



# **Changes to Data Editing Procedures and the Impact on Identifying Same-Sex Married Couples: 2004–2007 National Health Interview Survey**

**Division of Health Interview Statistics  
National Center for Health Statistics  
Hyattsville, Maryland**

**Centers for Disease Control and Prevention  
U.S. Department of Health and Human Services**

**April 2015**

## Introduction

This report describes an internal review of the data editing and error correction procedures used in the designation of sex for same-sex married couples in the 1997–2011 National Health Interview Survey (NHIS). This review was initiated in response to a data request by an outside researcher who was conducting a secondary analysis of the 1997–2010 NHIS public use data files. This researcher alerted the Division of Health Interview Statistics (DHIS) staff at the National Center for Health Statistics (NCHS) to the possibility of miscodes of the sex variables for same-sex married couples in the 2004–2007 NHIS, which the researcher had identified in the course of conducting an analysis of the health of same-sex couples. Specifically, the researcher observed a spike in the numbers of same-sex married couples in the 2004–2007 data compared to the numbers observed in the 1997–2003 and 2008–2010 data. In response to this inquiry, the DHIS staff conducted a review of NHIS data editing procedures since 1997 and confirmed that a change in data verification procedures from 2004 to 2007 may have led to uncorrected errors in identification of same-sex married couples in the 2004–2007 NHIS. While it is not possible to determine if any same-sex married couples were incorrectly identified as opposite-sex married couples, some opposite-sex married couples are believed to have been incorrectly identified as same-sex for data years 2004–2007. This report provides a general summary of the DHIS staff’s review of data editing and error correction procedures since 1997, with more detail on editing procedures for the sex and household/family relationship variables.

## NHIS Data Collection Procedures

The NHIS data are collected through in-person household interviews conducted by field representatives (interviewers) employed and trained by the U.S. Census Bureau, using procedures specified by NCHS (1). The Census Bureau, under contractual agreements with NCHS, has been the data collection agent for the NHIS (2) since the survey first went into the field in 1957. The main sampling unit in the NHIS is the household. An NHIS household contains one or more families, and an NHIS family is defined as an individual or a group of two or more related persons who are living together in the same household. Unmarried couples (same-sex and opposite-sex couples) are considered as belonging to the same family.

During the interview, a household roster is developed that includes all members of the household and identifies the members of each family in the household. One “household reference person” (usually the person who owns or rents the housing unit) is chosen for the household, and one “family reference person” is designated for each family in the household. In single-family households, the family reference person is the same person as the household reference person. Questions are asked to determine the relationship of each household member to the household reference person and, for each family, the relationship of each family member to the family reference person. Additional questions on relationships allow for the identification of all spouses or cohabiting partners residing in the household, regardless of whether one of them is a reference person.

## NHIS Data Editing and Verification Procedures

The NHIS data editing process consists of two major phases. Briefly, Phase 1 involves a series of preliminary data editing steps (pre-processing) that prepare the data for extensive testing and review in Phase 2.

In Phase 1, generic edits are performed to check for valid response codes for each questionnaire variable, convert data from “enter all that apply” questions to “mentioned/not mentioned” variables, and convert family-level data collected in the survey to person-level data. A review of major demographic variables such as sex, age,

date of birth, and household/family relationship information is performed to identify and resolve any potential relationship inconsistencies. Pre-processing of the NHIS data is performed on a quarterly basis. Missing data in the race and ethnicity fields are imputed, and tabulations and data files are created for weighting and further data editing in Phase 2.

Phase 2 of the data editing process involves a rigorous period of testing of the data, including review of frequency distributions for variables in specific sections of the questionnaire by the topic-area experts assigned to those sections, checks to determine that data values are within reasonable boundaries given the possible responses to a given question, and comparison of frequency distributions for selected variables with those from previous survey years to check for consistency.

For the purposes of this report, further discussion of the NHIS data editing process is limited to Phase 1, as this is the stage of editing in which changes occurred over time in the processing of the sex and relationship variables.

Because legal marriage of same-sex couples was largely banned in the United States before 2004 and still remains relatively rare, all records for same-sex married couples are reviewed and flagged as candidates for verification. However, the NHIS editing and verification processes underwent changes since 1997; these changes and their effects on sex codes for married couples are the focus of the remainder of this report. The [table](#) summarizes these changes and their impact on identifying same-sex married couples for three NHIS data collection time periods: **1997 through 2003**, in which sex codes for flagged married couples identified as same-sex were further reviewed by DHIS and, if necessary and possible, corrected; **2004 through quarter 2 of 2007**, in which sex codes for flagged married couples identified as same-sex were not further reviewed and therefore not corrected; and **starting in quarter 3 of 2007**, since when sex codes for flagged married couples identified as same sex have been further reviewed by the Census Bureau using additional verification procedures and, if necessary and possible, corrected. Also, since 2008, a verification screen in the NHIS instrument has been used to check certain sex and relationship codes during the actual interview. The rest of this section describes these changes in the NHIS data editing procedures in greater detail.

From 1997–2003, same-sex married couples were identified as part of one large-scale relationship edit. This included manual review by DHIS staff of confidential data files containing household members' names for all married couples identified as same-sex to try to confirm the sex of each spouse. If sex was believed to be incorrectly recorded for one spouse based on this information, the sex code of that spouse was changed. In some cases, the name provided adequate information to determine with high probability the person's correct sex. If the sex code was changed, the values of other correlated variables were adjusted to maintain consistency, if possible, but since that interview could not be conducted again, some variable outcomes had to be recoded as unknown to avoid inconsistencies. In other cases, names were gender-neutral or no other information was available to determine if the sex code was correct, so the relevant data were left unchanged.

In 2004, DHIS implemented a new NHIS instrument that utilized Windows-based computer-assisted interviewing software called Blaise. In addition, data production processes were streamlined and automated using SAS to limit the amount of manual review of individual records and thereby improve the timeliness of the release of the annual public use data file. The newly adopted practice was to send records back to the Census Bureau for verification only if the same data inconsistency occurred for 3% or more of the families in a given quarter. Public data release occurred more quickly because low-prevalence errors were not closely scrutinized. Records with specific data inconsistencies in household/family relationships were still noted and flagged during NHIS pre-processing, but because the number of flagged cases was so small relative to the whole sample, no such cases underwent a manual review from 2004 through the second quarter of 2007. However, DHIS staff continued to monitor these flagged cases during this period and considered approaches for verifying this information during the interview itself rather than as part of the post-interview data editing process.

---

During the summer of 2007, DHIS staff decided to send all flagged records with data inconsistencies for household/family relationships back to the Census Bureau for verification, starting with data from quarter 3 of 2007. Since then, and on a quarterly basis, DHIS staff asks the Census Bureau to try to determine if the sex code is correct for each spouse in a married couple identified as being same-sex, and to make any necessary and possible corrections. The Census Bureau verifies these data through several steps. First, Census Bureau staff reviews the case level notes for those identified cases to see if the interviewer included notes that used words such as “wife,” “husband,” “he,” or “she” that verify the sex of each person. If the case level notes are not informative, the Census Bureau Regional Office contacts the field supervisor and the specific interviewer in question to ask the interviewer to recall the interview. If the interviewer is unavailable or does not recall the details of the case, a Census Bureau representative at the Regional Office will usually phone the respondent to try to verify the case. Lastly, if the Census Bureau is unable to reach the respondent for verification, the Census Bureau will manually review confidential data files containing the household members’ names to try to confirm the sex of each same-sex spouse. As mentioned earlier, if sex was believed to be incorrectly recorded for one spouse based on this information, the sex code of that spouse was changed. In some cases, the name provided adequate information to determine with high probability the household member’s correct sex. However, in other cases, a correction could not be made because the name was gender-neutral and no other information was available to determine if the sex code was correct.

In addition, a verification screen for same-sex married couples was added to the NHIS instrument starting with the 2008 NHIS to be displayed to the interviewer shortly after the household roster is generated. [Figures 1](#) and [2](#) illustrate how this process works. [Figure 1](#) shows the relationship screen in the Blaise instrument that allows for a household member to be identified by his/her relationship to the household reference person. The household respondent is asked to identify all of the persons residing in the household and what each person’s relationship is to the household reference person. This information generates a series of follow-up questions designed to determine (1) the number of families in the household, (2) the composition of each family in the household, (3) the relationship of each family member to the family reference person, and (4) the presence in the family of spouses or cohabiting partners. If two family members are identified as being married to each other and are of the same sex, a verification screen ([Figure 2](#)) is triggered that asks the interviewer to verify that this information is correct. If it is not correct, the interviewer has the opportunity to go back and correct this information at the time of interview.

## Impact of Changes in NHIS Data Editing and Verification Procedures

Our review of data editing procedures indicated that there is likely to have been misclassification of sex in the NHIS data for 2004 through quarter 2 of 2007 such that some opposite-sex married couples were coded as being same-sex. As shown in the [table](#), for 1997–2003, a total of 49 same-sex married couples are on the final edited files. Then, for the single-year 2004, when verification of sex did not occur, 111 same-sex married couples are on the file. Later, for 2008, after the amount of verification had increased and when a verification screen was added to the NHIS instrument, only 30 same-sex married couples were on the file pre-edit; of these, 20 were found to be in error and the sex code of one partner was changed, leaving 10 same-sex married couples on the final edited file. Similarly, for 2009–2011, the percent of married couples identified pre-edit as being same-sex that were later found to have an incorrect sex code for one of the partners (which was therefore corrected) ranged from 68% to 73%.

Errors in sex codes were more likely to occur for persons who were not chosen as the “sample adult.” In each family in each interviewed household, a knowledgeable adult “family respondent” answers questions about all family members. Then, a knowledgeable adult answers questions about a randomly selected “sample child” (if there is at least one child in the family), and a randomly selected “sample adult” answers questions about

himself/herself. Our review suggests that misclassification of sex was more likely for persons who were not chosen as the sample adult because the sex of the sample adult is verified at the beginning of the non-proxy sample adult interview (unless the sample adult is also the family respondent).

## **Informal Review by DHIS Staff of Selected Sex Codes for Same-Sex Married Couples**

DHIS staff conducted an informal review of sex codes for all same-sex married couples in public-use NHIS microdata for quarter 1 of 2004 through quarter 2 of 2007. The review process and its results are described below.

The public use DHIS microdata files for the review period (January 2004 through June 2007) contain 273 same-sex married couples. By definition, members of such couples have either two sex codes signifying male or two sex codes signifying female. Two DHIS staff members compared the sex codes with the first names for both members of all 273 same-sex married couples on those files.

For each person in each couple, the two reviewers independently noted whether they thought the sex code could be incorrect because the person's first name was not gender-appropriate given their sex code. The reviewers also noted for each person their degree of certainty (sure or unsure) as to whether the sex code was incorrect. To make these decisions, they used their own judgment. Of course, sometimes the gender associated with a given name could not be determined by a reviewer because the name is appropriate for both sexes; misspelled; unfamiliar to the reviewer; missing from the data file; or of Chinese origin (because Chinese given names cannot be characterized as being male or female).

The two reviewers compared their opinions and found a number of inconsistencies between their opinions. For cases where the two reviewers disagreed or about which they were unsure, they consulted an Internet site that they had recently discovered—<http://www.genderchecker.com>—that provides an opinion of whether the gender associated with a given name is male (e.g., Richard), female (e.g., Mary), unisex (meaning either male or female, e.g., Sydney), or not on the database.

The two reviewers ultimately formed a consensus on which sex codes were likely to be incorrect. For 210 of the 273 same-sex married couples, one sex code was likely incorrect; if these sex codes were changed, those 210 couples (77% of the initial 273 same-sex married couples) would be counted as opposite-sex married couples. For two of the 273 same-sex married couples, both sex codes were likely to be incorrect; if these sex codes were changed, those two couples would still be counted as same-sex married couples. Thus, based on this informal review, if sex codes that were judged as likely incorrect were changed, the total number of same-sex married couples would be  $273 - 210 = 63$ .

## **Discussion**

Though NHIS data are subject to extensive quality review processes prior to public release, it is not possible to evaluate all data elements, and some errors remain. Analysts should be aware of the limitations of the data identifying same-sex married couples and use caution in analyzing them. Identification of sex and relationship status are both necessary to classify same-sex married couples, and measurement errors (especially misreporting of sex) have been noted previously for the U.S. Census (3). Similar errors are noted here for the NHIS, and they were not corrected from 2004 through quarter 2 of 2007. It is not feasible to re-edit the data for

2004 through quarter 2 of 2007 the way such data were edited before 2004, nor can the data for 2004 through quarter 2 of 2007 be re-edited the way such data have been edited since quarter 3 of 2007.

The number of same-sex married couples in the NHIS is small relative to the overall number of couples (same-sex and opposite-sex) in the NHIS sample. However, as noted in the [table](#), the number of incorrect sex codes identified through the verification process implemented in quarter 3 of 2007 is large compared to the number of married couples initially identified as being same-sex. Assuming that a similar number of errors occurred for data collected from 2004 through quarter 2 of 2007—an assumption that has merit given the results of the informal review of sex codes for same-sex married couples identified during this time period—these errors will affect analyses of this subgroup for the relevant years. Therefore, NCHS discourages analyses of this subgroup for those data years.

It is worth noting that the editing and verification procedures described here are limited to the sex codes for same-sex married couples. Since 1997 (and likely before then), verification of sex codes has not been done for same-sex cohabiting couples or for opposite-sex couples (married or cohabiting). Measurement error in responses to the relationship questions or the coding of relationship status may also occur, though such error was less common than misreporting of sex in the 2010 census (3). Verification of NHIS relationship codes has not been done for any couples, same-sex or opposite-sex. The extent to which any of these potential errors may affect analyses is unknown.

Since the number of same-sex married couples remains small and there is the potential for error in both sex and relationship codes, the analytic potential of the NHIS for studying same-sex married couples remains limited, even for data collected before 2004 or after 2007. Some researchers, however, are interested in these data because membership in a same-sex marriage or union is often used as a surrogate for sexual identity. Since January 2013, the NHIS has included questions on sexual identity (4). While these questions are asked only of sample adults and not about all adults in the family, this additional information may be useful for researchers interested in studying issues related to LGB health.

## References

1. National Center for Health Statistics. Survey Description, National Health Interview Survey, 2013. Hyattsville, MD. 2014. [ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Dataset\\_Documentation/NHIS/2013/srvydesc.pdf](http://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2013/srvydesc.pdf).
2. U.S. Census Bureau. National Health Interview Survey. CAPI Manual for NHIS Field Representatives. January 2013. [ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Survey\\_Questionnaires/NHIS/2013/frmanual.pdf](http://ftp.cdc.gov/pub/Health_Statistics/NCHS/Survey_Questionnaires/NHIS/2013/frmanual.pdf).
3. DeMaio TJ, Bates N, O'Connell M. Exploring measurement error issues in reporting of same-sex couples. *Public Opin Q* 77(S1):145-158. 2013.
4. Dahlhamer JM, Galinsky AM, Joestl SS, Ward BW. Sexual orientation in the 2013 National Health Interview Survey: A quality assessment. *Vital Health Stat* 2(169). 2014.  
[http://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_169.pdf](http://www.cdc.gov/nchs/data/series/sr_02/sr02_169.pdf).

**Table. Changes in data verification and their impact on identifying same-sex married couples in the 1997–2011 National Health Interview Survey**

Stage of processing	1997–2003 <sup>1</sup>	2004 <sup>2</sup>	2005 <sup>2</sup>	2006 <sup>2</sup>	Q1–Q2, 2007 <sup>2</sup>	Q3–Q4, 2007 <sup>3</sup>	2008 <sup>3</sup>	2009 <sup>3</sup>	2010 <sup>3</sup>	2011 <sup>3</sup>
Number of same-sex married couples identified by DHIS during pre-processing	unknown	111	68	68	26	19	30	52	76	59
Number of couples with corrected sex code of one spouse following Census Bureau verification	unknown	...	...	...	...	16	20	38	52	41
Number of same-sex married couples following Census Bureau verification	49	...	...	...	...	3	10	14	24	18
Percentage of same-sex married couples identified by DHIS during pre-processing that were corrected by Census Bureau	...	...	...	...	...	84.2%	66.7%	73.1%	68.4%	69.5%

... Not applicable.

<sup>1</sup>During National Health Interview Survey (NHIS) pre-processing, same-sex married couples are identified as part of the data editing process. From 1997–2003, Division of Health Interview Statistics (DHIS) staff identified same-sex married couples as part of one large-scale relationship edit and manually reviewed the spouses' names to try to confirm the sex of each spouse. If sex was believed to be incorrectly recorded for one spouse based on this information, the sex code of that spouse was changed.

<sup>2</sup>In 2004, DHIS implemented a new NHIS instrument that utilizes Windows-based computer-assisted interviewing software. In addition, data production processes were streamlined and automated. Records with data inconsistencies in relationships were flagged, but such records were only sent back to the Census Bureau for verification when inconsistencies occurred in 3% or more of families in a given quarter. As a result, from 2004 through quarter 2 (Q2) of 2007, the sex codes for married couples identified as being same-sex did not undergo further verification.

<sup>3</sup>Starting in quarter 3 (Q3) of 2007, DHIS staff began sending all flagged records back to the Census Bureau for verification. Census staff is asked to verify the sex of each spouse in a same-sex married couple and correct the sex code if in error. Starting in 2008, a verification screen for same-sex married couples was added to the instrument for display to the interviewer shortly after the household roster is generated.

Forms Answer Navigate Options Help Show Watch Window

NHIS Roster FAQ Status F10

H3 ? [F1]

What is Garry Public's relationship to you?

2. Spouse (husband/wife)       9. Grandparent  
(Grandmother/Grandfather)  
 3. Unmarried Partner       10. Aunt/Uncle  
 4. Child       11. Niece/Nephew  
(biological/adoptive/in-law/step/foste  
r)  
 5. Child of Partner       12. Other relative  
 6. Grandchild       13. Housemate/Roommate  
 7. Parent       14. Roomer/Boarder  
(biological/adoptive/in-law/step/foste  
r)  
 8. Brother/Sister       15. Other non-relative  
(biological/adoptive/in-law/step/foste  
r)

	Relationship to Ref Person	Child Type	Parent Type	Sibling Type
John Doe		1		
Garry Public		2		

Figure 1. Example of household relationship screen in the National Health Interview Survey instrument since 2004

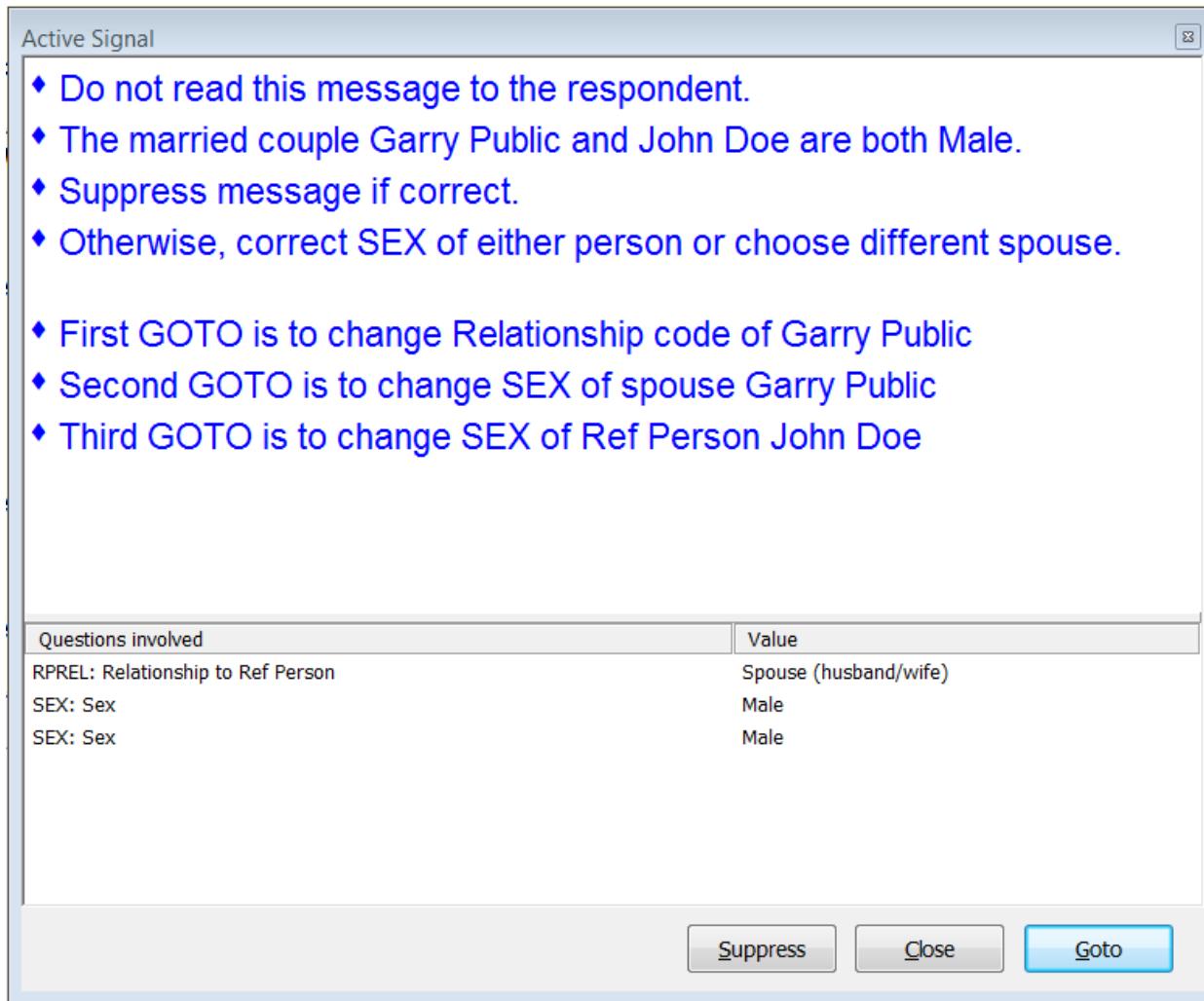


Figure 2. Example of verification screen added to the National Health Interview Survey instrument in the first quarter of 2008