs/data-linkage/index.htm. Linked data files enable researchers to examine factors that influence disability, chronic disease, health care utilization, and mortality. Linked data files also maximize scientific value by creating an efficient means to expand the analytic potential of both the survey and administrative data, so enabling analyses that would not be possible with either data source alone. To date, the NCHS survey data have been linked to the National Death Index, administrative data from CMS, the Social Security Administration, and the Department of Housing and Urban Development. This report uses the definition of linkage eligibility included in the NHIS–CMS linkage methodology (2,3).

Methods for Calculating Linkage Eligibility Bias

Bias was estimated using a method similar to a previous study of linkage eligibility (6). The method involves finding the difference in estimates made using the LE sample and full sample, then dividing by the estimate made using the full sample. Because this measure took both positive and negative values, the authors of this method, Sakshaug et. al, took the absolute value, and referred to the calculation as "absolute relative bias" (6). The main advantage of this method is that bias calculations are scaled by the estimate. As an example, consider two scenarios in which estimates differ by 5 percentage points. When the estimates are close to 100%, the bias is smaller: ((95% – 90%) / 90%) • 100 = 5.6%. When the estimates are close to 0%, the bias is greater: ((15% – 10%) / 10%) • 100 = 50.0%. As demonstrated by these two examples, a difference of 5 percentage points results in less bias when the true value is 90% and more bias when the true value is 10%. This mirrors real-world conditions, where an estimate of 90%, give or take 5 percentage points, may be more informative than an estimate of 10%, give or take 5 percentage points.

Reducing Bias With Weight Adjustment

When analyzing linked survey and administrative data, preliminary guidance from NCHS recommends adjusting survey weights for linkage eligibility using the same control total domains that are used to create the original weight: sex, age, and race and ethnicity (7). Adjusting weights for linkage eligibility is similar to nonresponse adjustment. LE participants are treated like respondents, and linkage-ineligible participants are treated like nonrespondents. This adjustment is expected to reduce bias for the variables used as control totals, but it is unclear how it will affect bias for other variables, such as the prevalence of diabetes, hypertension, and other health-related variables. By adding one or more of these variables as control totals in the weight adjustment procedure, it may be possible to reduce bias in estimates of health-related variables as well. This report compares bias estimates made using the original survey weights and adjusted survey weights.

Study Population

To account for changes to linkage eligibility criteria (e.g., family respondents and sample adults) over time, all analyses were limited to the sample adults who were at least age 18 at the time of the NHIS interview. All participants under age 18 were excluded because the consent process for children differs from the process for adults.

Percent Linkage-eligible

For each NHIS sample adult survey from 2000 to 2013, the percentage of participants who were eligible for linkage (i.e., "percent linkage-eligible") was calculated by dividing the number of LE sample adults by the total number of sample adults. The years 2000 to 2013 were chosen to account for the changes in questionnaire design made to improve linkage consent rates. The years leading up to the questionnaire change (2000 to 2006) highlight a time period when linkage consent was at its lowest, with less than 50% of survey participants eligible for linkage during some years. The design changes occurred in 2007 and, at the time this analysis was begun, 2013 was the most recent year of linked CMS data available for analysis.

Survey Weights

Original weight

The survey weights that accompanied the release of the NHIS public-use survey data files were used as the original weight. The original weight reflects the inverse of the probability of selection as an NHIS participant. The original weight was adjusted for nonresponse and accounted for the sample design.

Adjusted weight 1

The original weight was adjusted for linkage eligibility using PROC WTADJUST in SUDAAN software (8). More detailed information about adjusting survey weights for linkage eligibility using SUDAAN is available elsewhere (7). For this analysis, a model was fit using a three-way interaction among age category, race and ethnicity, and sex. Due to changes in how NHIS collected information on race and ethnicity during the study period, two different race and ethnicity variables were used. For 2000–2005, a three-level (Hispanic, non-Hispanic black, and non-Hispanic white or other) variable was used. For 2006–2013, a four-level (Hispanic, non-

Hispanic black, non-Hispanic Asian, and non-Hispanic white or other) variable was used (4). SAS code for adjusted weights for linkage eligibility derived using this model (referred to as "adjusted weight 1") is provided in the appendix to this report.

Adjusted weight 2

Preliminary analyses showed that adjusted weight 1 decreased bias for age, sex, and race and ethnicity estimates, but increased bias in estimates of the prevalence of some health conditions. The authors of this report hypothesized that this unintended consequence of the first weight adjustment could be addressed by adding one or more health variables to the weight adjustment model. Specifically, diabetes was added as a main effect in the model already containing a three-way interaction of age category, race and ethnicity, and sex used to create adjusted weight 1. This new set of weights (adjusted weight 2) was then used to re-estimate the prevalence of all health-related variables (including but not limited to diabetes), as well as age, sex, and education. Some data for the variable diabetes are missing (about 0.1% of the analytic sample). This was due to participants who refused or did not know how to answer the diabetes question. Following an example used in previous analyses, missing values were imputed before the weights were adjusted (9). Hot-deck imputation (PROC SURVEYIMPUTE in SAS) was used to fill in missing values for diabetes so weights could be calculated for all LE participants (10). The imputed values were not used to calculate the prevalence estimates of diabetes.

Bias Calculations Used in This Report

General relative bias calculation

Bias was analyzed by comparing weighted estimates made using LE sample adults to weighted estimates made using all sample adults. Full sample estimates were considered the true value in all comparisons. Point estimates and the corresponding variances for this analysis were calculated using PROC SURVEYMEANS in SAS, which accounts for the complex design of NHIS (11). Multiple survey years were combined for the presentation of findings related to bias. The groups used were 2000–2003, 2004–2006, 2007–2009, and 2010–2013. These groupings were selected to highlight the time period before and after the questionnaire design change in 2007. Single-survey-year estimates of bias are available by contacting the NCHS Data Linkage Team (datalinkage@cdc.gov).

Linkage eligibility bias was estimated using a modified version of the "absolute relative bias" method (6). The difference between estimates made using the LE sample and the full sample was divided by the estimate made using the full sample. All bias estimates were multiplied by 100 to improve interpretability (equation 1).

$$Relative\,Bias = \frac{(Linkage\,eligible\,sample\,estimate\ -\ Full\,sample\,estimate)}{Full\,sample\,estimate} \cdot 100$$

Because the LE sample was a subset of the full sample, statistical testing was not conducted, due to lack of independence. Bias was assessed based on magnitude. If bias was between 0.0% and 0.1%, the two estimates were considered similar (i.e., no bias) (12). Unlike previous analyses, this analysis reported whether bias was positive or negative (others have removed the sign and reported all results as positive values) (6). In this analysis, the maximum bias is reported.

Bias was examined for the following sociodemographic groups: age (18–44, 45–64, 65 and over), sex (female, male), and race and ethnicity (Hispanic; non-Hispanic black; non-Hispanic Asian, for 2006–2013 only; non-Hispanic white or other), education level (less than high school, high school diploma or GED, some college, bachelor's degree or higher). Bias also was estimated for the following health-related variables: diabetes, hypertension, smoked 100 cigarettes over entire lifetime, obesity, fair or poor health status, visit to the doctor in the past year. Health-related variables were dichotomized as "Yes" or "No." If a survey participant refused to answer, or if the response was not ascertained or the participant responded "don't know," that variable response was treated as missing. Body mass index (BMI) was calculated as weight in kilograms divided by height in meters squared, and obesity was defined as BMI greater than or equal to 30. For women who were pregnant at the time of the NHIS interview, BMI was set to missing. The wording of the health questions and SAS code for creating dichotomous health variables are provided in the appendix to this report. The questions were comparable over time.

Standard errors were calculated for all estimates and shown in Table 3.

Relative bias calculation, original weight

Relative bias in estimates made using the original weight was calculated as follows: The estimate made using the full sample with the original weight was subtracted from the estimate made using the LE sample with the original weight. The difference was then divided by the estimate made using the full sample with the original weight. The resulting measure was multiplied by 100 to improve interpretability (equation 2).

$$Relative\,Bias_{original\,weight} = \frac{(Linkage\,eligible\,sample\,estimate_{original\,weight}\ -\ Full\,sample\,estimate_{original\,weight})}{Full\,sample\,estimate_{original\,weight}} \cdot 100$$

Relative bias calculation, adjusted weight 1

Relative bias in estimates made using adjusted weight 1 was calculated as follows: The estimate made using the full sample with the original weight was subtracted from the estimate made using the linkage-eligible sample with adjusted weight 1. The difference was then divided by the estimate made using the full sample with the original weight. The resulting measure was multiplied by 100 to improve interpretability (equation 3).

$$Relative\,Bias_{adjusted\,weights1} = \frac{(Linkage\,eligible\,sample\,estimate_{adjusted\,weight\,1} - Full\,sample\,estimate_{original\,weight})}{Full\,sample\,estimate_{original\,weight}} \cdot 100$$

Relative bias calculation, adjusted weight 2

Preliminary results indicated that bias in the estimates of some variables, like diabetes, was not reduced by the use of adjusted weight 1. Consequently, as an example, a second set of adjusted weights was created. The second set included diabetes as a main effect in the weight adjustment model along with the same three-way interaction of age, race and ethnicity, and sex used to create adjusted weight 1. Then, bias was re-estimated for all variables using the second set of adjusted weights (adjusted weight 2). Similar to the other bias calculations in this report, the estimate made using the full sample with the original weight was subtracted from the estimate made using the linkage-eligible sample with adjusted weights. The difference was then divided by the estimate made using the full sample with the original weight (equation 4).

$$Relative\,Bias_{adjusted\,weights2} = \frac{(Linkage\,eligible\,sample\,estimate_{adjusted\,weight2} - Full\,sample\,estimate_{original\,weight})}{Full\,sample\,estimate_{original\,weight}} \cdot 100$$

Comparison of weights

As a final assessment, the distributions of the weights were compared, and the influence on variance inflation was assessed by comparing the estimated design effects due to the differential weighting. The estimated design effect due to differential weighting was defined as 1 plus the coefficient of variation (CV) squared ($1 + CV^2$) (13).

Results

Figure 2 shows the percentage eligible for linkage across all groups by year, using unweighted counts. Maximum linkage eligibility occurred in 2010, when 88.9% of the study sample was eligible for linkage.

Table 1 shows the percentage of the study population eligible for linkage by demographic and health-related variables, using unweighted counts for the time periods. Across all groups, the maximum linkage eligibility occurred during 2010–2013. In this time frame, at least 88% of NHIS sample adults were eligible for linkage. In contrast, the overall percentage eligible for linkage was as low as 46% before the data collection procedures and linkage eligibility criteria were revised in 2007.

Table 2 shows estimates of the weighted percentage of all demographic and health characteristics for the full and LE NHIS sample adults. Four estimates are presented for each variable and survey year:

1. An estimate made using the full sample with the original weight
2. An estimate made using only the LE sample with the original weight
3. An estimate made using only the LE sample with adjusted weight 1
4. An estimate made using only the LE sample with adjusted weight 2

Table 2 highlights population distribution differences resulting from linkage eligibility bias and the extent to which these discrepancies can be reduced using adjusted weights. It also shows three linkage eligibility bias estimates for each demographic characteristic, health characteristic, and survey year:

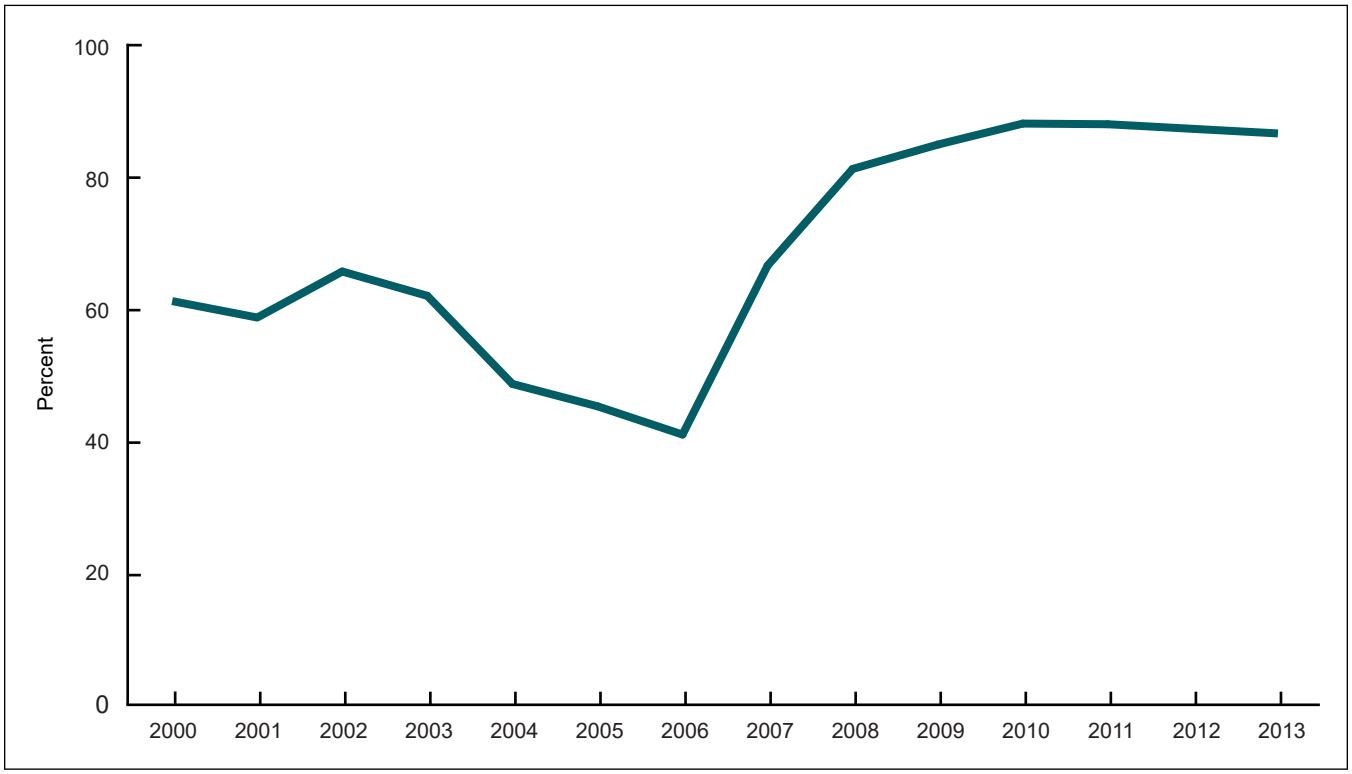
1. Bias when estimates are made with the original weight
2. Bias when estimates are made using adjusted weight 1
3. Bias when estimates are made using adjusted weight 2

In most cases, bias was smaller when estimates were made using either set of adjusted weights, compared with the original weight. This was true especially for the demographic variables used in the adjusted weights models, where the LE sample produced the same estimates as the full sample with the original weight, and no detectable bias was observed. For education and health-related variables, adjusted weights reduced bias in some, but not all, cases. More information about these two groups of variables is provided in the following sections.

Demographic Variables

For age, sex, and race and ethnicity, use of the original weight resulted in bias. However, no remaining detectable bias was observed when estimates were made using either set of adjusted weights. Specifically, for age groups, the maximum bias using the original weight was –4.6% (ages 45–64, 2004–2006) and –4.5% (ages 65 and over, 2000–2003). For sex, the maximum bias was –3.3% and 3.6% (females and males, respectively, 2000–2003). For race and ethnicity, the maximum bias was –34.1% (non-Hispanic Asian, 2006). Due to changes in the NHIS sample design, 2006 was

Figure 2. Percentage of National Health Interview Survey sample adults eligible for linkage, by survey year, 2000–2013



SOURCE: National Center for Health Statistics, National Health Interview Survey–Centers for Medicare & Medicaid Services Feasibility Files.

the only year of NHIS that included the non-Hispanic Asian race and ethnicity category before the change in linkage consent procedures, which changed beginning in 2007. However, the bias for the non-Hispanic Asian subgroup remained over 10% in the period immediately following the change in consent procedures (2007–2009). In most cases, the maximum bias for the demographic variables was observed during 2004–2006, and it often coincided with low linkage eligibility rates (Tables 1,2).

For education, the use of the original weight and adjusted weights resulted in bias. For the measures no high school diploma and bachelor's degree or higher, the maximum bias was observed during 2004–2006. The direction and magnitude of bias varied by education level and the weight used in estimation. Using the original weight, the maximum bias was 14.0% (no high school, 2004–2006) and –10.5% (bachelor's degree or higher, 2004–2006). Using adjusted weights, the maximum bias was 13.1% (no high school, 2004–2006, adjusted weight 1) and –8.9% (bachelor's degree or higher, 2004–2006, adjusted weight 1). For high school diploma and some college, bias was generally low, less than 5%, regardless of original or adjusted weights and time period (Table 2).

Health-related Variables

For diabetes, the use of the original weight and adjusted weight 1 both resulted in bias, but for most other health-related variables, little difference in bias was observed according to original or adjusted weights. No bias was seen in the prevalence of diabetes when estimates were made using adjusted weight 2, because diabetes was used as a main effect in the weight adjustment model. By design, this eliminated the bias since the control totals were fixed for the group with diabetes. Using the original weight, the maximum bias found within the health-related variables was 11.6% (fair or poor self-rated health, 2004–2006). The maximum bias increased to 14.0% (fair or poor self-rated health, 2004–2006) using adjusted weight 1, and decreased slightly to 12.3% with adjusted weight 2. For most health-related variables, bias was higher during 2004–2006, corresponding to lower linkage eligibility rates (Table 2).

Standard Errors and Confidence Intervals

The standard error for each estimate presented previously is shown in Table 3. In general, the minimum standard error was observed when estimates were made using the full sample, although this varied somewhat by time period. When estimates were made using the LE sample, little or no difference in standard errors was seen when each of the three weights was applied, although the standard errors generally increased (Table 3).

Comparison of Weights

The summary statistics for the different weights in Table 4 show similar estimated design effects due to differential weighting for the original weight and the adjusted weights, which implies that the weight adjustments had little effect on variance inflation. The mean and median values of the original weight are lower than the mean and median adjusted weights (approximately 8,000 compared with 12,000 for the mean, and 6,600 compared with 10,000 for the median). This is likely because the original weight was calculated for a larger number of participants and had a lower maximum value than either set of adjusted weights.

Discussion

For this study sample, changes in the data collection procedures and the way participants consent to record linkage resulted in higher linkage eligibility over time. Specifically, the percentage of NHIS sample adults eligible for record linkage increased from approximately 40% in 2006 to almost 90% starting in 2010, after the NHIS linkage questions and eligibility criteria changed in 2007. For many variables, the maximum observed bias occurred prior to 2007. For the variables used to create adjusted weight 1 (age, sex, race and ethnicity), bias was reduced to less than 0.1%. Use of adjusted weight 1 also resulted in lower bias for education and some, but not all, health-related variables.

To determine whether bias could be reduced further by adding an additional variable to the weight adjustment procedure, a dichotomous diabetes variable was included in the weight adjustment model with age, sex, and race and ethnicity. Using adjusted weight 2 reduced bias in the estimated prevalence of diabetes to less than 0.1% because diabetes was included as a main effect in the weight adjustment model. For other health-related variables, bias was similar when estimates were made using adjusted weights 1 and 2. These results suggest that specific health-related variables may be used for weight adjustment without affecting the inference on the other variables used in the weight adjustment model. However, ongoing research examines the use of adjusted survey weights for linked data. Additionally, NCHS and other federal statistical agencies are developing common frameworks for assessing the quality of integrated or linked data products and providing guidance to data users on appropriate statistical techniques to reduce potential biases such as those presented in this report (information on this progress is available from: https://nces.ed.gov/FCSM/pdf/Quality_Integrated_Data.pdf).

Although this report focuses on the linked eligible sample using the NHIS data, similar analyses and assessments could be conducted using other NCHS surveys that are linked to vital and other administrative data. Analysts using NCHS linked data should consider estimating linkage eligibility bias when selecting survey years for analysis and applying bias reduction strategies such as adjusting survey weights for linkage eligibility. In conclusion, changes in data collection procedures and linkage eligibility criteria were associated with decreased linkage eligibility bias, and residual bias may be reduced with weight adjustment, particularly for demographic characteristics.

References

1. Miller DM. Trends in record linkage refusal rates: Characteristics of National Health Interview Survey participants who refuse record linkage. In: Joint Statistical Meeting. Miami Beach, FL: American Statistical Association. 2011.
2. National Center for Health Statistics. The linkage of National Center for Health Statistics survey data to Medicaid enrollment and claims data: Methodology and analytic considerations. 2019. Available from: https://www.cdc.gov/nchs/data/datalinkage/nchs-medicaid-linkage-methodology-and-analytic-considerations.pdf.
3. National Center for Health Statistics. The linkage of National Center for Health Statistics surveys to Medicare enrollment and claims data: Methodology and analytic considerations. 2017. Available from: https://www.cdc.gov/nchs/data-linkage/cms/nchs_medicare_linkage_methodology_and_analytic_considerations.pdf.
4. Parsons VL, Moriarity C, Jonas K, Moore TF, Davis KE, Tompkins L. Design and estimation for the National Health Interview Survey, 2006–2015. National Center for Health Statistics. Vital Health Stat 2(165). 2014.
5. Botman SL, Moore TF, Moriarity CL, Parsons VL. Design and estimation for the National Health Interview Survey, 1995–2004. National Center for Health Statistics. Vital Health Stat 2(130). 2000.
6. Sakshaug JW, Huber M. An evaluation of panel nonresponse and linkage consent bias in a survey of employees in Germany. J Surv Stat Methodol 4(1):71–93. 2015.
7. Golden C, Driscoll AK, Simon AE, Judson DH, Miller EA, Parker JD. Linkage of NCHS population health surveys to administrative records from Social Security Administration and Centers for Medicare & Medicaid Services. National Center for Health Statistics. Vital Health Stat 1(58). 2015.
8. RTI International. SUDAAN (Release 11.0.0) [computer software]. 2012.
9. Hing E, Shimizu IM, Talwalkar A. Nonresponse bias in estimates from the 2012 National Ambulatory Medical Care Survey. National Center for Health Statistics. Vital Health Stat 2(171). 2016.

10. SAS Institute. SAS/STAT 14.1 user's guide: The SURVEYIMPUTE procedure. 2015.

11. SAS Institute. SAS/STAT 14.1 user's guide: The SURVEYMEANS procedure. 2015.

12. Lloyd PC, Helms VE, Simon AE, Golden C, Brittain J, Call E, et al. Linkage of 1999–2012 National Health Interview Survey and National Health and Nutrition Examination Survey data to U.S. Department of Housing and Urban Development administrative records. National Center for Health Statistics. Vital Health Stat 1(60). 2017.

13. Mirel L, Burt V, Curtin L, Zhang C. Different approaches for non-response adjustments to statistical weights in the continuous NHANES (2003–04). In: Proceedings of the Federal Committee on Statistical Methodology Research Conference. Washington, DC: Council of Professional Associations on Federal Statistics. 2009.

Table 1. Unweighted number and percentage of National Health Interview Survey sample adults eligible for linkage, by survey year and demographic and health-related variables, 2000–2013

Selected variables	2000–2003	2004–2006	2007–2009	2010–2013
Total				
Full sample	127,596	87,029	72,905	129,253
Linkage-eligible sample	79,810	39,982	57,375	114,074
Percent linkage-eligible	62.5	45.9	78.7	88.3
Ages 18–44:				
Full sample	64,859	41,953	33,831	58,170
Linkage-eligible sample	41,502	19,874	26,695	51,664
Percent linkage-eligible	64.0	47.4	78.9	88.8
Ages 45–64:				
Full sample	38,786	28,332	24,554	43,617
Linkage-eligible sample	23,979	12,477	19,396	38,390
Percent linkage-eligible	61.8	44.0	79.0	88.0
Ages 65 and over:				
Full sample	23,951	16,744	14,520	27,466
Linkage-eligible sample	14,329	7,631	11,284	24,020
Percent linkage-eligible	59.8	45.6	77.7	87.5
Female:				
Full sample	72,188	48,649	40,755	71,743
Linkage-eligible sample	43,855	22,247	31,963	63,142
Percent linkage-eligible	60.8	45.7	78.4	88.0
Male:				
Full sample	55,408	38,380	32,150	57,510
Linkage-eligible sample	35,955	17,735	25,412	50,932
Percent linkage-eligible	64.9	46.2	79.0	88.6
Hispanic:				
Full sample	21,681	15,312	13,034	22,827
Linkage-eligible sample	14,496	7,374	9,422	19,527
Percent linkage-eligible	66.9	48.2	72.3	85.5
Non-Hispanic Asian:				
Full sample	---	1,306	4,101	8,091
Linkage-eligible sample	---	409	2,950	6,926
Percent linkage-eligible	---	31.3	71.9	85.6
Non-Hispanic black:				
Full sample	17,533	12,766	11,554	20,129
Linkage-eligible sample	10,427	5,695	9,263	17,927
Percent linkage-eligible	59.5	44.6	80.2	89.1
Non-Hispanic white or other:				
Full sample	88,382	57,645	44,216	78,206
Linkage-eligible sample	54,887	26,504	35,740	69,694
Percent linkage-eligible	62.1	46.0	80.8	89.1
No high school:				
Full sample	25,323	16,489	12,732	20,935
Linkage-eligible sample	17,115	8,731	9,864	18,437
Percent linkage-eligible	67.6	53.0	77.5	88.1
High school or GED:				
Full sample	36,491	24,510	19,838	33,481
Linkage-eligible sample	22,974	11,097	15,598	29,674
Percent linkage-eligible	63.0	45.3	78.6	88.6
Some college:				
Full sample	35,768	24,288	21,094	39,148
Linkage-eligible sample	22,274	11,332	17,102	34,835
Percent linkage-eligible	62.3	46.7	81.1	89.0
Bachelor's degree or higher:				
Full sample	28,520	20,696	18,687	35,069
Linkage-eligible sample	17,014	8,600	14,566	30,671
Percent linkage-eligible	59.7	41.6	77.9	87.5

See footnote at end of table.

Table 1. Unweighted number and percentage of National Health Interview Survey sample adults eligible for linkage, by survey year and demographic and health-related variables, 2000–2013—Con.

Selected variables	2000–2003	2004–2006	2007–2009	2010–2013
Diabetes:				
Full sample	8,754	7,122	6,746	13,131
Linkage-eligible sample	5,786	3,568	5,501	11,826
Percent linkage-eligible	66.1	50.1	81.5	90.1
Hypertension:				
Full sample	32,626	24,425	22,485	42,047
Linkage-eligible sample	20,902	11,802	18,272	37,622
Percent linkage-eligible	64.1	48.3	81.3	89.5
Obesity:				
Full sample	27,899	20,782	18,940	35,429
Linkage-eligible sample	18,692	10,535	15,642	32,016
Percent linkage-eligible	67.0	50.7	82.6	90.4
Fair or poor self-rated health:				
Full sample	17,248	12,420	10,791	19,731
Linkage-eligible sample	11,193	6,361	8,747	17,673
Percent linkage-eligible	64.9	51.2	81.1	89.6
Doctor's office visit in past year:				
Full sample	101,017	68,652	57,875	102,101
Linkage-eligible sample	63,285	31,512	46,079	90,567
Percent linkage-eligible	62.6	45.9	79.6	88.7
Smoking:				
Full sample	56,135	36,080	30,069	52,244
Linkage-eligible sample	36,266	17,947	24,678	46,830
Percent linkage-eligible	64.6	49.7	82.1	89.6

--- Data not available.

NOTES: Eligibility for linkage was based upon consenting and having sufficient personally identifiable information. For 2000–2006, refusal for data linkage was defined as refusing to provide a Social Security Number (SSN) or Medicare health insurance claim number (HICN). For 2007–2013, refusal for data linkage was defined as not providing the last four digits of the SSN or HICN and responding "No" to the follow-up question to allow linkage without these personal identifiers. Due to sample design changes in 2006, estimates for non-Hispanic Asian persons for 2004–2006 were only from the 2006 National Health Interview Survey.

SOURCE: National Center for Health Statistics, linked National Health Interview Survey–Centers for Medicare & Medicaid Services Feasibility Files.

Table 2. Survey estimates and linkage eligibility bias for National Health Interview Survey sample adults, by survey year, sample weight, and demographic and health-related variables, 2000–2013

Selected variables	2000–2003	2004–2006	2007–2009	2010–2013
Ages 18–44		Percent		
Full sample:				
Original weight	52.8	50.7	49.1	47.6
Linkage-eligible sample:				
Original weight	54.0	52.4	49.4	48.0
Adjusted weight 1	52.8	50.7	49.1	47.6
Adjusted weight 2	52.8	50.7	49.1	47.6
Bias:				
Original weight	2.2	3.3	0.6	0.9
Adjusted weight 1	0.0	0.0	0.0	0.0
Adjusted weight 2	0.0	0.0	0.0	0.0
Ages 45–64				
Full sample:				
Original weight	31.0	33.2	34.4	34.9
Linkage-eligible sample:				
Original weight	30.6	31.6	34.4	34.7
Adjusted weight 1	31.0	33.2	34.4	34.9
Adjusted weight 2	31.0	33.2	34.4	34.9
Bias:				
Original weight	–1.4	–4.6	0.0	–0.6
Adjusted weight 1	0.0	0.0	0.0	0.0
Adjusted weight 2	0.0	0.0	0.0	0.0
Ages 65 and over				
Full sample:				
Original weight	16.1	16.1	16.5	17.5
Linkage-eligible sample:				
Original weight	15.4	16.0	16.2	17.3
Adjusted weight 1	16.1	16.1	16.5	17.5
Adjusted weight 2	16.1	16.1	16.5	17.5
Bias:				
Original weight	–4.5	–0.8	–1.8	–1.3
Adjusted weight 1	0.0	0.0	0.0	0.0
Adjusted weight 2	0.0	0.0	0.0	0.0
Female				
Full sample:				
Original weight	52.0	51.8	51.7	51.7
Linkage-eligible sample:				
Original weight	50.3	51.8	51.6	51.6
Adjusted weight 1	52.0	51.8	51.7	51.7
Adjusted weight 2	52.0	51.8	51.7	51.7
Bias:				
Original weight	–3.3	0.0	–0.3	–0.3
Adjusted weight 1	0.0	0.0	0.0	0.0
Adjusted weight 2	0.0	0.0	0.0	0.0
Male				
Full sample:				
Original weight	48.0	48.2	48.3	48.3
Linkage-eligible sample:				
Original weight	49.7	48.2	48.4	48.4
Adjusted weight 1	48.0	48.2	48.3	48.3
Adjusted weight 2	48.0	48.2	48.3	48.3
Bias:				
Original weight	3.6	0.0	0.3	0.3
Adjusted weight 1	0.0	0.0	0.0	0.0
Adjusted weight 2	0.0	0.0	0.0	0.0

See footnotes at end of table.

Table 2. Survey estimates and linkage eligibility bias for National Health Interview Survey sample adults, by survey year, sample weight, and demographic and health-related variables, 2000–2013—Con.

Selected variables	2000–2003	2004–2006	2007–2009	2010–2013
Hispanic		Percent		
Full sample:				
Original weight	11.2	12.7	13.6	14.5
Linkage-eligible sample:				
Original weight	11.8	13.2	12.4	14.1
Adjusted weight 1	11.2	12.7	13.6	14.5
Adjusted weight 2	11.2	12.7	13.6	14.5
Bias:				
Original weight	5.3	3.6	–8.8	–2.9
Adjusted weight 1	0.0	0.0	0.0	0.0
Adjusted weight 2	0.0	0.0	0.0	0.0
Non-Hispanic Asian				
Full sample:				
Original weight	---	1.5	4.6	5.1
Linkage-eligible sample:				
Original weight	---	1.0	4.2	4.9
Adjusted weight 1	---	1.5	4.6	5.1
Adjusted weight 2	---	1.5	4.6	5.1
Bias:				
Original weight	---	–34.1	–10.2	–3.4
Adjusted weight 1	---	0.0	0.0	0.0
Adjusted weight 2	---	0.0	0.0	0.0
Non-Hispanic black				
Full sample:				
Original weight	11.3	11.4	11.8	11.9
Linkage-eligible sample:				
Original weight	10.9	11.2	11.9	12.0
Adjusted weight 1	11.3	11.4	11.8	11.9
Adjusted weight 2	11.3	11.4	11.8	11.9
Bias:				
Original weight	–3.4	–1.7	1.0	0.7
Adjusted weight 1	0.0	0.0	0.0	0.0
Adjusted weight 2	0.0	0.0	0.0	0.0
Non-Hispanic white or other				
Full sample:				
Original weight	77.5	74.3	70.0	68.5
Linkage-eligible sample:				
Original weight	77.3	74.6	71.6	69.0
Adjusted weight 1	77.5	74.3	70.0	68.5
Adjusted weight 2	77.5	74.3	70.0	68.5
Bias:				
Original weight	–0.3	0.3	2.2	0.7
Adjusted weight 1	0.0	0.0	0.0	0.0
Adjusted weight 2	0.0	0.0	0.0	0.0
No high school				
Full sample:				
Original weight	17.2	16.7	15.3	14.1
Linkage-eligible sample:				
Original weight	18.4	19.0	15.0	14.1
Adjusted weight 1	18.3	18.9	15.4	14.2
Adjusted weight 2	18.2	18.8	15.4	14.2
Bias:				
Original weight	6.5	14.0	–1.8	–0.3
Adjusted weight 1	6.0	13.1	0.7	0.5
Adjusted weight 2	5.8	12.8	0.6	0.4

See footnotes at end of table.

Table 2. Survey estimates and linkage eligibility bias for National Health Interview Survey sample adults, by survey year, sample weight, and demographic and health-related variables, 2000–2013—Con.

Selected variables	2000–2003	2004–2006	2007–2009	2010–2013
High school or GED		Percent		
Full sample:				
Original weight	29.9	29.3	28.1	26.5
Linkage-eligible sample:				
Original weight	30.1	29.1	28.0	26.6
Adjusted weight 1	30.2	29.1	28.1	26.6
Adjusted weight 2	30.2	29.1	28.0	26.6
Bias:				
Original weight	0.8	–0.6	–0.4	0.4
Adjusted weight 1	1.1	–0.5	–0.3	0.4
Adjusted weight 2	1.0	–0.5	–0.3	0.4
Some college				
Full sample:				
Original weight	29.0	28.7	29.7	31.0
Linkage-eligible sample:				
Original weight	28.8	29.2	30.3	31.2
Adjusted weight 1	28.7	28.9	30.1	31.1
Adjusted weight 2	28.7	28.9	30.1	31.0
Bias:				
Original weight	–0.8	1.7	2.1	0.6
Adjusted weight 1	–1.1	0.8	1.2	0.3
Adjusted weight 2	–1.1	0.7	1.2	0.3
Bachelor's degree or higher				
Full sample:				
Original weight	23.9	25.3	26.9	28.4
Linkage-eligible sample:				
Original weight	22.8	22.7	26.6	28.2
Adjusted weight 1	22.9	23.1	26.5	28.2
Adjusted weight 2	22.9	23.1	26.5	28.2
Bias:				
Original weight	–4.7	–10.5	–0.9	–0.9
Adjusted weight 1	–4.3	–8.9	–1.4	–0.9
Adjusted weight 2	–4.2	–8.6	–1.3	–0.9
Prevalence of diabetes				
Full sample:				
Original weight	6.3	7.4	8.4	9.1
Linkage-eligible sample:				
Original weight	6.6	8.1	8.6	9.3
Adjusted weight 1	6.8	8.3	8.7	9.4
Adjusted weight 2	6.3	7.4	8.4	9.1
Bias:				
Original weight	4.0	8.7	3.2	2.1
Adjusted weight 1	6.4	12.1	4.1	2.9
Adjusted weight 2	0.0	0.0	0.0	0.0
Prevalence of hypertension				
Full sample:				
Original weight	23.9	26.1	28.5	29.9
Linkage-eligible sample:				
Original weight	24.4	27.4	29.3	30.2
Adjusted weight 1	25.1	28.1	29.4	30.4
Adjusted weight 2	24.9	27.9	29.3	30.3
Bias:				
Original weight	2.1	5.0	2.7	1.2
Adjusted weight 1	4.7	8.0	3.1	1.7
Adjusted weight 2	4.2	7.0	2.8	1.4

See footnotes at end of table.

Table 2. Survey estimates and linkage eligibility bias for National Health Interview Survey sample adults, by survey year, sample weight, and demographic and health-related variables, 2000–2013—Con.

Selected variables	2000–2003	2004–2006	2007–2009	2010–2013
Prevalence of obesity		Percent		
Full sample:				
Original weight	22.7	24.9	27.1	28.3
Linkage-eligible sample:				
Original weight	23.8	27.3	28.1	28.8
Adjusted weight 1	24.0	27.6	28.1	28.8
Adjusted weight 2	23.9	27.3	28.0	28.7
Bias:				
Original weight	5.1	9.4	3.6	1.7
Adjusted weight 1	6.0	10.8	3.6	1.8
Adjusted weight 2	5.5	9.8	3.2	1.5
Prevalence of fair or poor self-rated health				
Full sample:				
Original weight	12.0	12.3	13.1	13.1
Linkage-eligible sample:				
Original weight	12.4	13.7	13.4	13.3
Adjusted weight 1	12.6	14.0	13.6	13.4
Adjusted weight 2	12.5	13.8	13.5	13.3
Bias:				
Original weight	3.4	11.6	2.2	1.4
Adjusted weight 1	5.6	14.0	3.3	1.9
Adjusted weight 2	4.6	12.3	2.7	1.4
Prevalence of doctor's office visit in past year				
Full sample:				
Original weight	80.9	80.4	80.9	80.5
Linkage-eligible sample:				
Original weight	80.7	80.2	81.6	80.9
Adjusted weight 1	81.1	80.3	81.4	80.8
Adjusted weight 2	81.0	80.2	81.4	80.8
Bias:				
Original weight	–0.3	–0.2	0.9	0.4
Adjusted weight 1	0.2	0.0	0.7	0.4
Adjusted weight 2	0.2	–0.1	0.6	0.4
Prevalence of smoking				
Full sample:				
Original weight	44.7	42.2	42.1	40.5
Linkage-eligible sample:				
Original weight	45.8	45.7	43.5	41.0
Adjusted weight 1	45.9	45.9	43.3	41.0
Adjusted weight 2	45.9	45.9	43.3	41.0
Bias:				
Original weight	2.4	8.1	3.5	1.4
Adjusted weight 1	2.6	8.7	2.9	1.3
Adjusted weight 2	2.6	8.7	2.8	1.2

0.0 Quantity more than zero but less than 0.05.
--- Data not available.

NOTES: Eligibility for linkage was based upon consenting and having sufficient personally identifiable information. For 2000–2006, refusal for data linkage was defined as refusing to provide a Social Security Number (SSN) or Medicare health insurance claim number (HICN). For 2007–2013, refusal for data linkage was defined as not providing the last four digits of the SSN or HICN and responding "No" to the follow-up question to allow linkage without these personal identifiers. Due to sample design changes in 2006, estimates for non-Hispanic Asian persons for 2004–2006 were only from the 2006 National Health Interview Survey.

SOURCE: National Center for Health Statistics, linked National Health Interview Survey–Centers for Medicare & Medicaid Services Feasibility Files.

Table 3. Standard error of survey estimates for National Health Interview Survey sample adults, by survey year, linkage eligibility, sample weight, and demographic and health-related variables, 2000–2013

Selected variables	2000–2003	2004–2006	2007–2009	2010–2013
Ages 18–44				
Full sample:				
Original weight	0.25	0.27	0.36	0.27
Linkage-eligible sample:				
Original weight	0.28	0.36	0.40	0.28
Adjusted weight 1	0.28	0.36	0.40	0.28
Adjusted weight 2	0.28	0.36	0.40	0.28
Ages 45–64				
Full sample:				
Original weight	0.19	0.23	0.28	0.21
Linkage-eligible sample:				
Original weight	0.23	0.31	0.31	0.22
Adjusted weight 1	0.23	0.32	0.31	0.22
Adjusted weight 2	0.23	0.32	0.31	0.22
Ages 65 and over				
Full sample:				
Original weight	0.19	0.18	0.25	0.19
Linkage-eligible sample:				
Original weight	0.20	0.24	0.26	0.19
Adjusted weight 1	0.21	0.24	0.27	0.19
Adjusted weight 2	0.21	0.24	0.27	0.19
Women				
Full sample:				
Original weight	0.17	0.20	0.25	0.19
Linkage-eligible sample:				
Original weight	0.21	0.31	0.28	0.20
Adjusted weight 1	0.21	0.31	0.28	0.20
Adjusted weight 2	0.21	0.31	0.28	0.20
Men				
Full sample:				
Original weight	0.17	0.20	0.25	0.19
Linkage-eligible sample:				
Original weight	0.21	0.31	0.28	0.20
Adjusted weight 1	0.21	0.31	0.28	0.20
Adjusted weight 2	0.21	0.31	0.28	0.20
Hispanic				
Full sample:				
Original weight	0.22	0.22	0.29	0.25
Linkage-eligible sample:				
Original weight	0.26	0.32	0.28	0.26
Adjusted weight 1	0.25	0.31	0.31	0.27
Adjusted weight 2	0.25	0.31	0.31	0.27
Non-Hispanic Asian				
Full sample:				
Original weight	---	0.06	0.12	0.12
Linkage-eligible sample:				
Original weight	---	0.07	0.12	0.12
Adjusted weight 1	---	0.11	0.13	0.13
Adjusted weight 2	---	0.11	0.13	0.13
Non-Hispanic black				
Full sample:				
Original weight	0.25	0.22	0.28	0.23
Linkage-eligible sample:				
Original weight	0.28	0.27	0.32	0.24
Adjusted weight 1	0.29	0.28	0.31	0.24
Adjusted weight 2	0.29	0.28	0.31	0.24

See footnote at end of table.

Table 3. Standard error of survey estimates for National Health Interview Survey sample adults, by survey year, linkage eligibility, sample weight, and demographic and health-related variables, 2000–2013—Con.

Selected variables	2000–2003	2004–2006	2007–2009	2010–2013
Non-Hispanic white or other				
Full sample:				
Original weight	0.32	0.30	0.41	0.35
Linkage-eligible sample:				
Original weight	0.37	0.40	0.43	0.36
Adjusted weight 1	0.36	0.40	0.44	0.36
Adjusted weight 2	0.36	0.40	0.44	0.36
No high school				
Full sample:				
Original weight	0.21	0.22	0.25	0.21
Linkage-eligible sample:				
Original weight	0.26	0.33	0.26	0.21
Adjusted weight 1	0.26	0.32	0.26	0.21
Adjusted weight 2	0.26	0.32	0.26	0.21
High school or GED				
Full sample:				
Original weight	0.22	0.23	0.27	0.22
Linkage-eligible sample:				
Original weight	0.26	0.33	0.29	0.23
Adjusted weight 1	0.26	0.33	0.29	0.23
Adjusted weight 2	0.26	0.33	0.29	0.23
Some college				
Full sample:				
Original weight	0.18	0.22	0.26	0.21
Linkage-eligible sample:				
Original weight	0.23	0.33	0.29	0.23
Adjusted weight 1	0.22	0.33	0.28	0.22
Adjusted weight 2	0.22	0.33	0.28	0.22
Bachelor's degree or higher				
Full sample:				
Original weight	0.27	0.24	0.36	0.30
Linkage-eligible sample:				
Original weight	0.29	0.34	0.39	0.31
Adjusted weight 1	0.29	0.34	0.40	0.31
Adjusted weight 2	0.29	0.34	0.40	0.31
Diabetes				
Full sample:				
Original weight	0.09	0.11	0.14	0.10
Linkage-eligible sample:				
Original weight	0.10	0.16	0.16	0.11
Adjusted weight 1	0.10	0.17	0.16	0.11
Adjusted weight 2	0.10	0.15	0.15	0.11
Hypertension				
Full sample:				
Original weight	0.17	0.20	0.26	0.21
Linkage-eligible sample:				
Original weight	0.20	0.28	0.29	0.22
Adjusted weight 1	0.20	0.29	0.29	0.22
Adjusted weight 2	0.20	0.29	0.29	0.22
Obesity				
Full sample:				
Original weight	0.17	0.19	0.25	0.21
Linkage-eligible sample:				
Original weight	0.22	0.30	0.27	0.21
Adjusted weight 1	0.22	0.31	0.26	0.21
Adjusted weight 2	0.22	0.30	0.26	0.21

See footnote at end of table.

Table 3. Standard error of survey estimates for National Health Interview Survey sample adults, by survey year, linkage eligibility, sample weight, and demographic and health-related variables, 2000–2013—Con.

Selected variables	2000–2003	2004–2006	2007–2009	2010–2013
Fair or poor self-rated health				
Full sample:				
Original weight	0.14	0.15	0.20	0.14
Linkage-eligible sample:				
Original weight	0.18	0.22	0.21	0.15
Adjusted weight 1	0.18	0.23	0.21	0.15
Adjusted weight 2	0.18	0.23	0.21	0.15
Doctor's office visit in past year				
Full sample:				
Original weight	0.17	0.21	0.23	0.18
Linkage-eligible sample:				
Original weight	0.21	0.27	0.24	0.19
Adjusted weight 1	0.21	0.27	0.24	0.19
Adjusted weight 2	0.21	0.27	0.25	0.19
Smoking				
Full sample:				
Original weight	0.22	0.25	0.30	0.23
Linkage-eligible sample:				
Original weight	0.27	0.35	0.33	0.24
Adjusted weight 1	0.26	0.35	0.33	0.24
Adjusted weight 2	0.26	0.35	0.33	0.24

--- Data not available.

NOTES: Eligibility for linkage was based upon consenting and having sufficient personally identifiable information. For 2000–2006, refusal for data linkage was defined as refusing to provide a Social Security Number (SSN) or Medicare health insurance claim number (HICN). For 2007–2013, refusal for data linkage was defined as not providing the last four digits of the SSN or HICN and responding "No" to the follow-up question to allow linkage without these personal identifiers. Due to sample design changes in 2006, estimates for non-Hispanic Asian persons for 2004–2006 were only from the 2006 National Health Interview Survey.

SOURCE: National Center for Health Statistics, linked National Health Interview Survey–Centers for Medicare & Medicaid Services Feasibility Files.

Table 4. Summary statistics for sample weight for National Health Interview Survey sample adults, by survey year, 2000–2013

Weighting procedure	Number of respondents	Minimum weight	Maximum weight	Mean weight	Median weight	Estimated design effect due to differential weighting
2000–2003						
Original weight	127,596	667	85,548	6,461	5,639	1.4
Adjusted weight 1	79,810	932	125,419	10,329	9,255	1.4
Adjusted weight 2	79,810	932	125,561	10,329	9,240	1.4
2004–2006						
Original weight	87,029	728	127,899	7,506	6,514	1.5
Adjusted weight 1	39,982	1,442	207,889	16,338	13,999	1.5
Adjusted weight 2	39,982	1,320	195,150	16,338	13,993	1.5
2007–2009						
Original weight	72,905	900	121,989	9,269	7,498	1.6
Adjusted weight 1	57,375	1,085	161,508	11,778	9,501	1.6
Adjusted weight 2	57,375	1,085	159,680	11,778	9,495	1.6
2010–2013						
Original weight	129,253	168	93,300	7,220	5,701	1.6
Adjusted weight 1	114,074	194	104,481	8,181	6,478	1.6
Adjusted weight 2	114,074	197	104,500	8,181	6,479	1.6

SOURCE: National Center for Health Statistics, linked National Health Interview Survey–Centers for Medicare & Medicaid Services Feasibility Files.

Appendix. Survey Questions and SAS Codes

Wording of Survey Questions

Diabetes: “Have you EVER been told by a doctor or health professional that you have diabetes or sugar diabetes?”

Education level: “What is the HIGHEST level of school {person has} completed or the highest degree {person has} received? Please tell me the number from the card.”

Fair or poor health: “Would you say {your/ALIAS's} health in general is excellent, very good, good, fair, or poor?”

Hypertension: “Have you EVER been told by a doctor or other health professional that you had hypertension, also called high blood pressure?”

Obesity: “How tall are you without shoes? How much do you weigh without shoes?”

Body mass index (BMI): Calculated using the in-house version of the height and weight variables, which contain a greater range of height and weight values than are available on the public-use file. The range of possible BMI values listed (0001–9994) does not reflect actual calculated BMI values at the extremes of the range, but rather, allows for the theoretical possibility of such unlikely values. BMI was calculated as weight in kilograms (kg) divided by height in meters (m) squared, rounded 2 decimal places. Conversion factors: 1 kg = 2.20462 pounds; 1 m = 39.37008 inches. For both men and women, underweight is defined as BMI less than 18.5; healthy weight is BMI 18.5 to less than 25.0; overweight is BMI 25.0 to less than 30.0; obese is BMI 30.0 or greater.

Office visit: “DURING THE PAST 12 MONTHS, HOW MANY TIMES have you seen a doctor or other health care professional about your own health at a DOCTOR'S OFFICE, A CLINIC, OR SOME OTHER PLACE?”

Smoking: “Have you smoked at least 100 cigarettes in your ENTIRE LIFE?”

SAS Code for Creating Linkage Eligibility (LE) Flags

```
if CMS_MEDICARE_MATCH in (1,2) then LE = 1;
else if CMS_MEDICARE_MATCH = 9 then LE = 0;
else LE = 99;
```

For more information, see “NCHS–CMS Medicare Feasibility Files,” available from: https://www.cdc.gov/nchs/data/datalinkage/cms_medicare_feasibility_data_codebook.pdf.

SAS Code for Dichotomization of Health Variables

```
if DIBEV = 1 then DIBEV2 = 'Yes';
/*Note - "borderline" treated as "No"*/
else if DIBEV in (2,3) then DIBEV2 = 'No ';
else if DIBEV in (7,8,9) then DIBEV2 = '    ';
else DIBEV2 = '99 ';
if HYPEV  = 1 then HYPEV2 = 'Yes';
else if HYPEV in (2) then HYPEV2 = 'No ';
else if HYPEV in (7,8,9) then HYPEV2 = '    ';
else HYPEV2 = '99 ';
if SMKEV = 1 then SMKEV2 = 'Yes';
else if SMKEV in (2) then SMKEV2 = 'No ';
else if SMKEV in (7,8,9) then SMKEV2 = '    ';
else SMKEV2 = '99 ';
if PHSTAT in (4,5) then PHSTAT2 = 'Yes';
else if PHSTAT in (1,2,3) then PHSTAT2 = 'No ';
else if PHSTAT in (7,8,9) then PHSTAT2 = '    ';
else PHSTAT2 = '99 ';
if AHCNOYR2 in (1,2,3,4,5,6,7,8) then AHCNOYR22 = 'Yes';
else if AHCNOYR2 in (0) then AHCNOYR22 = 'No ';
else if AHCNOYR2 in (97,98,99) then AHCNOYR22 = ' ';
ELSE AHCNOYR22 = '99 ';
if BMI in (99.99 9999 100) then BMI = .;
if bmi   =   .  or  PREGNOW = 1  then bmi_cat  =  .;
else if bmi  < 18.5  then bmi_cat  =  1;
else if bmi ge 18.5 and bmi < 25 then bmi_cat  =  2;
else if bmi ge 25  and bmi < 30  then bmi_cat  =  3;
else if bmi ge 30  then bmi_cat  =  4;
```

```
if bmi_cat = 4  then obesity = 'Yes';
else if bmi_cat in (1,2,3) then obesity = 'No ';
```

SAS Code for Calculating Linkage Eligibility, Adjusted and Weighted

```
/*Create formats and age groupings*/
proc format;
      value raceeth4f
          1 = 'Non-Hisp Other'
          2 = 'Non-Hisp Black'
          3 = 'Hispanic'
          4 = 'Non-Hisp Asian';
      VALUE SEXF
        1 = "MALE"
        2 = "FEMALE"    ;
  VALUE XAGEF
          1 = "18–24 YEARS"
          2 = "25–34 YEARS"
          3 = "35–44 YEARS"
          4 = "45–54 YEARS"
          5 = "55–64 YEARS"
          6 = "65 YEARS and OVER"
          7 = "18–24 YEARS"
          8 = "25–44 YEARS"
          9 = "45–64 YEARS"
          10 = "65 YEARS & OVER"
          11 = "18–19 YEARS"
          12 = "20–24 YEARS"
          13 = "25–29 YEARS"
          14 = "30–34 YEARS"
          15 = "35–44 YEARS"
          16 = "45–49 YEARS"
          17 = "50–54 YEARS"
          18 = "55–64 YEARS"
          19 = "65–74 YEARS"
          20 = "75 YEARS & OVER";
      INVALUE SAAGHISF
           18–24 = 1
           25–34 = 2
           35–44 = 3
           45–54 = 4
           55–64 = 5
           65–HIGH = 6;
      INVALUE SAAGASIF
           18–24    = 7
           25–44    = 8
           45–64    = 9
           65–HIGH = 10;
      INVALUE SAAGGF
           18–19 = 11
           20–24 = 12
           25–29 = 13
           30–34 = 14
           35–44 = 15
           45–49 = 16
           50–54 = 17
           55–64 = 18
           65–74 = 19
           75–HIGH = 20;
        run;
/*Create variables needed for weight adjustment*/
data SPF;
merge samadult (in = a) personsx (in = b) MCARE_FEAS_NHIS (in = c);
```

Note: samadult and personsx are the names of data sets.

```
by PUBLICID;
if a and b and c;
if CMS_MEDICARE_MATCH in (1,2) then LE = 1;
else if CMS_MEDICARE_MATCH = 9 then LE = 0;
else LE = 99;
if ORIGIN_I = 1 and then xage = put(age_p, SAAGHISF.);
else if ORIGIN_I = 2 and RACRECI3 = 3 then xage = put(age_p, SAAGASIF.);
else if ORIGIN_I = 2 and RACRECI3 = 2 then xage = put(age_p, SAAGHISF.);
else if ORIGIN_I = 2 and RACRECI3 in (1,4) then xage = put(age_p, SAAGGF.);
else xage = 999;
race = racreci3;
if race> =  4 or race = 1 then race = 1; else if race = 2 then race = 2;
else if race = 3 then race = 5;
if ORIGIN_I = 1 then raceeth4 = 3;
else if ORIGIN_I = 2 and race = 1 then raceeth4 = 1;
else if ORIGIN_I = 2 and race = 2 then raceeth4 = 2;
```

```
else if ORIGIN_I = 2 and race = 5 then raceeth4 = 4;
else raceeth4 = 99;
run;
/*Adjust weights*/
proc wtadjust data = SP design = wr adjust = nonresponse;
nest STRAT_P PSU_P;
weight wtfa_SA;
class sex xage raceeth4;
model LE = sex*xage*raceeth4;
idvar LE sex xage raceeth4 PUBLICID1;
output / predicted = all filename = SP2 filetype = sas replace;
run;
```

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. 186
CS322081