Census 2000 Evaluation F.16, Revision 1

April 7, 2004

Evaluation of the Block Splitting Operation for Tabulation Purposes

FINAL REPORT

This evaluation study reports the results of research and analysis undertaken by the U.S. Census Bureau. It is part of a broad program, the Census 2000 Testing, Experimentation, and Evaluation (TXE) Program, designed to assess Census 2000 and to inform 2010 Census planning. Findings from the Census 2000 TXE Program reports are integrated into topic reports that provide context and background for broader interpretation of results.

Somonica Green and
Cindy Rothhaas
Planning, Research, and
Evaluation Division

USCENSUSBUREAU
Helping You Make Informed Decisions

Intentionally Blank

CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

This evaluation measures the percent of the country affected by collection blocks split for tabulation purposes, and the accuracy of that block splitting. The country refers to all areas of the United States **except** Puerto Rico, the Island Areas, Remote Alaska, and collection blocks without housing units or group quarters. This evaluation does not measure geocoding error.

Collection blocks are geographic areas that are usually defined by visible features, and used by the Census Bureau to conduct field operations. For Census 2000, the Census Bureau classified each collection block in the country into one of the following nine types of enumeration areas:

- Mailout/mailback
- Update/leave
- List/enumerate
- Remote Alaska
- Rural update/enumerate
- Military
- Urban update/leave
- Urban update/enumerate
- Update/leave originally assigned to mailout/mailback

Often, collection blocks cross governmental unit boundaries such as city and town or other required data tabulation boundaries. At the end of Census 2000, the Census Bureau redefined the census collection blocks by recognizing the boundaries of governmental units and other geographic entities required for tabulation of census data. One of the steps needed to achieve this involved the Geography Division using an automated system to split collection blocks in certain situations. This block splitting process was based on address ranges and map spot information in the Topologically Integrated Geographic Encoding and Referencing database. The address ranges and map spot information were the cumulative results of various geographic programs and field operations for Census 2000.

For Census 2000, out of more than five million blocks, the Census Bureau split 915,794 blocks for tabulation purposes. A total of 282,457 of these blocks formed the sampling universe we used to evaluate the block splitting process. The 633,337 split blocks excluded from our sampling universe were either located in remote Alaska, located in Puerto Rico, were split by the boundaries of special purpose governmental or administrative entities such as school districts, were split by the boundaries of statistical entities, or contained no housing units or group quarters. Remote Alaska and Puerto Rico were excluded from the evaluation to minimize cost. Boundaries of special purpose governmental, administrative, and statistical boundaries were excluded because this evaluation relied on the knowledge of residents of the block and we did not believe they would know where these types of boundaries existed in their blocks. Split blocks that contained no housing units or group quarters were excluded because the purpose of the evaluation was to measure the error associated with placing housing units and group quarters in the wrong tabulation block. The estimated number of blocks that fell into each of these categories was not available, but the sum total was 633,337 blocks. From the sampling universe, we selected a sample of 1,000 blocks for field verification. Field representatives determined whether the housing units in these collection blocks were allocated to the correct side of a tabulation boundary, allocated to the wrong side of a tabulation boundary, or unresolved.

Our key findings follow.

How much of the country was affected by block splitting for tabulation purposes?

About five and one half percent of the 5.1 million collection blocks in the country were split for tabulation purposes. A little more than ten percent of the 115.5 million housing units and about nine percent of the 186,000 group quarters in the country were located in these split collection blocks.

What percent of the country was allocated to the wrong side of a tabulation boundary?

About 26 percent of the 283,000 split collection blocks contained at least one housing unit allocated to the wrong side of a tabulation boundary. These collection blocks represent about two percent of the collection blocks in the country.

The percent in error was lower for housing units and group quarters. Less than four percent of the 12 million housing units affected by block splitting were allocated to the wrong side of a tabulation boundary. These errors represent less than half of one percent of the housing units in the country. For the group quarters in our sample, none were allocated to the wrong side of a tabulation boundary. Although our estimate of the number of group quarters allocated to the wrong side of a tabulation boundary was zero, we cannot conclude that there were no group quarters in error throughout the country. The preliminary August 2002 results from one of our administrative programs, in which the Census Bureau receives input from local governmental entities, showed that 1,867 group quarters in the country were in fact allocated to the wrong side of a tabulation boundary. This is less than one percent of the group quarters in the country.

Did the percent in error vary by Type of Enumeration Area grouping?

For this evaluation, we categorized collection blocks in the mailout/mailback, military, urban update/leave, and urban update/enumerate enumeration areas as “inside the blue line.” The term “inside the blue line” refers to areas where almost all mail delivery was to city-style addresses. City-style addresses contain a house number and street name. Except for Remote Alaska, all other types of enumeration areas were categorized as “outside the blue line.” This term refers to areas where mail delivery was to noncity-style addresses. Examples of noncity-style addresses include rural route, highway contract route, P.O. Box, and location descriptions. A mixture of city-style and noncity-style addresses occur in some types of enumeration areas, especially those "outside the blue line."

For the housing units affected by block splitting, the percent in error for enumeration areas "inside the blue line" was comparable to the percent in error for enumeration areas "outside the blue line" – about four percent of 8.1 million housing units compared to close to three percent of 3.8 million housing units, respectively. For the housing units in the country, the percent in error for enumeration areas "inside the blue line" was also comparable to the percent in error for enumeration areas "outside the blue line" – 0.35 percent of 92.5 million housing units compared to 0.45 percent of 23 million housing units, respectively.

Did the percent in error vary by address type?

For those affected by block splitting, fewer than four percent of the 11.1 million housing units with city-style addresses and fewer than three and one half percent of the 773,000 housing units with noncity-style addresses were allocated to the wrong side of a tabulation boundary. The percent in error for both address types in the country was the same – 0.37 percent. There were about 109 million housing units with city-style addresses and seven million housing units with noncity-style addresses in the country.

Did the percent in error vary by block split/misallocation status?

The Census Bureau conducted the Census 2000 block split/misallocation operation to ensure that housing units, special places, crews of vessels, and group quarters in collection blocks split by tabulation boundaries were assigned to the appropriate collection block and in the correct geographic entity (for tabulation purposes). Because there were limited resources for research, the Geography Division created a series of flags to identify the different scenarios for research, and then prioritized the cases for field review. High priority cases were defined as follows: cases with map spots within 300 feet of an invisible boundary, cases with missing map spots in the Topologically Integrated Geographic Encoding and Referencing database, cases with missing map spots in both the Master Address File and the Topologically Integrated Geographic Encoding and Referencing database, cases with conflicting geographic codes in the Master Address File, and cases with conflicting geographic codes between the Master Address File and the Topologically Integrated Geographic Encoding and Referencing database.

The Field Division conducted the block split/misallocation operation by either using detailed map and address reference sources, making telephone calls to knowledgeable respondents, or conducting field visits to determine the location of the housing units and group quarters in relation to a boundary. Most of the work was conducted by field visits. If a field visit was conducted, the address was coded as "worked." If an address was flagged to be visited but was not visited due to limited resources, the address was coded as "flagged to be worked but not worked." All other addresses were coded as "not flagged to be worked and not worked."

The percent in error varied by block split/misallocation status. Nearly ten percent of the 1.1 million housing units that were “worked” in the block split/misallocation operation were allocated to the wrong side of a tabulation boundary. This compared to around two percent of the four million housing units that were “flagged to be worked but not worked,” and less than half of one percent of the 110.5 million housing units that were “not flagged and not worked.”

What can we conclude?

A little more than ten percent of the housing units in the country were in blocks split for tabulation purposes. Less than one half of one percent of the housing units in the country were allocated to the wrong side of a tabulation boundary. Less than four percent of housing units in split blocks were allocated in error. The percent in error for housing units by type of enumeration area grouping was comparable. The percent in error by address type was comparable as well.

Unlike type of enumeration area grouping and address type, the percent in error by block split/misallocation status varied. Because the percent in error was very small for units “not flagged and not worked,” it appears that the Census Bureau did a good job of flagging potential problems. However, the percent in error remained relatively high for cases that were “worked” in the block split/misallocation operation. More research is needed to fully understand why the percent in error for addresses “worked” was so high. We suspect that limited resources contributed to these errors. Ample time and sufficient staff to plan, implement, and execute the block split/misallocation operation, as well as adequate tools like the Global Positioning System to determine the exact location of housing units and group quarters may have decreased the percent in error to the level of the other two groups. Even so, we suggest more upfront design/planning for the 2010 Census to enhance the success of this operation.

1. BACKGROUND

Collection blocks, hereafter referred to as blocks, are geographic areas that are usually defined by visible features, and used by the Census Bureau to conduct field operations. Often blocks cross governmental unit boundaries (such as city and town) or other boundaries required for tabulation of census data. At the end of Census 2000, the Census Bureau redefined the census blocks by recognizing the boundaries of governmental units and other geographic entities required for tabulation of census data. One of the steps needed to achieve this involved the Census Bureau conducting block split operations to provide for tabulation of data where the boundaries of governmental units and geographic areas delineated for statistical purposes do not conform to block boundaries. The resulting blocks are tabulation blocks. This evaluation measures the percent of the country affected by blocks split for tabulation purposes, and the accuracy of the processes used to allocate housing units (HUs) and group quarters (GQs) in these split blocks to the correct side of a tabulation boundary.

1.1 1990 Census

For the 1990 Census, the Census Bureau required a clerical block split operation to ensure the correct geocoding of any map spot located within 200 feet of a tabulation boundary that split a block. Some work was completed for an evaluation, but the analysis was not conducted due to limited resources.

1.2 Relationship between blocks and Type of Enumeration Areas (TEAs)

For Census 2000, the Census Bureau classified each block in the country into one of the following nine enumeration areas:

TEA 1:	Mailout/mailback
TEA 2:	Update/leave
TEA 3:	List/enumerate
TEA 4:	Remote Alaska
TEA 5:	Rural update/enumerate
TEA 6:	Military
TEA 7:	Urban update/leave
TEA 8:	Urban update/enumerate
TEA 9:	Update/leave originally assigned to mailout/mailback

For this evaluation, we categorized the enumeration areas into two groups - “inside the blue line” and “outside the blue line.” These two groups represent the two general approaches for how the Census Bureau built the address list. Blocks in TEAs 1, 6, 7, and 8 are “inside the blue line.” Blocks in TEAs 2, 3, 5, and 9 are “outside the blue line.” Blocks in TEA 4 are excluded from this evaluation. The “inside the blue line” and “outside the blue line” terms are explained in the next section.

1.3 Process for splitting blocks

The process for splitting blocks was performed by the Geography Division (GEO) with assistance from the geographic staff in the Regional Census Centers (RCC) and the National Processing Center (NPC). The criteria for splitting blocks took into account whether the block was "inside the blue line" or "outside the blue line." The term "inside the blue line" refers to blocks with HU mailing addresses that are mostly city-style, where the Census Bureau conducted most of the enumeration by mailing questionnaires. In these areas, the Census Bureau used the address list created from the 1990 Address Control File (ACF) and the Delivery Sequence Files (DSFs) from the United States Postal Service (USPS). The Census Bureau also received updates from local and tribal governments, and conducted a 100 percent block canvassing to attempt to validate the completeness of the addresses and the block assignment of the HUs. Blocks in TEAs 1, 6, 7, and 8 are "inside the blue line."

Conversely, blocks that have more addresses that are noncity-style where the Census Bureau conducted the census by hand-delivering questionnaires or by enumerating the census in person are referred to as "outside the blue line." For most of these areas, the Census Bureau believed that problems were likely with developing an accurate mailing list and delivering the questionnaires to the intended HU through the mail. In these areas, the Census Bureau used address listing and mapspotting to create the initial address list and to determine the block assignment of each HU, and the update/leave operation to update and improve the addresses and their block assignments. In the remaining areas, the Census Bureau developed the address list at the time of enumeration. Blocks in TEAs 2, 3, 5, and 9 are "outside the blue line."

For the purpose of this evaluation, an address was defined as city-style if a house number, street name, and ZIP Code were present. In addition, for units "inside the blue line," an address was also defined as city-style if the building name and ZIP Code were present. All other addresses that did not fit these definitions were defined as noncity-style. Examples of addresses that were noncity-style include rural route (RR), highway contract route (HCR), P.O. Box, and location descriptions.

1.3.1 Blocks "inside the blue line"

For blocks "inside the blue line," the GEO performed an automated block split operation using the address range data in the Topologically Integrated Geographic Encoding and Referencing (TIGER) database. These address ranges are the cumulative results of the initial address information used for Census 2000 operations, the block canvassing operation, the Local Update of Census Addresses (LUCA) field verification, the Boundary and Annexation Survey (BAS)[1], the block split/misallocation operation, and other programs for Census 2000.

1.3.2 Blocks "outside the blue line"

To determine the addresses on either side of a tabulation block boundary for blocks "outside the blue line," the computer relied on the following:

- The placement of map spots recorded in the TIGER database as noted by the listers during the address listing, LUCA 1999 relisting, update/leave, update/enumerate, list/enumerate, and block split/misallocation operations;
- The responses to the BAS, and the insertion of those boundaries accurately into the TIGER database; and
- The address ranges in the TIGER database for areas "outside the blue line" that have city-style addresses.

1.3.3 The block split/misallocation operation

Because we specifically used the results from the block split/misallocation operation to evaluate the process for splitting blocks, a detailed description of the block split/misallocation operation follows.

The Census Bureau conducted the Census 2000 block split/misallocation operation to ensure that HUs, special places, crews of vessels, and GQs in blocks split by tabulation boundaries were assigned to the appropriate block and in the correct geographic entity (for tabulation purposes). Because there were limited resources for research, the GEO created a series of flags to identify the different scenarios for research, and then prioritized the cases for field review. High priority cases were defined as follows: cases with map spots within 300 feet of an invisible boundary, cases with missing map spots in the TIGER database, cases with missing map spots in both the MAF and the TIGER database, cases with conflicting geographic codes in the MAF, and cases with conflicting geographic codes between the MAF and the TIGER database.

---

[1]The Boundary and Annexation Survey (BAS) is a Census Bureau survey of counties/county equivalents, minor civil divisions (MCDs), incorporated places, and federally recognized American Indian reservations (AIRs). Its purpose is to determine the complete inventory and the correct names, legal descriptions, official status, and official boundaries of the Nation's legal entities as of January 1 of the year of the survey.

The Field Division (FLD) conducted the block split/misallocation operation by either using detailed map and reference sources, making telephone calls to knowledgeable respondents, or conducting field visits to determine the location of the HUs and GQs in relation to a boundary. Most of the work was conducted by field visits. If a field visit was conducted, the address was coded as "worked." If an address was flagged to be visited but was not visited due to limited resources, the address was coded as "flagged to be worked but not worked." All other addresses were coded as "not flagged to be worked and not worked."

2. METHODS

2.1 Sample design

For Census 2000, out of more than five million blocks, the Census Bureau split 915,794 blocks for tabulation purposes. From these blocks that had at least one tabulation boundary that split the block, we selected a sample of 1,000 blocks for field verification. The purpose of the field verification was to determine if the HUs and GQs in these blocks were allocated to the correct side of a tabulation boundary defined as of January 1, 2000[2].

2.1.1 Blocks included in the sampling universe

The universe from which we selected the sample consisted of 282,457 blocks throughout the country that met all the following conditions:

- Located in TEA 1, 2, 3, 5, 6, 7, 8, or 9;
- Located in all areas of the United States (U.S.) except Puerto Rico or the Island Areas;
- Had at least one HU or GQ included in Census 2000; and
- Split by one or more governmental unit boundaries for tabulation purposes. For this evaluation, governmental unit boundaries included incorporated places, federally recognized American Indian reservations (AIRs), federally recognized off-reservation trust lands, and minor civil divisions (MCDs). Only MCDs in the Northeast and Midwest Census regions were included (See Appendix A).

The remaining 633,337 split blocks did not meet the conditions above. Refer to Section 3.1 for additional information about these blocks.

[2]The legal boundaries that the Census Bureau recognizes for Census 2000 are those in effect on January 1, 2000.

2.1.2 Stratification variables used

We stratified the blocks in the sampling universe by the following variables in the order they are listed.

- TEA grouping – "Inside the blue line" and "outside the blue line"

  A block was categorized as "inside the blue line" if the block was in either TEA 1, 6, 7, or 8. If the block was in TEA 2, 3, 5, or 9, the block was categorized as "outside the blue line."

- Address type – City-style and noncity-style

  A block was defined as city-style if the majority of the addresses contained a house number, street name, and ZIP Code. If the majority of addresses in a block used RR, HCR, P.O. Box, or location descriptions, the block was defined as noncity-style.

- Block split/misallocation status – "Worked," "flagged to be worked but not worked," and "not flagged and not worked"

  A block was defined as "worked" if a field visit was conducted for at least one address in the block during the block split/misallocation operation. A block was defined as "flagged to be worked but not worked" if no addresses in the block were visited but at least one address in the block was flagged to be visited. All remaining blocks were defined as "not flagged and not worked."

2.2 Field procedures

For this evaluation, the FLD sent maps and listings of all addresses in the sampled split blocks to the Regional Offices (ROs). The ROs conducted the field verification from May 17, 2001 to June 29, 2001. Field verification was conducted by either observation, use of a knowledgeable source, or both. For each HU and GQ on the address listing page, the field representatives (FRs) entered one of the following valid action codes:

- "V" for addresses allocated to the correct side of a tabulation boundary,
- "C" for addresses allocated to the wrong side of a tabulation boundary, or
- "N" for addresses that the FR could not locate or resolve.

The "V" and "C" action codes are self-explanatory, but the "N" action code requires more explanation. For the "N" action code, the following scenarios could have occurred.

Scenario 1. The FR could not find the address because it never existed or no longer existed.
Scenario 2. When the FR went to the block, the FR could not find the unit because it existed

in a different block (geocoding error).

Scenario 3. After exhausting all resources, the FR still could not determine if the unit was allocated to the correct side of a tabulation boundary.

For the reasons listed in scenarios 1, 2, and 3, we considered all HUs with an "N" action code (about three percent of the HUs in the sample) to be out-of-scope, and we removed them from the analysis.

2.3 Variance estimation and statistical significance testing

2.3.1 Variance and standard error calculations

We used VPLX and the stratified jackknife method to calculate the variances and standard errors for the estimates. VPLX is a Fortran program developed by the Census Bureau to calculate variances for complex sample designs through replication. The stratified jackknife method calculates the variances and standard errors of the estimates by taking into account the stratification in the sample design.

2.3.2 Statistical significance testing

To determine statistical significance, we constructed 90 percent confidence intervals using the standard errors of the estimates and the critical value 1.645. If the confidence intervals contained zero or did not overlap, the estimates were significantly different at the α=.10 level. Otherwise, the estimates were not significantly different.

2.4 Applying quality assurance procedures

We applied quality assurance procedures throughout the creation of this report. They encompassed how we determined evaluation methods, created specifications for project procedures and software, designed and reviewed computer systems, developed clerical and computer procedures, analyzed data, and prepared this report.

3. LIMITS

3.1 Blocks excluded from the sampling universe

Based on the objectives of the evaluation and operational issues such as cost, we excluded 633,337 split blocks from the sampling universe. These blocks fell into one or more of the following categories[3]:

[3]The estimated number or percent of blocks that fell into each category is not available.

- Located in TEA 4 (Remote Alaska)
- Located in Puerto Rico or the Island Areas
- Contained no HUs or GQs
- Split by the boundaries of special-purpose governmental or administrative entities (such as school and voting districts) or statistical entities for tabulation purposes

By definition, split blocks that contain no HUs or GQs have no errors. We do not make inferences about blocks split by governmental boundaries in Remote Alaska or Puerto Rico. Furthermore, we do not attempt to make inferences about the boundaries of special-purpose and statistical entities that often are not visible on the ground or known by the residents.

3.2 Redefined boundaries

Actual governmental unit boundaries may have been legally changed between the time the GEO split the blocks for Census 2000 data tabulation and the time the FRs conducted field verification. For this reason, although we asked the FRs to identify boundaries defined as of January 1, 2000, this may have been difficult to operationalize. In some cases, results may reflect boundaries as they existed at the time of field verification.

3.3 Data by type of tabulation geography

Although this evaluation looks at blocks split for tabulation purposes, we do not present data by type of tabulation geography (state, county, MCD, incorporated places, and the like). The evaluation was not designed to produce estimates by type of tabulation geography.

3.4 Block statistics

This evaluation only measures errors in tabulation statistics for HUs and GQs that are in the correct block. The F.15 evaluation (Quality of Geocodes Evaluation) will measure errors related to allocating HUs and GQs to the wrong block. These two evaluations together will provide an assessment of all geographic errors in tabulation statistics.

4. RESULTS

In general, this evaluation measures the percent of the country that was affected by blocks split for tabulation purposes, and the accuracy of the processes used to allocate HUs and GQs in these split blocks to the correct side of a tabulation boundary. For this evaluation, the "country" refers to all areas of the U.S. **except** Puerto Rico, the Island Areas, Remote Alaska, and blocks without either HUs or GQs. Additionally, even though the numbers are presented for informational purposes, we removed all HUs coded "N" (could not locate or resolve) from the analysis for the reasons stated in Section 2.2.

Before answering the major questions of the evaluation, an overview of the distribution of split blocks based on TEA, representativeness of the selected sample, and the results of the field operation follow.

4.1 Distribution of split blocks based on TEA

As previously stated, TEA grouping was one of the variables we used to stratify the sampling universe. For each TEA, Table 1 gives the number and percent of split blocks in the sampling universe and blocks in the country.

Table 1. Distribution of blocks by TEA

TEA	Split blocks in the sampling universe		Blocks in the country	
	Number	Percent of total	Number	Percent of total
1	130,828	46.32	3,076,928	60.41
2	143,960	50.97	1,848,401	36.29
3	1,961	0.69	47,978	0.94
5	3,279	1.16	71,861	1.41
6	34	0.01	378	0.01
7	237	0.08	7,685	0.15
8	307	0.11	4,099	0.08
9	1,851	0.66	36,131	0.71
Total	**282,457**	**100.00**	**5,093,461**	**100.00**

As shown in Table 1, a little more than 97 percent of the split blocks in the sampling universe were in TEAs 1 and 2 alone. This percentage is consistent with the TEA percentages of blocks in the country - almost 97 percent. The TEA percentages for HUs and GQs similarly followed these results (see Appendix B).

4.2 Representativeness of the selected sample based on the estimated number of split blocks, HUs, and GQs

The sample slightly underestimated the number of split blocks and overestimated the number of HUs and GQs in the sampling universe. For this reason, we applied coverage factors to the blocks, HUs, and GQs to produce estimates that sum to the sampling universe. Table 2 gives the sampling universe, the number in the sample, the rate of selection, the sample estimate, and the coverage factor for the blocks, HUs, and GQs.

Table 2. Coverage factors for the sample estimates

Category	Sampling universe	In the sample	Rate of selection	Sample estimate[4]	Coverage factor[5]
Blocks	282,457	1,000	282.413	282,413	1.0002
HUs	12,219,396	43,470	282.413	12,276,493	0.9953
GQs	21,785	87	282.413	24,570	0.8867

Except for GQs, the coverage factors were very close to one. This means that the sample was more accurate in estimating the number of split blocks and HUs than GQs. This may be due to the limited occurrence of GQs in the split blocks, which makes it harder to estimate because the sample was selected based on such blocks.

Hereafter, all estimates presented and discussed in the tables and text are adjusted by the rate of selection and the applicable coverage factor.

4.3 Results of the field operation

When the FRs visited the sampled 1,000 split blocks, they determined whether the HUs and GQs in each block were allocated to the correct side of a tabulation boundary. If so, the address received a "V" action code. If not, the address was coded "C". All addresses that the FRs could not locate or resolve were coded "N".

Tables 3, 4, and 5 give the number of blocks, HUs, and GQs in the sample, and the estimated number and percent of these categories affected by block splitting based on the results of the field operation.

[4]The sample estimate is calculated by multiplying the number in the sample by the rate of selection.

[5]The coverage factor is calculated by dividing the sampling universe by the sample estimate.

Table 3. Results of the field operation - Blocks[6]

	Number in sample	Affected by block splitting	
		Number	**Percent**
Total	**1,000**	**282,457**	**100**
With action code of "V"	973	274,831	97.30
With action code of "C"	264	74,569	26.40
With action code of "N"	301	85,020	30.10

More than 97 percent, or 973, of the 1,000 sampled blocks had at least one HU or GQ that was verified to be allocated to the correct side of a tabulation boundary. Of the remaining 27 blocks, almost half contained only one HU or GQ. More than 26 percent of split blocks contained at least one HU allocated to the wrong side of a tabulation boundary. About 30 percent of the split blocks contained at least one HU or GQ that could not be located or resolved.

Table 4. Results of the field operation - HUs

	Number in sample	Affected by block splitting	
		Number	**Percent**
Total	**43,470**	**12,219,396**	**100**
With action code of "V"	40,528	11,392,401	93.23
With action code of "C"	1,534	431,207	3.53
With action code of "N"	1,408	395,788	3.24

Of the HUs in the sample, more than 93 percent were allocated to the correct side of a tabulation boundary. Additionally, about 3.5 percent were allocated to the wrong side of a tabulation boundary, and about 3.2 percent could not be located or resolved. The national estimate of erroneous enumerations in the census - 2.3 percent[7] - is on target with the HU results. The reasons for the erroneous enumerations in the census (geocoding error, not a HU, duplicates, and unresolved) were similar to the "could not locate" and "unresolved" reasons for this evaluation.

---

[6]These counts represent the number of blocks containing HUs or GQs or both coded either "V", "C", or "N". Therefore, the counts are not mutually exclusive and do not sum to the total.

[7]Hogan, Howard L., *Executive Steering Committee on Accuracy and Coverage Evaluation Policy II Report Number 17: Census 2000 Housing Unit Coverage Study*, Decennial Statistical Studies Division Census 2000 Procedures and Operations Memorandum Series #U-10revised, Bureau of the Census, October 17, 2001.

Table 5. Results of the field operation - GQs

		Affected by block splitting	
	Number in sample	Number	Percent
Total	**87**	**21,785**	**100**
With action code of "V"	67	16,777	77.01
With action code of "C"	0	0	0
With action code of "N"	20	5,008	22.99

All of the GQs that could be located (about 77 percent) were allocated to the correct side of a tabulation boundary. However, 23 percent of the GQs in the sample could not be located or resolved. Half of these GQs contained the same address for a marina in Florida, providing evidence that sampling variability yielding a small number of GQs in the sample is plausible.

None of the GQs in our sample of split blocks were allocated to the wrong side of a tabulation boundary. Although our estimate of the number of GQs allocated to the wrong side of a tabulation boundary was zero, we cannot conclude that there are no GQs in error throughout the country. The preliminary August 2002 results from one of our administrative review programs, Count Question Resolution (CQR)[8], showed that 1,867 GQs identified by local governmental entities were in fact allocated to the wrong side of a tabulation boundary.

We will now examine the following: how much of the country was affected by block splitting for tabulation purposes; the percent of the country in split blocks allocated to the wrong side of a tabulation boundary; and, the percent in error by TEA grouping, address type, and block split/misallocation status. For these calculations, we classified all HUs and GQs with an action code of "N" as out-of-scope, and we removed them from the data analysis.[9]

4.4 How much of the country was affected by block splitting for tabulation purposes?

Table 6 highlights the percent of blocks, HUs, and GQs affected by block splitting for tabulation purposes.

[8]The CQR Program is a planned administrative review program that handles external challenges to particular official Census 2000 counts of housing units and group quarters population received from state, local, or tribal officials of governmental entities or their designated representatives in the U.S. and Puerto Rico. The program occurs June 30, 2001 through September 30, 2003.

[9]Refer to Section 2.2 for a detailed explanation.

Table 6. Block splitting effects

Category	Number affected by block splitting	Number in the country	Out-of-scope cases ("Ns")	Number affected by block splitting (adjusted)[10]	Number in the country (adjusted)[11]	Percent of the country (adjusted)*
Blocks	282,457	5,093,461	NA	282,457	5,093,461	**5.55** **(0.03)**
HUs	12,219,396	115,877,639	395,788	11,823,608	115,481,851	**10.24** **(0.60)**
GQs	21,785	190,573	5,008	16,777	185,565	**9.04** **(2.50)**

* Standard errors are in parentheses.

As shown in Table 6, about five and one half percent of blocks in the country were split for tabulation purposes. A little more than ten percent of the HUs and about nine percent of the GQs in the country were in these split blocks. The percent of blocks in the country affected by block splitting for tabulation purposes is significantly different from that of the HUs. This is to be expected because split blocks tend to occur more often in urban areas where the average block size is larger than in more rural areas. The percent of GQs in the country affected by block splitting for tabulation purposes is comparable to the percent of blocks and HUs affected by block splitting for tabulation purposes.

4.5 What percent of the country was allocated to the wrong side of a tabulation boundary?

Table 7 highlights the percent of blocks, HUs, and GQs allocated to the wrong side of a tabulation boundary.

---

[10]The adjusted number affected by block splitting is calculated by subtracting the number of out-of scope cases from the number affected by block splitting. This is not applicable for blocks because entirely out-of-scope blocks are omitted from the universe.

[11]The adjusted number in the country is calculated by subtracting the number of out-of-scope cases from the number in the country.

Table 7. Percent of blocks, HUs, and GQs in error

Category	Estimated number in error ("Cs")	Affected by block splitting (adjusted)		In the country (adjusted)	
		Number	Percent in error (*)	Number	Percent in error (*)
Blocks	74,569	282,457	**26.40 (1.38)****	5,093,461	**1.46 (0.08)****
HUs	431,207	11,823,608	**3.65 (0.56)****	115,481,851	**0.37 (0.06)****
GQs	0	16,777	**0.0 (----)**	185,565	**0.0 (----)**

* Standard errors are in parentheses.
** Estimates are statistically significant at the α=.10 level.

Table 7 shows that 26.4 percent of split blocks contained at least one HU allocated to the wrong side of a tabulation boundary. These blocks represent 1.46 percent of the blocks in the country. The percent in error was lower for the HUs and GQs than for blocks. Of the HUs affected by block splitting, 3.65 percent were allocated to the wrong side of a tabulation boundary. These errors represent 0.37 percent of the HUs in the country. For the GQs in our sample, none were allocated to the wrong side of a tabulation boundary. Again, although our estimate of the number of GQs allocated to the wrong side of a tabulation boundary was zero, the preliminary August 2002 CQR results showed that 1,867 GQs in the country were allocated to the wrong side of a tabulation boundary.

As expected, the percent of blocks in error is significantly different from the percent of HUs in error. Because a block was defined to be in error if at least one HU or GQ in the block was allocated to the wrong side of a tabulation boundary, the proportion of blocks in error was significantly higher than the proportion of HUs in error. Also, the small number of blocks compared to the large number of HUs was a contributing factor as well.

Hereafter, all tables in this report present data at the HU level only.

4.6 Did the percent in error vary by TEA grouping?

We categorized TEAs into the following two groups:

- "Inside the blue line" consisting of TEAs 1, 6, 7, and 8; and
- "Outside the blue line" consisting of TEAs 2, 3, 5, and 9.

Table 8 shows that the percent in error of HUs affected by block splitting and the percent in error of HUs in the country were comparable by TEA grouping. It also shows that there were significantly more HUs affected "inside the blue line" than "outside the blue line." This is because there were significantly more HUs in TEA 1 than in the other TEAs combined (see Appendix B).

Table 8. Percent in error by TEA grouping

TEA grouping	Estimated number in error ("Cs")	Affected by block splitting (adjusted)		In the country (adjusted)	
		Number	Percent in error(*)	Number	Percent in error(*)
"Inside the blue line"	327,481	8,064,464	**4.06 (0.80)**	92,545,537	**0.35 (0.07)**
"Outside the blue line"	103,726	3,759,144	**2.76 (0.40)**	22,936,314	**0.45 (0.06)**

* Standard errors are in parentheses.
<Estimates are not statistically significant at the α=.10 level>

For the HUs affected by block splitting, the percent in error for TEAs "inside the blue line" was comparable to the percent in error for TEAs "outside the blue line" - about four percent compared to almost three percent, respectively. The same was true for HUs in the country. Both percentages were less than one half of one percent. Because the estimates are not statistically significant, it appears that the methodology the Census Bureau used to split blocks did not differentially affect the percent in error by TEA grouping.

4.7 Did the percent in error vary by address type?

In addition to TEA grouping, we categorized the addresses in the sample as one of two types: city-style or noncity-style. Like the TEAs "inside the blue line," significantly more city-style addresses than noncity-style addresses were affected because most addresses in TEA 1 were categorized as city-style. The percent in error of HUs affected by block splitting and the percent in error of HUs in the country were comparable by address type. Table 9 highlights the percent in error by address type.

Table 9. Percent in error by address type

Address type	Estimated number in error ("Cs")	Affected by block splitting (adjusted)		In the country (adjusted)	
		Number	Percent in error(*)	Number	Percent in error(*)
City-style	406,189	11,051,146	**3.68 (0.60)**	108,644,123	**0.37 (0.06)**
Noncity-style	25,018	772,461	**3.24 (0.80)**	6,837,728	**0.37 (0.08)**

* Standard errors are in parentheses.
<Estimates are not statistically significant at the α=.10 level>

Less than four percent of HUs with either city-style or noncity-style addresses affected by block splitting were allocated to the wrong side of a tabulation boundary. Additionally, the percent in error for HUs in the country for both address types was the same - 0.37 percent. Like TEA grouping, the estimates are not statistically significant. It appears that the methodology the

Census Bureau used to split blocks did not differentially affect the percent in error by address type.

4.8 Did the percent in error vary by block split/misallocation status?

As stated in Section 1.3.3, the addresses in the block split/misallocation operation were either "worked," "flagged to be worked but not worked," or "not flagged and not worked." Table 10 shows that the percent in error for the country varied by block split/misallocation status. Unlike the preceding tables, Table 10 does not give the number or percent in error for HUs affected by block splitting because the data were not available.

Table 10. Percent in error by block split/misallocation status

Block split/misallocation status	Estimated number in error ("Cs")	In the country (adjusted)	
		Number	Percent in error(*)
"Worked"	107,099	1,090,556	**9.82 (3.37)****
"Flagged but not worked"	80,957	3,963,672	**2.04 (0.33)****
"Not flagged/not worked"	243,151	110,427,623	**0.22 (0.05)****

* Standard errors are in parentheses.
** Estimates are statistically significant at the α=.10 level.

Nearly ten percent of the HUs that were "worked" in the block split/misallocation operation were allocated to the wrong side of a tabulation boundary. This compares to two percent of the HUs that were "flagged to be worked but not worked," and less than one half of one percent of the HUs that were "not flagged and not worked." All three estimates are significantly different from each other.

Limited resources available to plan, implement, and execute the block split/misallocation operation may have been a contributing factor to the percent in error of the "worked" addresses. It appears that the efforts the Census Bureau used for the worked addresses did not result in decreasing the percent in error to the level of the other two groups. Even so, it appears that the Census Bureau correctly differentiated the addresses and accurately flagged problematic areas of the country. We suspect that if the Census Bureau had not worked the problematic addresses in the block split/misallocation operation, the percent in error would have been higher.

4.9 What can we conclude?

A little more than ten percent of the HUs in the country were in blocks split for tabulation purposes. Less than one half of one percent of the HUs in the country were allocated to the wrong side of a tabulation boundary. Less than four percent of HUs in split blocks were allocated in error. The percent in error for HUs by TEA grouping was comparable. The percent in error by address type was comparable as well.

Unlike TEA grouping and address type, the percent in error by block split/misallocation status varied. Because the percent in error was very small for units "not flagged and not worked," it appears that the Census Bureau did a good job of flagging potential problems. However, the percent in error remained relatively high for cases that were "worked" in the block split/misallocation operation. More research is needed to fully understand why the percent in error for addresses "worked" was so high. We suspect that limited resources contributed to these errors. Ample time and sufficient staff to plan, implement, and execute the block split/misallocation operation, as well as adequate tools like the Global Positioning System (GPS) to determine the exact location of HUs and GQs may have decreased the percent in error to the level of the other two groups. Even so, we suggest more upfront design/planning for the 2010 Census to enhance the success of this operation.

References

Angueira, Teresa, *Program Master Plan: Census 2000 Block Split/Misallocation Operation*, Census 2000 Informational Memorandum No. 115, Bureau of the Census, November 14, 2001.

"Census 2000 Geographic Handbook, Appendix C - Glossary of Terms Related to Field and Geographic Operations," *Field Intranet at the Bureau of the Census*, n.d., <http://gsacbweb.fld.census.gov> (January 4, 2002).

"Census 2000 Type of Enumeration Areas (TEAs)," *Geography Intranet at the Bureau of the Census*, n.d., <http://www.geo.census.gov/mob/homep/teas.html> (November 30, 2001).

"Count Question Resolution (CQR) Fact Sheet," *U.S. Census Bureau website*, January 28, 2002, <http://www.census.gov/dmd/www/cqrfsheet.htm> (March 27, 2002).

Fay, Robert E., "VPLX Handbook," *Statistical Research Division Intranet at the Bureau of the Census*, n.d., <http://vplx.srd.census.gov/index2.html> (March 25, 2002).

Fifield, Scott P. <scott.p.fifield@census.gov>, "Source for F.16 Results," February 21, 2002, office communication.

"Geographic Areas Reference Manual," *Geography Intranet at the Bureau of the Census*, n.d., <http://www.census.gov/geo/www/garm.html> (March 6, 2002).

Hogan, Howard L., *Executive Steering Committee on Accuracy and Coverage Evaluation Policy II Report Number 17: Census 2000 Housing Unit Coverage Study*, Decennial Statistical Studies Division Census 2000 Procedures and Operations Memorandum Series #U-10revised 2, Bureau of the Census, October 17, 2001.

Killion, Ruth Ann, *Specification for Producing the File of Blocks Split for Tabulation Purposes*, Planning, Research, and Evaluation Division TXE/2010 Memorandum Series: MAF-TAB-S-01, Bureau of the Census, March 2, 2001.

Killion, Ruth Ann, *Specifications for Producing the Analysis File and Counts for the Evaluation of Block Splitting Operation for Tabulation Purposes*, Planning, Research, and Evaluation Division TXE/2010 Memorandum Series: MAF-TAB-S-2, Bureau of the Census, March 6, 2002.

Killion, Ruth Ann, *Study Plan for Evaluation of the Block Splitting Operation for Tabulation Purposes - F.16*, Planning, Research, and Evaluation Division TXE/2010 Memorandum Series: MAF-STP-O-07, Bureau of the Census, April 6, 2001.

Pees, Elva M. <elva.marie.pees@census.gov>, "CQR Info Relating to GQs," September 10, 2002, office communication.

Ramirez, Lourdes <lourdes.ramirez@census.gov>, "Number of Split Blocks," March 29, 2001, office communication.

Tomassoni, Robert G., "Block Split Evaluation Field Lister's Manual," *Field Intranet at the Bureau of the Census*, n.d., <http://gsacbweb.fld.census.gov> (January 28,2002).

Tomassoni, Robert G. <robert.g.tomassoni@census.gov>, " BSE Dates," October 9, 2001, office communication.

Vitrano, Frank A., *Overview of the Quality of the Geocodes Evaluation*, Planning, Research, and Evaluation Division TXE/2010 Memorandum Series: MAF-GC-O-01-R1, Bureau of the Census, October 22, 2001.

Vitrano, Frank A., *Sample Design for the Evaluation of the Block Splitting Operation for Tabulation Purposes (F.16)*, Planning, Research, and Evaluation Division TXE/2010 Memorandum Series: MAF-TAB-G-01, Bureau of the Census, April 5, 2001.

Appendix A: States in the Northeast and Midwest Census Regions

FIPS State Code	State Name
09	Connecticut
17	Illinois
18	Indiana
19	Iowa
20	Kansas
23	Maine
25	Massachusetts
26	Michigan
27	Minnesota
29	Missouri
31	Nebraska
33	New Hampshire
34	New Jersey
36	New York
38	North Dakota
39	Ohio
42	Pennsylvania
44	Rhode Island
46	South Dakota
50	Vermont
55	Wisconsin

Appendix B: Distribution of Housing Units and Group Quarters by TEA

Table B-1. Distribution of HUs by TEA

TEA	HUs in split blocks in the sampling universe		HUs in the country	
	Number	Percent of total	Number	Percent of total
1	8,363,723	68.45	92,451,759	79.78
2	3,645,052	29.83	21,335,678	18.41
3	39,471	0.32	392,235	0.34
5	88,522	0.72	886,231	0.76
6	12,440	0.10	50,656	0.04
7	12,598	0.10	238,216	0.21
8	10,894	0.09	69,983	0.06
9	46,696	0.38	452,881	0.39
Total	**12,219,396**	**100.00***	**115,877,639**	**100.00***

*TEA percentages may not sum to 100 due to rounding.

Table B-2. Distribution of GQs by TEA

TEA	GQs in split blocks in the sampling universe		GQs in the country	
	Number	Percent of total	Number	Percent of total
1	14,386	66.04	153,460	80.53
2	6,776	31.10	32,840	17.23
3	82	0.38	839	0.44
5	105	0.48	1,090	0.57
6	297	1.36	1,197	0.63
7	45	0.21	492	0.26
8	32	0.15	102	0.05
9	62	0.28	553	0.29
Total	**21,785**	**100.00**	**190,573**	**100.00**