

## DATE OF REFERENCE FOR AGE AND BIRTH DATE USED BY RESPONDENTS OF CENSUS 2000

Nathan Carter<sup>1</sup>

United States Bureau of the Census, Washington, D.C. 20233

**KEY WORDS:** Self Response, Reporting Error, Questionnaire Design

Followup took place from April 26, 1990 through July 30, 1990.

### Introduction

The goal of this paper was to see how well respondents answered the Census as of Census Day, April 1, 2000. One way to do this is to look at how respondents answered the age and date of birth question. The way respondents answer these questions could be influenced by whether or not they are using Census Day as their date of reference.

The Census 2000 questionnaire was modified significantly from the form used in 1990. The format of the form was the most significant change. The wording of the age question changed, so that it specifically stated that the respondent should report age as of April 1, 2000. This change was designed to reduce the discrepancy between the reported age and the actual age. Also, instead of just asking the respondent to provide a year of birth, the entire date of birth was requested.

### Background

The 1990 Census questionnaire asked for the age and year of birth for each person in the household. No instruction was given for the respondent to answer the question in reference to Census Day, April 1, 1990. Some discrepancy resulted between the reported age and the actual age calculated from the year of birth. The Census Bureau staff examined this discrepancy using the following method:

The timing of the questionnaire delivery in Census 2000 was earlier than in the 1990 Census. The delivery of the Census 2000 questionnaires took place from March 13, 2000 to March 15, 2000. The time frame for Nonresponse Followup enumeration was from April 27, 2000 to June 26, 2000.

“April 1, 1990 is the 90<sup>th</sup> day of the year and therefore, 24.7 percent of the year (containing 365 days). For most birth years about 24.7 percent of respondents should have had a birthday before April 1, assuming birthdays are equally distributed throughout the year. In such cases the person’s age added to the year of birth always equals “1990.” For the other 75.3 percent of respondents the persons age added to their year of birth always add up to “1989.” In 1990, 34.3 percent of the respondents’ age added to their birth year, equaled 1990. This number was not consistent with 24.7 percent that was expected from looking at April 1, 1990. What day would be consistent with the 34.3 percent observed in the 1990 Census? The answer was May 5, 1990, which is 34.2 percent of a 365 day year. The connection was made that this would represent the true 1990 Census Moment (Spencer, 1997).”

### Source File and the Universe Creation

The data file used for this analysis was the Hundred Percent Census Unedited File (HCUF). This file included some housing units that were later removed during the unduplication process. A total of 1,392,686 housing units in the United States and Puerto Rico were removed during this process and were not included in this analysis. As a result, the persons from these housing units were also not included in this analysis.

The time at which the enumeration took place may have affected responses to the age question. The time frame for the 1990 Census questionnaire delivery was approximately on March 23, 1990. Nonresponse

The HCUF was used so that analysis could be done on data solely provided by the respondent prior to the editing and imputation process. This file included data that were blank or invalid values. Persons were removed from the analysis if any of the following conditions were met.

- Age, month, day or year of birth was left blank,
- Month or day of birth was an invalid value,
- Age reported by respondents was greater than 115, or
- Age calculated from date of birth was less than 0 or greater than 115.

---

<sup>1</sup> This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

The cases where the first bullet apply, meaning the respondent left one or more of the parts of the date of birth or the age question blank, were removed from the data file first. The cases where the last three bullets apply, meaning the respondent provided some information that was considered to be invalid, were removed from the data file during a subsequent step. Table 1 contains a breakdown of persons, with the duplicates removed, by whether or not they were included in the analysis and the reason for exclusion.

**Table 1. Results from Performing Edits**

	Number	Percent
Total	271,541,738	100.0
Included in the Analysis	252,490,497	93.0
Blank Data	18,196,157	6.7
Invalid Data	855,084	0.3

As shown in Table 1, 93.0 percent of persons were included in further analysis. This also means that 7.0 percent of persons were not included in the analysis. This breaks down to 6.7 percentage points being excluded from the analysis because of some data being blank and 0.3 percentage points were excluded because of some of the data being invalid values. The 252,490,497 persons, 93.0 percent, is the base universe in the analysis.

**The Census Moment or “Average” Date of Reference**

The methodology for computing the Census Moment had been modified from what was used in 1990. As stated in the previous section, the Census 2000 questionnaire asked for respondents to provide the entire date of birth. This allowed for a distribution of the number of persons born on each day throughout the year with valid data to be calculated. Therefore, the assumption that was made for the 1990 Census analysis, that dates of birth were equally distributed through the year, was not necessary.

A person’s age added to his or her date of birth would show whether that person’s age had incremented for that year or not, or in other words the person’s age implies having had a birthday. For example, if a person was born on March 25, 1975 and the age was reported at 25, then the sum of the year of birth and age would be 2000. On the other hand, if the age was reported as 24, then the sum would be 1999. The sum of 2000 shows the age having been incremented for the year of 2000, while 1999 shows that the age has not yet been incremented. This sum was done for every person included in the

analysis.

If every person’s age was correctly reported, the proportion of sums that equaled 2000 would be equal to the proportion of persons who have a birth between January 1 and April 1. If the proportion was different, it indicated that some date other than April 1, 2000 was used as a reference date. If the proportion that was observed was matched up to a distribution of dates of birth throughout the year, the day corresponding to the percentage indicated the “average” date of reference.

The concept of a date of reference referred to whatever date the respondent was referring to when he or she was answering the age question. The questionnaire asked the respondents to use April 1, 2000 as the date of reference for the age question.

To calculate the Census Moment or “average” date of reference, the initial step was to sum the year or birth and the age reported by the respondent. As stated in the methodology section, the expected values from calculating this sum are 2000 and 1999. A sum of 2000 would mean that the person’s age had been incremented for the year, while 1999 would mean that the person’s age has not yet changed for the year. Table 2 summarizes the results from summing of age and year of birth.

**Table 2. Sum of Year of Birth and Age**

	Number	Percent
Total	252, 490, 497	-
1999	171,056,027	70.1
2000	73,109,542	29.9
Some Other Sum*	8,324,928	-

\*This category is not included in the calculation of the percentage.

As shown in Table 2, there were 8,324,928 persons when the sum was computed had a sum that was a value other than 2000 or 1999. These persons could not be included in the calculation of the Census Moment or “average” date of reference. Of the remaining people, 29.9 percent of them had an observed sum of 2000. These are persons whose age had incremented for the year, meaning their age reflected having had a birthday. The remaining 70.1 percent had a sum of 1999.

The final step in calculating the Census Moment or “average” date of reference is to comparing the 29.9 percent from the pervious step to the distribution date of birth. The 29.9 percent falls between two days,

April 19, with a proportion of 29.8 percent, and April 20, with a proportion of 30.0 percent. The 29.9 percent for the sum of 2000 corresponds to April 20. This is quite a big difference from May 5, which was observed in 1990. There are a couple of reasons why this change may have occurred. The first is the change to the questionnaire so that respondents are asked to report age as of April 1, 2000. The second reason would be the earlier dates for delivery of mail questionnaires and the completion of Nonresponse Followup in 2000 compared to 1990.

**Final Mail Return Rates and Date of Reference**

If the date in which a respondent is completing his or her form affects how he or she reports age, then at the state level, the return rate would be related to the states' date of reference. Most mail response happens early in the Census, the most often precluded the Housing Unit from being enumerated in Nonresponse Followup, which would have the respondent's enumeration at a date, after April 1, 2000. This means that the expected effect would be that as the return rate increases the date of reference for the state would be earlier in the year. A discussion of the final return rate follows.

Final mail return rates were also used in the analysis. It is a measure of respondent cooperation in mailback areas. It refers to the number of occupied housing units with corresponding non-blank questionnaires checked in through the end of the year (December 31, 2000) over the number of occupied housing units. The calculation of these rates was restricted to housing units that were in one of the mailback types of enumeration areas - Mailout/Mailback, Update/Leave, Military, Urban Update/Leave, or Mailout/Mailback converted to Update/Leave.

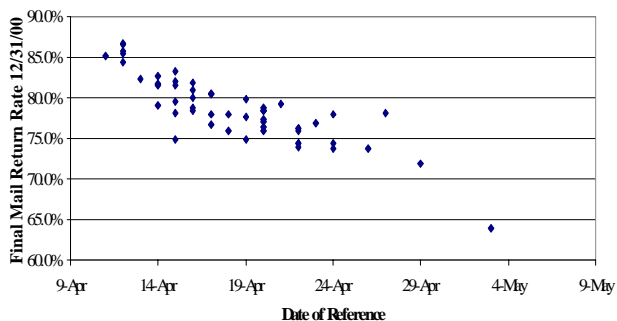
To be included in the final return rate denominator, an address must have been an occupied housing unit, in a mailback type of enumeration area, and not a Census undeliverable as addressed questionnaire. A Census undeliverable as addressed is a questionnaire in the Mailout/Mailback universe that was never successfully delivered to an address, either by the U.S. Postal Service or by Census Bureau employees. Deleted addresses in update/leave and urban update/leave also were excluded from the mail return rate denominator. Additionally, any address included in the denominator must have been added to the Decennial Master Address File extract through an operation that occurred prior to Nonresponse Followup. The March 2001 Master Address File extract was used to determine whether an address was added in one of those pre-Nonresponse Followup operations.

In order to have been included in the final return rate numerator, an address must have been in the

denominator and have a non-blank mail return data capture. Those non-blank questionnaires included actual mail return questionnaires, Be Counted Forms, Internet returns, and responses via Telephone Questionnaire Assistance or Coverage Edit Followup. The existence of a data capture was determined using information from the Decennial Response File - Stage 2 (U.S. Bureau of the Census, 2002).

Figure 1 is a scatter plot of the return rate as of December 31, 2000 for each state and Puerto Rico versus the corresponding date of reference for that state and Puerto Rico.

**Figure 1: Mail Return Rates for Each State and Puerto Rico**



As shown in Figure 1, there is a clear relationship between date of reference and final mail return rate. As the final mail return rate increases the date of reference moves earlier in the year. Therefore, states with higher final mail return rates have dates of reference that are closer to April 1, which should be the date of reference when reporting age. Note that no state (including Puerto Rico) had a reference day before April 11.

**Age Misreporting at the Person Level**

The Census 2000 questionnaire asked for respondents to provide a completed date of birth. This allowed for analysis that was not possible with the 1990 Census data. A calculated age was computed as of April 1, 2000 from the date of birth provided by the respondent. A person's age was considered to have been misreported if the age reported for that person differed from the age calculated from date of birth. As stated previously, the assumption made is that date of birth is always correctly reported. This means that if there is a discrepancy between the reported age and the calculated age, it is due to the respondent misreporting age. Table 3 gives the results of the comparison of the calculated age to the age reported.

**Table 3. Outcome of Reporting Age Compared to Calculated Age**

	Number	Percent
Total	252,490,497	100.0
Under by More than 1 Year	2,949,505	1.2
Under by 1 Year	4,601,172	1.8
Ages are Consistent	226,762,801	89.8
Over by 1 Year	15,227,068	6.0
Over by More than 1 Year	2,949,951	1.2

As shown in Table 3, 89.8 percent of persons had their reported age consistent with calculated age, 3.0 percent of persons had an under reported age, and 7.2 percent had an over reported age. The evaluation is concerned with the date of reference affecting the reporting of age. The concept behind this is that persons responding to the Census before April 1, 2000 might have a tendency to under report their age by a year. For example, a person with the birthday of March 25, 1975 who is filling out the Census 2000 questionnaire on March 20, 2000 might report his or her age as 24 rather than 25, which would have been the correct age as of April 1, 2000.

On the other hand those persons responding to the Census after April 1, 2000 would have a tendency to over report their age by a year. For example, a person with a birthday of May 20, 1975 who is being interviewed during Nonresponse Followup on May 25, 2000 might report his or her age as 25 rather than 24, which would have been the correct age as of April 1, 2000. This theory does not explain why some people misreported their age by more than a year. The only explanation for the 2.4 percent of persons from Table 3 who had an age misreported by more than a year is simple misreporting error. The 5,899,456 such cases will not be included in the next table.

The date at which a respondent is answering the Census may influence how age is reported. The closest proxy for the date at which a respondent answers the Census is the date at which the questionnaire is checked in. This means that there are really three dates to consider: the date of birth, the date of check-in, and April 1, 2000. The following are the six possible ways to order these three dates within a year:

- Birthday/Check-in/April 1,
- Check-in/Birthday/April 1,

- Birthday/April 1/Check-in,
- Check-in/April 1/Birthday,
- April 1/Birthday/Check-in, and
- April 1/Check-in/Birthday.

Only in two of these possible situations, we expected respondents may have had difficulty in reporting age correctly. They are Check-in/Birthday/April 1 and April 1/Birthday/Check-in.

In the first case, respondents would have provided their age before they had a birthday. This means the respondents may have reported age without having incremented it for the year, but age should have been incremented if reported as of April 1, 2000.

In the second case, the respondents would have provided their age after they had a birthday. This means the respondents may have reported age having incremented if for the year, but age should not have been incremented if reported as of April 1, 2000.

In all the other cases, we expected that respondents would have reported their age correctly. Table 4 gives the outcome of age reporting broken down by each of the different date orders. Table 4 is located in the appendix.

Looking at Table 4, there are a few trends worth noting. In the two situations where we expected respondents may have had difficulty in reporting age correctly, there are anomalies in the percent of person misreporting age. In the Check-in/Birthday/April 1 situation, there is 10.3 percent of persons in this category who under reported their age by a year, which is the trend that was expected. It is also higher than what was observed for the other situations. In the April 1/Birthday/Check-in, there is 40.1 percent of persons in this category who over reported their age. This is much higher than what was observed for the other situations. For the majority of these persons, their data were collected by an enumerator during a personal visit interview. If the enumerators did not emphasize that age should be reported as of April 1, 2000, it may explain why this particular category is so high.

Another trend that can be observed in Table 3 is that the first three categories all have the birthday happening before April 1, while the last three have the birthday happening after April 1. For the categories with the birthday before April 1, age tended to be more under reported, while categories with birthday after April 1 tended to have age over reported.

**Age Misreporting at the Household Level**

The census is usually responded to by one person at each housing unit and all of the persons on each form are enumerated at the same time. This would mean that misreporting of age should be grouped because of these reasons.

The next table will examine misreporting of age at the household level. To be categorized as Age Under Reported in Table 5, at least one person had to have his or her age under reported but no one had their age over reported. To be categorized as Age Over Reported in Table 5, at least one person had to have his or her age over reported but no one had their age under reported. To be categorized as Age Under and Over Reported in Table 5, at least one person had to have his or her age under reported and at least one person had to have his or her age over reported. To be categorized as Age Correctly Reported in Table 5, every person in the household had to correctly report their birthday.

**Table 5. Outcome of Age Reporting at the Household Level**

	Number	Percent
Total	99,724,760	100.0
Under Reported	5,487,486	5.5
Correctly Reported	80,144,563	80.4
Over Reported	12,717,132	12.8
Both Over and Under	1,375,579	1.4

From Table 5, 80.4 percent of households had every person’s age correctly reported. This also means that 19.6 percent of households had at least one person’s age misreported.

This breaks down to 5.5 percent of households had at least one person with his or her age under reported, 12.8 percent that had at least one person with his or her age over reported, and 1.4 percent of with at least a person with under reported age and also at least one person with his or her age over reported.

By way of reminder, from Table 3, 89.8 percent of persons had his or her age correctly reported, and 10.2 percent had his or her age incorrectly reported.

**Limitations**

In data collection, it is impossible to know if the data provided by respondents was correctly reported. For this analysis this issue was important with respect to discrepancies between age and date of birth. It is

important to note that there was an assumption being made throughout this report, that date of birth was correctly reported. Therefore, all reported discrepancies were attributed to the respondent failing to correctly report their age.

**Conclusions and Recommendations**

The Census Moment or “average” date of reference moved from May 5 in 1990 to April 20 in 2000. This improvement may be due to the change in questionnaire design and in the enumeration time frame. The 2010 Census questionnaire should ask the respondents to provide their age as of Census Day, April 1, 2010. This will help respondents not misreport age. Also, a compressed Census enumeration time frame may aid respondents to correctly report age.

Respondents enumerated by personal visit tended to have a tendency to over report age. Enumerators should have this problem explained to them and training should stress the importance of Census Day. Enumerators should know that respondents need to be reminded of April 1, 2010, so they can correctly provide their information.

**References**

Spencer, Gregory. *What Was The “True” Census Moment of the April 1, 1990 Census and the March 2, 1996 National Content Survey?*, Memorandum for The Record, January 27, 1997.

U.S. Bureau of the Census, 2002, *Census 2000 Response and Return Rates - National and State by Form Type*, DSSD Census 2000 Procedures and Operations Memorandum Series L-10, Bureau of the Census, February 12, 2002.

U.S. Bureau of the Census, Draft, *Final Draft Report for Date of Reference for Age and Birth Date used by Respondents of Census 2000 - H.10*, Draft.

APPENDIX

**Table 4. Outcome of Reporting Age as Compared to Calculated Age by Each Date Order\***

	Total	Under By One Year	Age Reported Correctly	Over By One Year
Total				
Number	246,591,041	4,601,172	226,762,801	15,227,068
Percent	100.0	1.9	92.0	6.2
Birthday/Check-in/April 1				
Number	34,298,599	1,095,163	33,003,120	200,316
Percent	100.0	3.2	96.2	0.6
Check-in/Birthday/April 1				
Number	4,221,921	433,386	3,758,746	29,789
Percent	100.0	10.3	89.0	0.7
Birthday/April 1/Check-in				
Number	22,902,535	1,119,952	21,542,610	239,973
Percent	100.0	4.9	94.1	1.0
Check-in/April 1/Birthday				
Number	116,725,492	1,021,466	110,231,015	5,473,011
Percent	100.0	0.9	94.4	4.7
April 1/Birthday/Check-in				
Number	10,694,363	117,760	6,285,046	4,291,557
Percent	100.0	1.1	58.8	40.1
April 1/Check-in/Birthday				
Number	57,748,131	813,445	51,942,264	4,992,422
Percent	100.0	1.4	89.9	8.6

\*Percentages may not add to 100.0 percent because of rounding.