SAFER • HEALTHIER • PEOPLE™

Monitoring the Nation's Health

National Health and Nutrition Examination Survey: National Youth Fitness Survey Estimation Procedures, 2012

Vital and Health Statistics
Series 2, Number 168
November 2014

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

Suggested citation

Johnson CL, Dohrmann SM, Van de Kerckhove W, et al. National Health and Nutrition Examination Survey: National Youth Fitness Survey estimation procedures, 2012. National Center for Health Statistics. Vital Health Stat 2(168). 2014.

Library of Congress Cataloging-in-Publication Data

National Health and Nutrition Examination Survey. National Youth Fitness Survey estimation procedures, 2012.
p. ; cm.— (Vital and health statistics. series 2, data evaluation and methods research ; number 168) (DHHS publication ; no. (PHS) 2015–1368)
National Youth Fitness Survey estimation procedures, 2012
Includes bibliographical references and index.
ISBN 978–0-8406–0669-3 (alk. paper)—ISBN 0–8406-0669–9 (alk. paper)
I. National Center for Health Statistics (U.S.), issuing body. II. Title: National Youth Fitness Survey estimation procedures, 2012. III. Series: Vital and health statistics. Series 2, Data evaluation and methods research ; no. 168. IV. Series: DHHS publication ; no. (PHS) 2015–1368. 0276–4733
[DNLM: 1. National Youth Fitness Survey (U.S.) 2. Health Surveys—United States. 3. Physical Fitness—United States. 4. Adolescent—United States. 5. Child—United States. 6. Motor Activity—United States. 7. Statistics as Topic—United States. W2 A N148vb no.168 2015]
RA407.3
614.4'273—dc23 2014036998

For sale by the U.S. Government Printing Office
Superintendent of Documents
Mail Stop: SSOP
Washington, DC 20402–9328
Printed on acid-free paper.

Vital and Health Statistics

Series 2, Number 168

National Health and Nutrition Examination Survey: National Youth Fitness Survey Estimation Procedures, 2012

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
November 2014
DHHS Publication No. 2015–1368

Contents

Acknowledgments iv

Abstract 1

Introduction 1

Sample Design Summary 2
- Sample Selection Stages 2
- Sampling Rates 3

Weighting Sample Data 3
- Calculating Base Weights 4
- Nonresponse Adjustment 5
- Trimming 6
- Poststratification 6
- Computing Final Weights 7
- Combined NHANES/NNYFS Weights 7

Variance Estimation 9
- Variance Estimation for Publicly Released NNYFS Data 10
- Variance Estimation for Combined NHANES/NNYFS Data in RDC 10

Cautionary Notes 10

References 11

Appendix I. Glossary 12

Appendix II. Supporting Tables 15

Appendix Tables

I. Target sample sizes, screening amounts, and sampling rates: NNYFS, 2012 15

II. Variables used to form nonresponse adjustment cells for weighting interview and examination samples: NNYFS, 2012 16

III. Design effects, effective sample sizes, and final compositing factors used to combine annual samples: National Health and Nutrition Examination Survey and NHANES National Youth Fitness Survey, 2011–2012 17

Acknowledgments

The authors gratefully acknowledge the assistance of Ryne Paulose and Michele Chiappa in the preparation of this report.

Abstract

Background

The National Health and Nutrition Examination Survey's (NHANES) National Youth Fitness Survey (NNYFS) was conducted in 2012 by the Centers for Disease Control and Prevention's National Center for Health Statistics (NCHS). NNYFS collected data on physical activity and fitness levels to evaluate the health and fitness of children aged 3–15 in the United States. The survey comprised three levels of data collection: a household screening interview (or screener), an in-home personal interview, and a physical examination. The screener's primary objective was to determine whether any children in the household were eligible for the interview and examination. Eligibility was determined by preset selection probabilities for desired sex-age subdomains. After selection, the in-home personal interview collected demographic, health, physical activity, and nutrition information about the child as well as information about the household. The examination included physical measurements and fitness tests.

Objectives

This report provides background on the NNYFS program and summarizes the survey's sample design specifications. The report presents NNYFS estimation procedures, including the methods used to calculate survey weights for the full sample as well as a combined NHANES/NNYFS sample for 2012 (accessible only through the NCHS Research Data Center). The report also describes appropriate variance estimation methods. Documentation of the sample selection methods, survey content, data collection procedures, and methods to assess nonsampling errors are reported elsewhere (available from: http://www.cdc.gov/nchs/nhanes.htm).

Keywords: sampling • weighting • variance estimation

National Health and Nutrition Examination Survey: National Youth Fitness Survey Estimation Procedures, 2012

by Clifford L. Johnson, M.S.P.H., National Center for Health Statistics; Sylvia M. Dohrmann, M.S., Wendy Van de Kerckhove, M.S., Mamadou S. Diallo, M.Sc., Jason Clark, M.S., and Leyla K. Mohadjer, Ph.D., Westat; and Vicki L. Burt, Sc.M., R.N., National Center for Health Statistics

Introduction

The National Health and Nutrition Examination Survey (NHANES) is one of a series of health-related programs conducted by the Centers for Disease Control and Prevention's (CDC) National Center for Health Statistics (NCHS). It provides information on the health and nutritional status of the U.S. population. In 2012, NCHS also conducted the NHANES National Youth Fitness Survey (NNYFS).

NNYFS was designed to collect data on physical activity and fitness levels for U.S. children aged 3–15 (1). NNYFS was linked to NHANES by using the same primary sampling units and the same operational procedures as NHANES. NNYFS data were collected using an in-home personal interview; fitness tests and a dietary recall were done in a separate, single-trailer mobile examination center (MEC). In addition, results from questions and examination components common to both NNYFS and NHANES 2011–2012 were collected using the same protocol so that they could be combined for certain analyses.

In October 2008, the federal government issued the first Physical Activity Guidelines for Americans to provide science-based guidance on the types and amounts of physical activity that provide substantial health benefits (2). Guidelines for children and teenagers recommend 60 minutes or more of aerobic, muscle-strengthening, or other physical activity daily. The inclusion of standardized tests of core strength, upper and lower body strength, cardiovascular capacity, and gross motor skills in NNYFS for ages 3–15 provides additional information with which to evaluate the health and fitness of this age group.

NNYFS provides information on the noninstitutionalized resident population of children aged 3–15 in the United States. NNYFS excludes all children in supervised care or custody in institutional settings, children of active-duty military personnel living overseas, and any other U.S. citizens residing outside of the 50 states and the District of Columbia.

NNYFS included three levels of data collection: a household screening interview (or screener), an in-home personal interview, and a standardized physical examination conducted in the MEC that included selected objective measures of fitness. The primary objective of the screener was to determine whether any children in the household were eligible for the in-home personal interview and the MEC examination. The personal interview collected demographic, health, physical activity, and nutrition information about the child as well as information about the household. The examination included physical measurements, a dietary interview, and fitness tests;

conducting the examinations in the MEC helped to standardize their administration.

NNYFS was conducted in parallel with the 2012 data collection of the 2011–2012 survey cycle of NHANES. Some questionnaire items and physical measurements were collected in the same manner in both surveys for 2012. In some cases, this required a modification of NHANES 2011 questionnaire items or physical measures to create comparability for both surveys in 2012. Thus, it is possible to combine samples for the two surveys for 2012. This is especially useful if the NNYFS sample sizes are small for the specific analysis of interest. Because single years of NHANES data are not released as public-use data files, the combined 2012 NHANES and NNYFS file is available for analysis only in the NCHS Research Data Center (RDC) (visit http://www.cdc.gov/rdc/ for more details). Specifically, sample weights were created for use with this combined NHANES and NNYFS sample of children aged 3–15.

Less survey content overlap occurs between NNYFS and 2011 NHANES because the first year of the NHANES 2011–2012 survey cycle was fielded before changes were implemented to make the 2012 NHANES and the 2012 NNYFS more comparable.

A combined 2011–2012 NHANES and NNYFS data file for children aged 3–15 is also available for use in RDC, but the relatively small number of questionnaire items and physical measures comparably collected in NHANES 2011, NHANES 2012, and NNYFS limits the analytic usefulness of these data despite the file's increased sample size, number of primary sampling units (PSUs), and stability of estimates.

The "Sample Design Summary" section briefly summarizes the sample design specifications for NNYFS, with the remainder of the report providing the estimation procedures. The "Weighting Sample Data" section describes the creation of weights for the entire NHANES sample and the combined NHANES and NNYFS sample; the "Variance Estimation" section presents the appropriate variance estimation methods; and the "Cautionary Notes" section presents some analytic limitations of the combined NNYFS and NHANES data.

General information related to the planning, sample design, and field operations of NNYFS, as well as general analytical guidance, can be found in "National Health and Nutrition Examination Survey: National Youth Fitness Survey Plan, Operations, and Analysis, 2012" (1). Documentation of the survey content, data collection procedures, and methods to assess nonsampling errors are reported elsewhere (available from: http://www.cdc.gov/nchs/nhanes.htm). For more details on the 2011–2014 NHANES sample design, see "National Health and Nutrition Examination Survey: Sample Design, 2011–2014" (3).

Sample Design Summary

The NNYFS sample represents the total noninstitutionalized resident population of children aged 3–15 in the United States. The survey was conducted in the same primary and secondary sampling units as NHANES during 2012. The four-stage sample design used in NHANES was also used in NNYFS. A description of the stages of selection and the calculation of sampling rates follows.

Sample Selection Stages

The first stage of the NHANES 2011–2014 sample design consisted of selecting the PSUs from a sampling frame of all U.S. counties. The PSUs in the first stage were mostly counties; in a few cases, adjacent counties were combined to keep PSUs above a certain minimum size. NHANES PSUs were selected with probabilities proportionate to a measure of size (PPS). After selection, 15 PSUs were allocated to each of the 4 years of the study period randomly, so that each year contained a nationally representative sample (an NHANES design requirement). The NHANES PSUs allocated to 2012 were also the NNYFS PSUs.

The second stage of selection for the 2012 NHANES and NNYFS samples included a single sample of area segments, comprising census blocks or combinations of blocks. The sample was designed to produce approximately equal sample sizes per PSU for both NHANES and NNYFS. PSUs selected with certainty (with a probability of one) may have more or fewer than 24 segments to ensure appropriate representation in the sample. Noncertainty PSUs have 24 segments. The NNYFS segments were formed and selected with PPS using the same methods as in NHANES, with one exception: The segments were enlarged to ensure they would provide enough samples of dwelling units (DUs) for both surveys.

The third stage of sample selection consisted of DUs, including noninstitutional group quarters (i.e., group quarters that do not provide formally authorized, supervised care or custody in institutional settings). These include college residence halls, group homes intended for adults, residential treatment facilities for adults, workers' group living quarters and Job Corps centers, and religious group quarters. In a given PSU, following the selection of segments, a listing of all DUs in the sampled segments was prepared. A subsample of these was selected, and a random subset of these was designated for screening to identify potential sampled participants for NNYFS. All other DUs were designated for screening to identify potential sampled participants for NHANES.

The fourth stage of sample selection consisted of persons within occupied DUs, or households. All eligible children within a household were listed, and a subsample was selected based on domains defined by sex and age. The sampling procedures for NNYFS were generally the same as those used in NHANES.

Sampling Rates

NNYFS was conducted in conjunction with NHANES, using some of the same screening and interviewing staff. The overall target number of examinations for the study was 1,500, with approximately equal sample sizes within each sex and single year of age combination. To meet this end, sampling rates were developed for six sex-age subdomains: males aged 3–5, 6–11, and 12–15, and females aged 3–5, 6–11, and 12–15. The subsampling rates and designation of potential sampled participants within screened households were arranged to provide approximately self-weighting samples for each subdomain while simultaneously maximizing the average number of sampled participants per household.

The original target sample distribution for NNYFS is shown in Table I. To calculate sampling rates, an overall 80% response rate was assumed. This table also gives the projected amount of screening required to a) obtain one examined person in each domain and b) attain the target number of examined children in each domain.

The amount of screening needed for NNYFS was originally 4,145 households, corresponding to the number of occupied housing units needed to identify the targeted number of females aged 12–15. All screened children in this subdomain were retained in the sample. Screened children in the other subdomains were subsampled to bring the sampling rates for those subdomains down to desired levels. The extent of this subsampling is shown in the last column of Table I.

A derivation of the NNYFS maximum sampling rate (sampling rate for the screening sample) is:

1. Screening sample size for 1 year = 4,145 households
2. A 50% reserve (additional 2,073 households) sample size for 1 year = 6,218 households
3. Projected total households in the United States during 2011–2014 = 118,410,614 households
4. Therefore,
   Maximum sampling rate = 6,218 /118,410,614 ≈ 1/19,044

The PSUs and segments for NNYFS were selected as efficient sampling units for NHANES, which, unlike NNYFS, oversamples persons by race and Hispanic origin. As a result, in some areas more DUs were needed than expected; although they contained a sufficient number of minorities to meet the NHANES sample targets, they did not contain a sufficient number of children. Consequently, NNYFS selection probabilities changed twice during 2012 to increase the sample yield so that the overall annual target of 1,500 examinations could be reached.

The change was implemented after DUs were selected for the first two study locations. The selected 50% reserve did not appear sufficient for future study locations. The rates were subsequently changed to include a 100% reserve sample, changing the maximum sampling rate to 8,291/118,410,614 ≈ 1/14,282.

Children in different sampling domains were originally given different sampling rates so that approximately the same number of examinations would be obtained for each single year of age. For the final six survey locations, in an attempt to improve efficiency and obtain more examinations, the rates for all children were set to the maximum so that all 3- to 15-year-olds living in selected DUs would be selected for the study.

Weighting Sample Data

The goal of NNYFS was to produce data representative of the noninstitutionalized U.S. population of children aged 3–15. The weighting of sample data permits analysts to produce estimates of statistics they would have obtained if the entire sampling frame had been surveyed. Sample weights can be considered as measures of the number of persons represented by the particular sampled participant. Weighting takes into account several survey features: the differential probabilities of selection for the sampling domains, nonresponse to survey instruments, and differences between the final sample and the total population.

NNYFS samples were weighted to the following objectives:

1. Compensate for differential probabilities of selection among subgroups defined by sex and age.
2. Reduce biases arising from the fact that nonrespondents may differ from respondents.
3. Fix weighted sample data to match an independent estimate from the U.S. Census Bureau of the target population totals.
4. Compensate, to the extent possible, for inadequacies in the sampling frame (resulting from omissions of some housing units in the listing of area segments, omissions of persons with no fixed address, and others).
5. Reduce variances in the estimation procedure by using auxiliary information that is known with a high degree of accuracy.

The sample weighting was carried out in three steps. The first step involved the computation of weights to compensate for unequal probabilities of selection (objective 1). The second step adjusted for nonresponse (objective 2). In the third step, the sample weights were poststratified to census estimates of the U.S. population to simultaneously accomplish objectives 3–5.

These steps were performed for respondents at each stage of the survey: the screener, personal interview, and examination. The weights described in ‘‘Calculating Base Weights’’ were the starting point for the screener weight calculation. Those weights were then adjusted for nonresponse to the screener and then poststratified. The resulting weights became the starting point for calculating the interview weights, which were then adjusted for nonresponse to the interview, inspected for extreme weights, and again poststratified. Finally, those poststratified interview weights were the starting point for calculating the examination weights, which were adjusted for nonresponse to the

examination, inspected for extreme weights, and then poststratified.

Note that extreme variability in the weights results in reduced reliability (increased sampling error) of some survey estimates. The NNYFS sample was designed to minimize variability in the weights, subject to operational and analytic constraints. Additionally, measures such as weight trimming were implemented to reduce variability in the NNYFS weights. The impact of weight variability is minimal when estimates are applied to demographic subdomains used in the design; however, when estimates are instead designed for domains aggregated across design domains (for example, an estimate for the total population), then the impact of weight variability is greater.

Calculating Base Weights

The overall selection probability for a person selected in sampling domain k for NHANES is

$$P_h \cdot P_{hj} \cdot \frac{\max\{r_k^N\}}{P_h \cdot P_{hj}} \cdot \frac{r_k^N}{\max\{r_k^N\}} = r_k^N$$

where

P_h = Pr{select PSU h}

P_{hj} = Pr{select segment hj | select PSU h}

$\frac{\max\{r_k^N\}}{P_h \cdot P_{hj}}$ = Pr{select DU in segment hj | select segment hj}

$\frac{r_k^N}{\max\{r_k^N\}}$ = Pr{domain (k) flagged for selection in DU | DU in segment hj selected}

r_k^N = Sampling rate for NHANES participant in sampling domain k

$\max\{r_k^N\}$ = Maximum sampling rate for NHANES across all sampling domains.

With the addition of NNYFS in the PSUs and segments selected for NHANES, the NHANES DU selection probability was adjusted to account for the selection of a sample large enough for both surveys as well as the additional stage of selection for subsampling the larger DU sample in the specific NHANES and NNYFS samples. As a result, the expression for the overall NHANES selection probability was changed to

$$P_h \cdot P_{hj} \cdot \left[\frac{\max\{r_k^N\}}{P_h \cdot P_{hj}} \cdot \text{RsrvAdj}\right] \cdot \frac{1}{\text{RsrvAdj}} \cdot \frac{r_k^N}{\max\{r_k^N\} \cdot \text{RsrvAdj} \cdot \frac{1}{\text{RsrvAdj}}} = r_k^N$$

where RsrvAdj is the increase of the DU sample size to meet the needs of both the NHANES and NNYFS surveys compared with the originally planned NHANES DU sample size for 2012, and $\frac{1}{\text{RsrvAdj}}$ is the subsampling rate for the NHANES DU sample.

Consequently, the overall NNYFS selection probability for a child in sex-age group k is

$$P_h \cdot P_{hj} \cdot \left[\frac{\max\{r_k^N\}}{P_h \cdot P_{hj}} \cdot \text{RsrvAdj}\right] \cdot \left[1 - \frac{1}{\text{RsrvAdj}}\right] \cdot \frac{r_k^Y}{\max\{r_k^N\} \cdot \text{RsrvAdj} \cdot \left[1 - \frac{1}{\text{RsrvAdj}}\right]} = r_k^Y$$

where

$\frac{\max\{r_k^N\}}{P_h \cdot P_{hj}} \cdot \text{RsrvAdj}$ = Pr{select DU in segment hj | select segment hj} for inclusion in either the NHANES or NNYFS sample

$\left[1 - \frac{1}{\text{RsrvAdj}}\right]$ = Subsampling rate for NNYFS DU sample

$\frac{r_k^Y}{\max\{r_k^N\} \cdot \text{RsrvAdj} \cdot \left[1 - \frac{1}{\text{RsrvAdj}}\right]} =$ Pr{domain (k) flagged for selection in DU | DU in segment hj selected} for NNYFS

r_k^Y = Sampling rate for NNYFS participant in sampling domain k.

The base weight for a sampled child is simply the reciprocal of the sampling rate for the domain of the sampled participant, r_k^Y. These sampling rates are provided in Table I, and their derivation is described in the preceding section. For NNYFS, the base weight was adjusted further to account for:

- Proportion of DUs released, $f_{i(\text{release})}$
- Within-segment adjustments to the selection probabilities, $f_{i(\text{wsa})}$
- Release of an extra reserve sample, $f_{i(\text{res})}$

The final base weight was calculated as

$$w_{i(\text{base})} = \frac{1}{r_k^Y}\left(f_{i(\text{release})} f_{i(\text{wsa})} f_{i(\text{res})}\right)$$

where i indicates the sampled participant. The following briefly describe each component of this calculation.

Adjustments for number of sampled DUs released to field—The first component, the release factor $f_{i(\text{release})}$ was introduced to reflect the procedures used to obtain a relatively fixed sample size within each study location in NNYFS. The sampled child base weight was adjusted according to the proportion of the total sample released to the field. The release factor was calculated as

$$f_{i(\text{release})} = \frac{1}{D_i}$$

where D_i represents the proportion of sampled DUs released for screening in the location from which sampled participant i was selected. If response rates approached predicted values and the measures of size used during sampling were current, the subsample factor would be approximately 1.5 for the first two study locations and 2.0 for all other locations. That is, approximately two-thirds of the sampled cases were expected to be released for the earlier locations, and approximately one-half of the sampled cases were expected to be released for later study locations (see ‘‘Adjustment for release of extra reserve sample’’ for more details).

Adjustment to increase sample size within segments—Some study locations had relatively small sample sizes after DU selection, and the sample might not have provided enough DUs to reach the target number of identified sample persons. For these study locations, the DU sample size within each segment

was increased, but the DU selection probability was limited to 1.0. Most segments had sample sizes increased by the same percentage, but segments that had most or all of the DUs already sampled had smaller increases.

To accomplish this increase, the combined NHANES and NNYFS DU sample size was increased by an adjustment factor (AdjFac) so that the new DU selection probability was

$$\left[\frac{\max\{r_k^N\}}{P_h \cdot P_{hj}} \cdot \text{RsrvAdj}\right] \cdot \text{AdjFac}$$

To ensure that all of the increased sample size was used for the NNYFS sample, the new NNYFS DU subsampling rate was calculated as

$$\left[\frac{\max\{r_k^N\}}{P_h \cdot P_{hj}} \cdot \text{RsrvAdj}\right] \cdot \text{AdjFac} \cdot \left[1 - \frac{1}{\text{RsrvAdj} \cdot \text{AdjFac}}\right]$$

Solving algebraically, the percent increase in the NNYFS DU sample becomes

$$\frac{\text{RsrvAdj} \cdot \text{AdjFac} - 1}{\text{RsrvAdj} - 1}$$

The factor to be applied to the base weights, $f_{i(\text{wsa})}$, is then calculated as

$$f_{i(\text{wsa})} = \frac{\text{RsrvAdj} - 1}{\text{RsrvAdj} \cdot \text{AdjFac} - 1}$$

Adjustment factors other than 1.0 were used in three study locations.

Adjustment for release of extra reserve sample—Some study locations, even after increasing the sample size within segments, required a larger sample to achieve the target number of identified sample persons. For these study locations, an extra reserve sample was created, but it was not loaded into the system unless the entire original sample was released and more sampled DUs were necessary.

Increasing the sample size in this way—by increasing the sampling rate in each segment—causes different factors to be created for segments where the sampling rate was raised to 100%. Having different factors by segment can add to variability in the weights, so this extra reserve was released only if necessary to achieve the target number of identified sample persons.

The extra reserve sample was created in the same way as the adjustment to increase the sample size within segments—by increasing the combined sample by a certain factor for each segment (RsrvFac). Applying this extra reserve factor as above, the factor to be applied to the base weights, $f_{i(\text{res})}$, is calculated as

$$f_{i(\text{res})} = \frac{\text{RsrvAdj} \cdot \text{AdjFac} - 1}{\text{RsrvAdj} \cdot \text{AdjFac} \cdot \text{RsrvFac} - 1}$$

The extra reserve sample was loaded in only one study location. For all other study locations, this factor is 1.0. Thus, for one study location only, the combined factor for increasing the sample size is

$$f_{i(\text{wsa})} \cdot f_{i(\text{res})} = \frac{\text{RsrvAdj} - 1}{\text{RsrvAdj} \cdot \text{AdjFac} - 1} \cdot \frac{\text{RsrvAdj} \cdot \text{AdjFac} - 1}{\text{RsrvAdj} \cdot \text{AdjFac} \cdot \text{RsrvFac} - 1} = \frac{\text{RsrvAdj} - 1}{\text{RsrvAdj} \cdot \text{AdjFac} \cdot \text{RsrvFac} - 1}$$

The RsrvFac value differs by segment within only one study location.

Nonresponse Adjustment

If every selected household had agreed to complete the screener, and every selected child had agreed to complete the interview and the examination, weighted estimates (using the base weights described in ''Calculating Base Weights'') would be approximately unbiased estimates of characteristics for the noninstitutionalized U.S. population. But in reality, some of the sampled participants who were screened refused to be interviewed (interview nonresponse), and some of the interviewed participants refused the examination (examination nonresponse). Thus, nonresponse bias may result. Bias in the survey estimates occurs when the characteristics of nonrespondents are very different from those of respondents. The best approach to minimizing nonresponse bias is to plan and implement field procedures that maintain high cooperation rates. For NNYFS, the payment of cash incentives and repeated callbacks for refusal conversion are very effective in reducing nonresponse and, thus, nonresponse bias. Yet some nonresponse occurs even with the best strategies; therefore, adjustments are always necessary to minimize potential nonresponse bias.

A multistage procedure for nonresponse adjustment was carried out to adjust for nonresponse to the screener, interview, and examination. The nonresponse adjustment procedure consists of computing adjustment factors and applying these factors to the survey weights separately by nonresponse cell. Nonresponse adjustment reduces bias if response rates and survey characteristics vary from cell to cell, and if respondents and nonrespondents sharing the same characteristics are in the same cell. The nonresponse adjustment factors are the reciprocals of the weighted response rates within the selected cells.

A negative effect of nonresponse adjustment is that it increases the variability of the weights, which in turn increases the variance. When the nonresponse cells contain a sufficient number of cases and the adjustment factors are not too large, the effect on variances is modest. A large adjustment factor in a cell is usually the result of the small number of respondents in that cell. To avoid having nonresponse adjustments based on very small sample sizes, or having large nonresponse adjustment factors, cells are usually collapsed to form larger cells. The following criteria were used in NNYFS to determine whether to collapse cells:

- Minimum of 30 respondents in each cell
- Maximum adjustment factor of 1.35

Nonresponse adjustments were carried out separately for screener nonresponse, interview nonresponse, and examination nonresponse. In general,

nonresponse adjustment cells were generated using variables with known values for both respondents and nonrespondents. A few variables with low item nonresponse rates were considered when creating nonresponse adjustment cells. For the screener nonresponse adjustment, cells were defined by segments within each location. For the interview and examination nonresponse adjustments, a logistic regression model was run to determine which variables were important in predicting response propensity. Once the significant variables were determined, a program was run to classify cases into cells while maximizing the difference in response rates between groups. As with NHANES weights, a classification program based on the University of Michigan algorithm Search was used. This software uses measures based on a chi-squared statistic; see Table II for the variables used to form the nonresponse adjustment cells.

The nonresponse adjustment factors, $f_{i(\mathrm{NR})}$, were calculated as

$$f_{i(\mathrm{NR})} = \frac{\sum_{i=1}^{n_{a_s}} w_{i(\mathrm{base})}}{\sum_{i=1}^{n_{a_r}} w_{i(\mathrm{base})}}$$

where

$w_{i(\mathrm{base})}$ = Base weight for i-th sampled participant in a-th cell

n_{a_s} = Total sample size in a-th nonresponse adjustment cell.

n_{a_r} = Number of respondents in a-th cell

The summation was carried out separately for each cell. Thus, the nonresponse-adjusted weights, $w_{i(\mathrm{NR})}$, were calculated as

$$w_{i(\mathrm{NR})} = w_{i(\mathrm{base})} f_{i(\mathrm{NR})}$$

Trimming

Nonresponse adjustments can contribute to extreme weights; therefore, trimming of the weights was considered. Extreme weights may also occur when units are sampled to yield fixed sample sizes within a PSU, which occurred in NHANES and NNYFS. Even a few unexpectedly large weights can seriously inflate the variance of survey estimates. Thus, weight trimming procedures may be used to reduce the impact of any such large sampled participant weights on the estimates produced from the sample.

Because trimming introduces a bias in the estimates, the resulting reduction in variances is also expected to decrease the mean squared error. The inspection method was used for trimming weights in NNYFS. This method involves inspecting the distribution of weights in the sample, and it applies to samples (or subsets of samples) that were originally designed to be self-weighting.

The subdomains for trimming were the age category (3–5, 6–11, and 12–15) and sex sampling domains. Once the weights to be trimmed had been identified, the weights of the nontrimmed cases were also adjusted so that the weights for each sampling domain summed to the corresponding weighted sum prior to trimming. This is referred to as ‘‘preserving weighted totals,’’ an important feature because failure to preserve weighted totals may lead to serious understatements in estimated totals.

The trimming factors, $f_{i(\mathrm{TR})}$, were calculated as

$$f_{i(\mathrm{TR})} = \frac{\sum_{i=1}^{n_k} t_i}{\sum_{i=1}^{n_k} w_{i(\mathrm{base})} f_{i(\mathrm{NR})}}$$

where n_k is the sample size of the k-th sex-age sampling domain, and t_i is equal to $w_{i\,(\mathrm{base})} f_{i(\mathrm{NR})}$, provided that this product does not exceed the threshold and is set to be equal to the threshold otherwise. The trimmed weights, $w_{i(\mathrm{TR})}$, were calculated as

$$w_{i(\mathrm{TR})} = w_{i(\mathrm{NR})} f_{i(\mathrm{TR})}$$

Poststratification

The final step in the weighting procedure was poststratification to known population totals, to compensate for undercoverage or overcoverage of certain demographic groups, and for any residual differential nonresponse among these groups. Poststratification of sample weights to independent population estimates is used for several purposes. In most household surveys, certain demographic groups in the U.S. population (e.g., children aged 4 and under) experience fairly high rates of undercoverage in survey efforts. Besides partially compensating for such undercoverage and any differential nonresponse, poststratification to census estimates can help reduce resulting bias in the survey estimates, reduce the variability of sample estimates, and achieve consistency with accepted U.S. figures for various subpopulations.

Poststratification involves applying a ratio adjustment to the survey weights. Broad classes, called poststratification cells or poststrata, are constructed using auxiliary data, and a single ratio adjustment factor is applied to all units in a given poststratification cell. The numerator of the ratio is a ‘‘control total’’ obtained from a secondary source; the denominator is a weighted total obtained using the survey weights. Therefore, at the poststratum level, estimates obtained using the poststratified survey weights will correspond to the control totals used. Since poststratification is a ratio adjustment, this process improves the efficiency of estimates, provided that the variables used in constructing poststratification cells are associated with the analysis variables of interest. Such gains in efficiency are most evident in the case of linear estimates such as means or totals; for ratio estimates, the ratio adjustments cancel each other out at the poststratum level, and the overall gains in efficiency due to poststratification tend to be small.

A major effect of poststratification is that it implicitly imputes for nonresponse of survey characteristics for the missed persons. The assumption is that these missed persons not covered by the survey have the same distribution of characteristics as interviewed persons within the poststratification cells. This is obviously an oversimplification; the missed persons are likely to be different. However, in the absence of any detailed information on the characteristics of the missed persons, poststratification appears to be the only reasonable technique available for reducing bias due to undercoverage and nonresponse.

The control totals were obtained using weights from the 2011 American

Community Survey (ACS). These ACS weights have undergone poststratification to the latest census estimates of the total U.S. noninstitutionalized civilian population, including those not counted in surveys or the most recent decennial census. Poststratification, therefore, brings the weighted totals up to the level of the presumed number of noninstitutionalized children in the United States aged 3–15.

The poststratification factors, $f_{i(\mathrm{PS})}$, were calculated as

$$f_{i(\mathrm{PS})} = \frac{N_j}{\sum_{i=1}^{n_j} w_{i(\mathrm{TR})}}$$

where N_j is the control total and n_j is the sample size of the poststratification cell. Thus, the poststratified weights, $w_{i(\mathrm{PS})}$, were calculated as

$$w_{i(\mathrm{PS})} = w_{i(\mathrm{NR})} f_{i(\mathrm{PS})}$$

Computing Final Weights

The final weight for each sampled child at each stage was calculated as the product of the base weight and the nonresponse adjustment, trimming, and poststratification factors; that is,

$$w_i = w_{i(\mathrm{base})} f_{i(\mathrm{NR})} f_{i(\mathrm{TR})} f_{i(\mathrm{PS})}$$

More specifically, the final screening weight was calculated as

$$w_{i(\mathrm{S})} = w_{i(\mathrm{base})} f_{i(\mathrm{NR,S})} f_{i(\mathrm{TR,S})} f_{i(\mathrm{PS,S})}$$

and the final interview weight was calculated as

$$w_{i(\mathrm{I})} = w_{i(\mathrm{base})} f_{i(\mathrm{NR,S})} f_{i(\mathrm{TR,S})} f_{i(\mathrm{PS,S})} f_{i(\mathrm{NR,I})} f_{i(\mathrm{TR,I})} f_{i(\mathrm{PS,I})}$$

so that the final examination weight was calculated to be

$$w_{i(\mathrm{E})} = w_{i(\mathrm{base})} f_{i(\mathrm{NR,S})} f_{i(\mathrm{TR,S})} f_{i(\mathrm{PS,S})} f_{i(\mathrm{NR,I})} f_{i(\mathrm{TR,I})} f_{i(\mathrm{PS,I})} f_{i(\mathrm{NR,E})} f_{i(\mathrm{TR,E})} f_{i(\mathrm{PS,E})}$$

Only the interview and examination weights were released to the public.

Any sampled participant who did not respond to the interview was assigned an interview weight of zero. These sampled participants were considered ineligible for the examination and assigned an examination weight of zero as well. Their records were not released to the public. Sampled participants who completed the interview and were eligible for the examination, but did not respond, were assigned examination weights of zero, and their records are included in the public release.

The interview weight should be used for analyses of data from the household interview only. The examination weights should be used for analyses of data from the examination exclusively, or in conjunction with the household interview data.

Combined NHANES/ NNYFS Weights

The 2012 NNYFS was conducted not only at the same time as the 2012 NHANES but also in the same PSUs and segments. In addition, some questionnaire items and some physical measurement tests were collected in the same manner in both surveys. Therefore, it is possible to combine the samples to increase the sample sizes for these common items. The weights of the two surveys were combined to form new sample composite weights that could be used to analyze the combined sample.

As noted earlier, a much smaller set of 2011 NHANES questionnaire and examination items were comparable to those in the 2012 NHANES and 2012 NNYFS. This happened because the first year of the NHANES 2011–2012 survey cycle was fielded before changes were implemented to make the 2012 NHANES and the 2012 NNYFS more comparable. Regardless, it is possible to combine the 2012 NNYFS with the 2011–2012 NHANES for these limited items. Sample weights were also created for this combined data set.

Only data from the NNYFS sample are released to the public. Due to confidentiality restrictions, data from either of the combined files described above are available only through RDC.

Combined 2012 weights

The process of compositing began with the final interview and examination weights for the separate 2012 NHANES and NNYFS samples. These weights were then combined into one file and the weights adjusted within subgroups, or compositing domains, defined by race and Hispanic origin, sex, and age. The method of calculating the adjustment factors, known as compositing factors, was designed so that estimates from the combined sample would result in the lowest variance for key statistics.

The combined interview and examination weights were then each poststratified to the same totals for the civilian noninstitutionalized population from the 2011 ACS used in the NNYFS weighting.

Creation of compositing factors

Calculation of the compositing factors was based on examination data and weights from the two annual samples. The compositing factor was calculated as

$$\alpha_j = \frac{n^N_{eff,j}}{n^N_{eff,j} + n^Y_{eff,j}}$$

where $n^N_{eff,j}$ represents the effective sample size in compositing domain j from NHANES, and $n^Y_{eff,j}$ represents the effective sample size in compositing domain j from NNYFS. The effective sample size is defined as the sample size divided by the design effect. It captures aspects of the sample design that are likely to affect variance, regardless of the choice of statistic. The design effect for each sample (S) was approximated as the product of the design effect due to clustering ($\mathrm{Deff}^S_{C,j}$) times the design effect due to unequal weighting ($\mathrm{Deff}^S_{W,j}$), as in

$$\mathrm{Deff}^S_j = \left(\mathrm{Deff}^S_{C,j}\right)\left(\mathrm{Deff}^S_{W,j}\right)$$

Eighteen values of α were calculated, one for each compositing domain defined by race and Hispanic origin (Hispanic, non-Hispanic black, and non-Hispanic white and other), sex, and age category (3–5, 6–11, and 12–15). These compositing factors, along with the design effects and effective sample sizes, are shown in Table III.

Adjusted weights

A total of 1,253 children aged 3–15 responded to the NHANES interview in 2012, and 1,640 children responded to the NNYFS interview, resulting in a final total of 2,893 children when the two samples were combined. Each final NHANES interview weight was adjusted by the value of α_j corresponding to the sampled child's compositing domain shown in Table III; each final NNYFS interview weight in compositing domain j was adjusted by one minus the value of α_j. That is, the initial combined interview weights $w_{i(\text{base,I})}^{C_{12}}$ were initially set at

$w_{ij(\text{I})}^{N_{12}} \propto_j$, for child i in compositing domain j in 2012 NHANES sample

and

$w_{ij(\text{I})}^{Y_{12}}(1-\propto_j)$, for child i in compositing domain j in NNYFS sample

For example, the NHANES interview weights for non-Hispanic black males aged 3–5 from that sample were multiplied by 0.5715, while the NNYFS weights for non-Hispanic black males aged 3–5 from that sample were adjusted by 1 – 0.5715 = 0.4285.

These initial weights were poststratified to totals of the civilian noninstitutionalized population from the 2011 ACS, which was also the source for the 2012 NHANES and NNYFS annual weights. The weights were adjusted to the compositing domains, which are the same as the 18 poststrata used for poststratifying the NNYFS sample. Although Asian persons were oversampled for NHANES, the sample size for that group in the combined sample was not large enough to separate them from non-Hispanic white and other persons for poststratification.

The poststratification factors, $f_{i(\text{PS,I})}^{C_{12}}$, were calculated as

$$f_{i(\text{PS,I})}^{C_{12}} = \frac{N_j}{\sum_{i=1}^{n_j} w_{i(\text{base,I})}^{C_{12}}}$$

where N_j is the control total and n_j is the sample size of the poststratification cell. Thus, the final poststratified interview weights for the combined 2012 NHANES/NNYFS sample were calculated as

$$w_{i(\text{I})}^{C_{12}} = w_{i(\text{base,I})}^{C_{12}} f_{i(\text{PS,I})}^{C_{12}}$$

For NHANES, 1,186 of the 1,253 interview respondents aged 3–15 completed an examination in 2012. For NNYFS, 1,576 of the 1,640 interview respondents completed an examination. This resulted in 2,762 examination respondents in the combined sample. The weights for the combined sample were calculated in a manner similar to that used for the interview weights. Each final NHANES examination weight in domain j was multiplied by the value of α_j corresponding to the sampled child's domain shown in Table III; each final NNYFS examination weight in domain j was adjusted by 1 minus the value of α_j corresponding to the sampled child's domain. That is, the initial combined examination weights $w_{i(\text{base,E})}^{C_{12}}$ were initially set at

$w_{ij(\text{E})}^{N_{12}} \propto_j$, for child i in compositing domain j in 2012 NHANES sample

and

$w_{ij(\text{E})}^{Y_{12}}(1-\propto_j)$, for child i in compositing domain j in NNYFS sample

The combined NHANES/NNYFS examination weights were then poststratified to the same totals as the interview weights, using the same poststrata. The poststratification factors, $f_{i(\text{PS,E})}^{C_{12}}$, were calculated as

$$f_{i(\text{PS,E})}^{C_{12}} = \frac{N_j}{\sum_{i=1}^{n_j} w_{i(\text{base,E})}^{C_{12}}}$$

where N_j is the control total and n_j is the sample size of the poststratification cell. Thus, the final poststratified examination weights for the combined 2012 NHANES/NNYFS sample were calculated as

$$w_{i(\text{E})}^{C_{12}} = w_{i(\text{base,E})}^{C_{12}} f_{i(\text{PS,E})}^{C_{12}}$$

Combined 2011–2012 NHANES/2012 NNYFS weights

To produce the 2-year weights, the final weights for children aged 3–15 from the 2011 NHANES and the final composited weights for the 2012 NHANES/NNYFS were divided by 2 and combined. The combined weights were then reviewed for extreme weights that might require trimming. After trimming, the combined interview and examination weights were then each poststratified to the same totals used in the NNYFS weighting for the civilian noninstitutionalized population from the 2011 ACS.

The 1,387 children aged 3–15 who responded to the NHANES 2011 interview, 1,253 children who responded to the NHANES 2012 interview, and 1,640 children who responded to the NNYFS interview resulted in a total of 4,280 when the samples were combined. The final interview weights for the 2011 NHANES sample and the final composited interview weights for the 2012 NHANES/NNYFS were combined, and the initial combined interview weights $w_{i(\text{base,I})}^{C_{11,12}}$ were initially set at

$w_{i(\text{I})}^{N_{11}}$, for child i in 2011 NHANES sample

and

$w_{i(\text{I})}^{C_{12}}$, for child i in 2012 NHANES or NNYFS samples

The combined weights were reviewed for extreme values that might need trimming. Trimming thresholds were based on the mean weight for each compositing domain. These were defined by race and Hispanic origin (Hispanic, non-Hispanic black, and non-Hispanic white and other), sex, and age category (3–5, 6–11, and 12–15). Any weight exceeding five times the domain mean was trimmed down to that level. The excess weight was then distributed within the same domains so that the weights for each domain summed to the weighted sum prior to trimming, with the exception of the subdomains in the non-Hispanic white and other group. These were further

divided by Asian/non-Asian and low income/not low income, with the children selected through the NNYFS sample considered to be not low income. Non-Hispanic non-black Asian persons, and low-income white and other persons, were oversampled in 2011–2012 NHANES. In the full sample, 10 weights required trimming.

The trimming factors, $f_{i(\mathrm{TR,I})}^{C_{11,12}}$, were calculated as

$$f_{i(\mathrm{TR,I})}^{C_{11,12}} = \frac{\sum_{i=1}^{n_b} t_i}{\sum_{i=1}^{n_b} w_{i(\mathrm{base,I})}^{C_{11,12}}}$$

where n_b is the sample size of the b-th trimming domain and t_i is equal to $w_{i(\mathrm{base,I})}^{C_{11,12}}$, provided that this product does not exceed the threshold and is set to be equal to the threshold otherwise. The trimmed weights were calculated as

$$w_{i(\mathrm{TR,I})}^{C_{11,12}} = w_{i(\mathrm{base,I})}^{C_{11,12}} f_{i(\mathrm{TR,I})}^{C_{11,12}}$$

Once again, the adjusted weights were poststratified to totals of the civilian noninstitutionalized population from the 2011 ACS, the same source used for the 2011–2012 NHANES and 2012 NNYFS weights. The weights were adjusted to the same poststrata used for poststratifying the NNYFS and combined 2012 NHANES/NNYFS samples. Although non-Hispanic non-black Asian persons were oversampled for NHANES, again the sample size for that group in the combined sample was not large enough to separate them from the non-Hispanic white and other persons for this adjustment. The poststratification factors, $f_{i(\mathrm{PS,I})}^{C_{11,12}}$, were calculated as

$$f_{i(\mathrm{PS,I})}^{C_{11,12}} = \frac{N_j}{\sum_{i=1}^{n_j} w_{i(\mathrm{TR,I})}^{C_{11,12}}}$$

where N_j is the control total and n_j is the sample size of the poststratification cell. Thus, the final poststratified interview weights for the combined 2011–2012 NHANES/NNYFS sample were calculated as

$$w_{i(\mathrm{I})}^{C_{11,12}} = w_{i(\mathrm{TR,I})}^{C_{11,12}} f_{i(\mathrm{PS,I})}^{C_{11,12}}$$

For NHANES 2011, 1,330 of the 1,387 interview respondents aged 3–15 completed an examination; for NHANES 2012, 1,186 of the 1,253 interview respondents aged 3–15 completed an examination. For NNYFS, 1,576 of the 1,640 interview respondents completed an examination. This resulted in 4,092 examination respondents in the combined sample. The final examination weights for the 2011 NHANES sample and the final composited examination weights for the 2012 NHANES/NNYFS were combined, and the initial combined examination weights, $w_{i(\mathrm{base,E})}^{C_{11,12}}$, were initially set at

$w_{i(\mathrm{E})}^{N_{11}}$, for child i in the 2011 NHANES sample

and

$w_{i(\mathrm{E})}^{C_{12}}$, for child i in the 2012 NHANES or NNYFS sample

The combined weights were reviewed for extreme values that might need trimming. The methodology described for the interview weights was also used for the examination weights. In the full sample, eight weights required trimming.

The combined NHANES/NNYFS examination weights were then poststratified to the same totals as the interview weights, using the same poststrata, to arrive at the final poststratified examination weights for the combined 2011–2012 NHANES/NNYFS sample, $w_{i(\mathrm{E})}^{C_{11,12}}$.

Variance Estimation

Sampling errors should be calculated for all survey estimates to aid in determining the statistical reliability of those estimates. For complex sample surveys, exact mathematical formulas for variance estimates are usually not available. Variance approximation procedures are needed to provide reasonable, approximately unbiased and design-consistent estimates of variance. Although the NNYFS sample is nationally representative, it was selected from only 15 PSUs, and the sample sizes for some subdomains may be small. The small number of PSUs also poses challenges for variance estimation. With a small number of PSUs, direct design-based variance estimates may be unstable for some measures. In addition, because variance computations must incorporate the NNYFS design, standard statistical software routines (i.e., software packages that assume a simple random sample) should not be used for computing variances for NNYFS. This section introduces design-based methods of variance estimation for complex sample survey data and describes the creation of variables necessary for variance estimation on the public- and restricted-use data files for the NNYFS sample.

Two variance approximation procedures that account for the complex sample design and allow the computation of design effects are replication methods and Taylor series linearization.

Replication methods provide a general means for estimating variances for the types of complex sample designs and weighting procedures usually encountered in practice. The basic idea behind the replication approach is to select subsamples repeatedly from the whole sample, to calculate the statistic of interest (or replicates) for each of these subsamples, and then to use the variability among these replicate statistics to estimate the variance of the full-sample statistic. The jackknife and balanced repeated replication (BRR) methods are two common procedures for deriving replicates from a full sample. The jackknife procedure retains most of the sample in each replicate, whereas the BRR approach retains a portion of the sample in each replicate.

For the linearization approach, nonlinear estimates are approximated by linear ones for estimating variance. The linear approximation is derived by taking the first-order Taylor series approximation for the estimator. Standard variance estimation methods for linear statistics are then used to estimate the variance of the linearized estimator. Currently, NCHS recommends using Taylor series linearization methods for variance estimation in analyses of all

NNYFS data. SUDAAN, Stata, R, and SAS survey procedures can be used to obtain variance estimated by this method.

Variance Estimation for Publicly Released NNYFS Data

For the NNYFS sample, the 15 PSUs are considered to be from one stratum in order to estimate sampling error using the Taylor series linearization approach. The small number of PSUs in the NNYFS sample, geographic data and other characteristics of the area on the data files, and local publicity campaigns while the survey is in the field all pose a risk for data disclosure. As a result, masked variance units (MVUs) are provided for use with the public-use file to reduce the chance of an intruder being able to match PSUs in the sample to PSUs in the population, while minimizing the bias in the variance caused by altering the PSU structure. MVUs can be used as if they were pseudo-PSUs to estimate sampling errors, as in NHANES.

The MVUs, or pseudo-PSUs, on the data file are not the ''true'' design PSUs. They are instead a collection of secondary sampling units (SSUs) aggregated into groups for variance estimation. They produce variance estimates that closely approximate the variances that would have been estimated using the true design PSUs.

Many surveys swap data values between cases for disclosure limitation. Rather than swapping individual values, however, the procedure used in NNYFS, described by Park et al. (4), swapped entire segments (SSUs) between PSUs. That is, for two similar segments in different PSUs, the PSU and variance stratum identifiers for all sampled cases were swapped. Any PSUs with swapped segments are no longer completely associated with a single real PSU; thus, the chance of correctly matching a given individual within the PSU is limited. The point estimates of the overall population means do not change under this PSU masking, but the variance estimates may change slightly.

To identify which segments to swap in NNYFS, estimates were first calculated for all of the segments in all of the study locations for comparative purposes. These estimates provide general descriptions of the segments, such as the percentage of sampled participants with a particular race or Hispanic origin or obesity prevalence that should be similar for swapped segments. Then study locations that were the most at risk for data disclosure (locations with smaller populations or in rural areas) were identified.

Within each of these at-risk locations, each segment was paired with all segments from the other study locations (including other at-risk locations), and a distance measure was calculated to determine the effect on variance by swapping the pair. The distance measure was calculated as

$$D = \sum_{l=1}^{q} \left| \frac{v(\bar{x}_l|S^r) - v(\bar{x}_l|S)}{v(\bar{x}_l|S)} \right|$$

where q is the number of variables used to calculate the estimates, l is an individual estimate, $\bar{x}_1$ is the mean of that estimate, $v(\bar{x}_1|S^r)$ is the variance of the estimate after swapping, and $v(\bar{x}_1|S)$ is the variance of the estimate before swapping.

Within each at-risk location, the segments were sorted by smallest distance measure achieved, and some segments were selected to be swapped. Generally, pairs with the smallest distances were swapped, but if any two pairs included the same segment, one pair was not used for swapping. In this way, a single segment was swapped only once. Consideration was also given to pairs of segments that came from at-risk study locations; swapping of such pairs was minimized where possible.

Further research by Park (5) indicated that variance estimates generally tended to increase as more segments were swapped, although the variance for specific analysis variables could also be underestimated after swapping. For this reason, the amount of swapping (i.e., the number of study locations determined to be at risk and the number of segments swapped per location) is limited.

Variance Estimation for Combined NHANES/ NNYFS Data in RDC

Special unmasked PSU and stratum codes (which differ from the MVU codes provided for public-use files) were created and are available for use in the RDC for variance estimation with data from the combined 2012 NHANES/NNYFS and the combined 2011–2012 NHANES/NNYFS files. These true (unmasked) design codes are needed when true geographic linkage of the above combined files with some external data sets is required.

More information on the RDC and lists of special NHANES data files are available from the NHANES website: http://www.cdc.gov/nchs/nhanes/participant.htm. Information on RDC proposals is also available from the NHANES website.

Cautionary Notes

The purpose of compositing the NHANES and NNYFS samples was to join two surveys representing the same population to form a single sample by adjusting the weights so that the combined sample may represent the same population.

While the method used to combine the NHANES and NNYFS annual samples would produce unbiased estimates for any set of compositing factors, the optimum values for any particular statistic are the ones that result in the lowest variance for that statistic. However, because only one set of factors was created, the values will not be optimal for a single given variable. For some statistics, estimates may be more precise if made on either sample alone.

Despite the fact that the sample sizes have increased in the combined NHANES/NNYFS 2011–2012 compared with NHANES 2011–2012 alone, the variability of the weights has also increased. This is essentially due to the fact that there is an important weight differential between NHANES 2011 and NHANES/NNYFS 2012. These two samples represent the same population,

but the latter sample has more than twice the number of cases as the former. As a consequence, weights from NHANES/NNYFS 2012 are much smaller than the corresponding NHANES 2011 weights, and the average weight reduction by compositing domain varies from 33.7% to almost 79.0%. This weight differential results in increased design effects and other measures of precision.

For most cases, combining the two samples improves the reliability of some statistics, but this is not the case for all situations. Other methods for combining the NHANES and NNYFS samples are available that may improve some results but adversely affect others. Given the sample design differences across the two samples, in addition to the factors mentioned in the ''Weighting Sample Data'' section, the weights for the combined sample are quite variable. Note the potential influence that cases with large weights can have on analyses, especially when extreme weights are associated with extreme data points. Additionally, analysts should be aware of potential differences between PSUs as a cause of high variances for specific analytic variables of interest.

In summary, awareness of these analytic limitations is advised when using either of the combined NNYFS and NHANES data files. This is especially true for the combined NNYFS/NHANES 2011–2012 data set, which has a very limited number of data items that can be analyzed and increased variability in sample weights.

References

1. Borrud L, Chiappa MM, Burt VL, et al. National Health and Nutrition Examination Survey: National Youth Fitness Survey plan, operations, and analysis, 2012. National Center for Health Statistics. Vital Health Stat 2(163). 2014.
2. U.S. Department of Health and Human Services. 2008 physical activity guidelines for Americans. 2008. Available from: http://www.health.gov/PAGuidelines/pdf/paguide.pdf.
3. Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National Health and Nutrition Examination Survey: Sample design, 2011–2014. National Center for Health Statistics. Vital Health Stat 2(162). 2014.
4. Park I, Dohrmann S, Montaquila J, Mohadjer L, Curtin LR. Reducing the risk of data disclosure through area masking: Limiting biases in variance estimation. Proceedings of the Survey Research Methods Section, American Statistical Association; p 1761–7. 2006.
5. Park I. Primary sampling unit (PSU) masking and variance estimation in complex surveys. Surv Methodol 34(2):183–94. 2008.

Appendix I. Glossary

Centers for Disease Control and Prevention (CDC)—One of the major operating components of the U.S. Department of Health and Human Services.

Department of Health and Human Services (HHS)—The U.S. government's principal agency for protecting the health of all Americans and providing essential human services, especially for those least able to help themselves. CDC, including the National Center for Health Statistics, operates under HHS authority.

Domain—A demographic group of analytic interest (analytic domain). Analytic domains may also be sampling domains if a sample design is created to meet goals for specific demographic groups. For NHANES National Youth Fitness Survey (NNYFS), sampling domains are defined by sex and age. See *Sampling domain*.

Dwelling unit (DU), housing unit—The house, apartment, mobile home or trailer, group of rooms, or single room occupied as separate living quarters (see *Group quarters*) or, if vacant, intended for occupancy as separate living quarters. Separate living quarters are those in which the occupants live separately from other persons in the building and which have direct access from outside the building or through a common hall. In this report, the term generally means those DUs that are eligible for the survey (i.e., excluding institutional group quarters), or that could become eligible (e.g., vacant at the time of sampling but which could be occupied once screening begins).

Group quarters—A place where people live or stay that is normally owned or managed by an entity or organization providing housing or services for the residents. These services may include custodial or medical care as well as other types of assistance, and residency is commonly restricted to those receiving these services. People living in group quarters usually are not related to each other. Group quarters include college residence halls, residential treatment centers, skilled nursing facilities, group homes, military barracks, correctional facilities, workers' dormitories, and facilities for people experiencing homelessness. These are generally grouped into two categories: institutional group quarters and noninstitutional group quarters.

Institutional group quarters—Group quarters providing formally authorized supervised care or custody in institutional settings, such as correctional facilities, nursing and skilled nursing facilities, inpatient hospice facilities, mental health or psychiatric hospitals, and group homes and residential treatment centers for juveniles. Institutional group quarters and are not included in the NHANES sample.

Noninstitutional group quarters—Group quarters that do not provide formally authorized supervised care or custody in institutional settings. These include college or university housing, group homes and residential treatment facilities for adults, workers' group living quarters and Job Corps centers, and religious group quarters. Noninstitutional group quarters are included in the NHANES and NNYFS samples.

Household—The group of persons living in an occupied dwelling unit.

Low income—Beginning in 2000, NHANES split the sampling domains for white and other persons based on their income status into low income and non-low income. Low-income persons are those at or below 130% of the poverty level. The poverty threshold used in this determination was based on the most recent poverty guidelines published by HHS; these thresholds are updated annually by the U.S. Census Bureau.

Masked variance units (MVUs)—A collection of secondary sampling units aggregated into groups for variance estimation, designed to not reveal the identity of the selected primary sampling units (PSUs). For NHANES, rather than using the units as sampled, some pseudo-units are created by swapping segments between PSUs. The resulting units produce variance estimates that closely approximate the ''true'' design variance estimates. MVUs have been created for all 2-year survey cycles, from NHANES 1999–2000 through 2009–2010. They can also be used for analyzing any combined 4-, 6-, or 8-year data set.

Maximum sampling rate ($\max\{r_k\}$)—The largest probability of selection assigned to a demographic group within a survey design. This value within certain strata and demographic groups was used in determining the sample size and other sampling parameters in NHANES and NNYFS.

Measure of size (MOS)—A value assigned to every sampling unit in a sample selection, usually a count of units associated with the elements to be selected. For NHANES and NNYFS, the MOS used for PSU and segment selection is actually a weighted average of estimates of population counts for the NHANES race-Hispanic origin-income groups of interest.

National Center for Health Statistics (NCHS)—The nation's principal health statistics agency, which designs, develops, and maintains a number of systems that produce data related to demographic and health concerns. These include data on registered births and deaths collected through the National Vital Statistics System, National Health Interview Survey or NHIS, National Health and Nutrition Examination Survey or NHANES, National Health Care Surveys, and National Survey of Family Growth or NSFG, among others. NCHS is one of 13 centers within CDC, which is part of HHS.

Noninstitutional group quarters—See listing under *Group quarters*.

Noninstitutionalized civilian population—Includes all people living in households, excluding institutional

group quarters and those persons on active duty with the military. This is the target population for NHANES and NNYFS.

Primary sampling unit (PSU)—The first-stage selection unit in a multistage area probability sample. In NHANES and NNYFS, PSUs are counties or groups of counties in the United States. Some PSUs have such a large MOS that they are selected into the survey with a probability of one. These are referred to as PSUs selected with certainty (certainty PSUs); all other PSUs are selected without certainty (noncertainty PSUs).

Probability proportionate to size (PPS) sampling—In this method, the probability of selecting any unit varies with the size of the unit, giving larger units a greater probability of selection and smaller units a lower probability. NHANES and NNYFS use PPS sampling in the selection of primary and secondary sampling units (PSUs and segments).

Public-use file—An electronic data set containing respondent records from a survey with a subset of variables collected in the survey that have been reviewed by analysts within NCHS to ensure that the respondents' identities are protected. NCHS disseminates this file to encourage widespread use of the survey data.

Race and Hispanic origin—The term used in this report as it was used in the NHANES sample selection covering four groups: Hispanic, non-Hispanic black, non-Hispanic non-black Asian, and a fourth group consisting of all others.

Replicates—Subsamples selected repeatedly from a sample used in some variance estimation approaches. The statistic of interest is calculated for each subsample, and the variability among the replicate statistics is used to estimate the variance of the full-sample statistic. The jackknife and balanced repeated replication or BRR methods are two common procedures for deriving replicates from a full sample.

Respondent—A person selected into a sample who agrees to participate in all aspects of a survey. In NNYFS, persons agreeing to complete the in-home interviews are interview respondents. Persons agreeing to complete the in-home interviews and an examination at a mobile examination center (MEC) are MEC respondents.

Response rate—The number of survey respondents divided by the number of persons selected into the sample. Response rates in this report are MEC response rates, calculated as the number of people receiving examinations in the MEC divided by the total number of people sampled.

Restricted-use file—An electronic data set of survey respondent records containing some information that may, if released to the public, risk disclosing individual survey respondents. The data are available only through the NCHS Research Data Center. These data sets include a) NHANES data items collected for an odd number of calendar years (1, 3, or 5 years); b) data geographically linked to other contextual data files (often supplied by the data user); and c) data items determined to be too sensitive or detailed to be released to the public due to confidentiality restrictions.

Sample weight—For each NNYFS respondent, the sample weight is the estimated number of persons in the target population that he or she represents. For example, if a boy in the sample represents 12,000 boys in his age group, then his sample weight is 12,000. The NNYFS sample weights were adjusted for different sampling rates (of the sex-age groups), different response rates, and different coverage rates among persons in the sample, so that accurate national estimates can be made from the sample. Because it is the product of all of these adjustments, it is sometimes called the ''final'' sample weight.

Sampling domain—NNYFS includes six sampling domains, and Table I in this report contains the specific sampling domains; see also *Domain*.

Sampling rate—The rate at which a unit is selected from a sampling frame. For NNYFS, the rates required for sampling persons in the sex-age domains were designed to achieve the designated number of MEC examinations in each of those domains. The sampling rates are the driving force in all stages of sampling.

Screener—An interview (usually short) containing a set of questions asked of a household member to determine whether the household contains anyone eligible for the survey. In NNYFS, the screener, or screening interview, consisted of a household roster collecting the income level of the household and the sex and age of all members. In NNYFS, only persons aged 18 and over could answer the screener.

Screening—The process of conducting, or attempting to conduct, the screening interview in selected dwelling units. Occupied dwelling units (households) are ''screened'' through the screening interview. Other units can also be screened; the process for these units is verification that they are either vacant or not DUs. See *Screener*.

Secondary sampling unit (SSU)—The second-stage selection unit in a multistage area probability sample. For NHANES and NNYFS, these are typically referred to as ''segments.''

Segment—A group of housing units located near each other, all of which were considered for selection into the sample. For NNYFS, segments consisted of a census block or groups of blocks selected at the second stage of sampling. Within each segment, a sample of DUs was selected.

Self-weighting sample—A sample for which each elementary unit in the population has the same nonzero chance of selection into the sample; that is, they are selected with the same constant probability. Higher-stage sampling units may be selected with differing probabilities, but such differences in selection probabilities at various stages cancel out. NNYFS is a self-weighting sample of persons within each sampling domain.

Strata, stratification—The partitioning of a population of sampling units into mutually exclusive categories (strata).

Typically, stratification is used to increase the precision of survey estimates for subpopulations important to the survey's objectives.

Study location—The set of segments within a PSU that were fielded together, with all MEC examinations conducted at the same physical location. The distinction between a PSU and a study location is necessary because some large certainty PSUs were divided into multiple study locations and fielded at different times.

Target population—The population to be described by estimates from the survey. In NNYFS, the target population was the resident civilian noninstitutionalized population of the United States, which excluded all children in supervised care or custody in institutional settings, active-duty family members living overseas, and any other persons residing outside of the 50 states and District of Columbia.

Undercoverage—The result of failing to include all of the target population in the sampling frame.

Variance—A measure of the dispersion of a set of numbers. In this report, the variance is specifically the sample variance, which is a measure of the variation of a statistic, such as a proportion or mean, calculated as a function of the sampling design and the population parameter being estimated. Many common statistical software packages compute population variances by default, which may underestimate the sampling variance because they do not incorporate any effects of having taken a sample compared with collecting data from every person in the full population. Estimating the variance in NNYFS requires special software, as discussed in this report.

Variance stratum—The cluster of variance units used when forming a replicate for variance estimation.

Variance unit—A collection of SSUs aggregated into groups and excluded when forming a replicate for variance estimation.

Weight—See *Sample weight*.

Appendix II. Supporting Tables

Table I. Target sample sizes, screening amounts, and sampling rates: NNYFS, 2012

Age (years) and sex	Projected population in 2011–2014[1]	Target number of NNYFS examinations	Projected amount of screening required to obtain one examined person[2]	Projected amount of screening required to obtain target examinations in self-weighting area sample[2]	Numerator of sampling rate[3]
Male			Number of households		
3–5	6,418,335	173	23	3,991	0.9629
6–11	12,983,222	346	11	3,946	0.9520
12–15	8,519,830	231	17	4,009	0.9672
Female					
3–5	6,418,335	173	23	3,991	0.9629
6–11	12,477,784	346	12	4,106	0.9906
12–15	8,239,995	231	18	4,145	1.0000
Total	55,057,503	1,500	. . .	. . .	. . .

. . . Category not applicable.

[1]Population projection created for NHANES 2011–2014 primary sampling unit (PSU) selection.

[2]Estimated number of occupied households is 118,410,614, as created for NHANES 2011–2014 PSU selection.

[3]For a sample including a 50% reserve; denominator is 19,044.

NOTE: NNYFS is National Health and Nutrition Examination Survey's (NHANES) National Youth Fitness Survey.

Table II. Variables used to form nonresponse adjustment cells for weighting interview and examination samples: NNYFS, 2012

Variables considered for nonresponse	Categories of variables cross-classified	
	Interview	Examination
Race and ethnicity of sampled person	Non-Hispanic black, Non-Hispanic non-black Asian, Hispanic, other	. . .
State grouping based on health characteristics of population (used in PSU stratification)[1]	Healthy states, California, somewhat healthy states, fairly healthy states, poorly healthy states	. . .
Urbanicity	Urban areas with population over 3 million, all other urban areas, suburban areas, rural areas	. . .
Sex of household reference person	Male, female	. . .
Age (years) of sampled person	3–5, 6–11, 12–15	. . .
Age (years) of household reference person	Under 30, 30–39, 40–49, 50 and over	. . .
Census region	Northeast, Midwest, South, West	Northeast, Midwest, South, West
Number of sampled persons in household	. . .	1, 2, 3, 4 or more

. . . Category not applicable.

[1]State health-related variables used to derive health ranking are death rate, infant mortality rate, percentage of adults with high blood pressure, percentage of adults overweight or obese, percentage of adults with poor nutrition, and percentage of adults who smoke. PSU is primary sampling unit.

NOTE: NNYFS is National Health and Nutrition Examination Survey's (NHANES) National Youth Fitness Survey.

Table III. Design effects, effective sample sizes, and final compositing factors used to combine annual samples: National Health and Nutrition Examination Survey and NHANES National Youth Fitness Survey, 2011–2012

Domain (*j*), by race and ethnicity, sex, and age (years)	Design effect		Effective sample size		α_j
	NHANES	NNYFS	NHANES	NNYFS	
Non-Hispanic black					
Male:					
3–5	1.14	1.11	42.18	31.62	0.5715
6–11	1.17	1.31	67.46	67.93	0.4983
12–15	1.09	1.23	41.45	49.58	0.4554
Female:					
3–5	1.08	1.18	29.62	34.79	0.4599
6–11	1.18	1.27	79.93	75.87	0.5130
12–15	1.11	1.21	34.20	42.87	0.4438
Hispanic					
Male:					
3–5	1.12	1.14	44.84	54.28	0.4524
6–11	1.23	1.32	79.90	87.75	0.4766
12–15	1.10	1.19	39.09	55.28	0.4142
Female:					
3–5	1.11	1.21	36.13	49.65	0.4212
6–11	1.19	1.25	66.14	89.06	0.4262
12–15	1.13	1.17	45.90	56.39	0.4487
Non-Hispanic white and other					
Male:					
3–5	1.87	1.23	30.94	66.63	0.3171
6–11	1.81	1.29	66.83	118.38	0.3609
12–15	2.08	1.26	34.55	97.32	0.2620
Female:					
3–5	1.61	1.25	36.08	57.70	0.3847
6–11	1.91	1.33	67.57	125.87	0.3493
12–15	1.91	1.30	26.19	95.73	0.2148

NOTE: NHANES is National Health and Nutrition Examination Survey, and NNYFS is NHANES National Youth Fitness Survey.

Vital and Health Statistics Series Descriptions

ACTIVE SERIES

Series 1. **Programs and Collection Procedures**—This type of report describes the data collection programs of the National Center for Health Statistics. Series 1 includes descriptions of the methods used to collect and process the data, definitions, and other material necessary for understanding the data.

Series 2. **Data Evaluation and Methods Research**—This type of report concerns statistical methods and includes analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory. Also included are experimental tests of new survey methods, comparisons of U.S. methodologies with those of other countries, and as of 2009, studies of cognition and survey measurement, and final reports of major committees concerning vital and health statistics measurement and methods.

Series 3. **Analytical and Epidemiological Studies**—This type of report presents analytical or interpretive studies based on vital and health statistics. As of 2009, Series 3 also includes studies based on surveys that are not part of continuing data systems of the National Center for Health Statistics and international vital and health statistics reports.

Series 10. **Data From the National Health Interview Survey**—This type of report contains statistics on illness; unintentional injuries; disability; use of hospital, medical, and other health services; and a wide range of special current health topics covering many aspects of health behaviors, health status, and health care utilization. Series 10 is based on data collected in this continuing national household interview survey.

Series 11. **Data From the National Health Examination Survey, the National Health and Nutrition Examination Surveys, and the Hispanic Health and Nutrition Examination Survey**—In this type of report, data from direct examination, testing, and measurement on representative samples of the civilian noninstitutionalized population provide the basis for (1) medically defined total prevalence of specific diseases or conditions in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics, and (2) analyses of trends and relationships among various measurements and between survey periods.

Series 13. **Data From the National Health Care Survey**—This type of report contains statistics on health resources and the public's use of health care resources including ambulatory, hospital, and long-term care services based on data collected directly from health care providers and provider records.

Series 20. **Data on Mortality**—This type of report contains statistics on mortality that are not included in regular, annual, or monthly reports. Special analyses by cause of death, age, other demographic variables, and geographic and trend analyses are included.

Series 21. **Data on Natality, Marriage, and Divorce**—This type of report contains statistics on natality, marriage, and divorce that are not included in regular, annual, or monthly reports. Special analyses by health and demographic variables and geographic and trend analyses are included.

Series 23. **Data From the National Survey of Family Growth**—These reports contain statistics on factors that affect birth rates, including contraception and infertility; factors affecting the formation and dissolution of families, including cohabitation, marriage, divorce, and remarriage; and behavior related to the risk of HIV and other sexually transmitted diseases. These statistics are based on national surveys of women and men of childbearing age.

DISCONTINUED SERIES

Series 4. **Documents and Committee Reports**—These are final reports of major committees concerned with vital and health statistics and documents. The last Series 4 report was published in 2002. As of 2009, this type of report is included in Series 2 or another appropriate series, depending on the report topic.

Series 5. **International Vital and Health Statistics Reports**—This type of report compares U.S. vital and health statistics with those of other countries or presents other international data of relevance to the health statistics system of the United States. The last Series 5 report was published in 2003. As of 2009, this type of report is included in Series 3 or another series, depending on the report topic.

Series 6. **Cognition and Survey Measurement**—This type of report uses methods of cognitive science to design, evaluate, and test survey instruments. The last Series 6 report was published in 1999. As of 2009, this type of report is included in Series 2.

Series 12. **Data From the Institutionalized Population Surveys**—The last Series 12 report was published in 1974. Reports from these surveys are included in Series 13.

Series 14. **Data on Health Resources: Manpower and Facilities**—The last Series 14 report was published in 1989. Reports on health resources are included in Series 13.

Series 15. **Data From Special Surveys**—This type of report contains statistics on health and health-related topics collected in special surveys that are not part of the continuing data systems of the National Center for Health Statistics. The last Series 15 report was published in 2002. As of 2009, reports based on these surveys are 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, and so compilations of Advance Data reports are no longer needed.

Series 22. **Data From the National Mortality and Natality Surveys**—The last Series 22 report was published in 1973. Reports from these sample surveys, based on vital records, are published in Series 20 or 21.

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, and so compilations of reports 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 5419
Hyattsville, MD 20782

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

U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES

Centers for Disease Control and Prevention
National Center for Health Statistics
3311 Toledo Road, Room 5419
Hyattsville, MD 20782

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

DHHS Publication No. 2015–1368, Series 2, No. 168
CS250447