March 28, 2012

2012 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES #ACS12-RER-17

DSSD 2012 American Community Survey Memo Series #ACS12-UC-04

MEMORANDUM FOR   ACS Research and Evaluation Steering Committee

From: David C. Whitford */Signed/*
                      Chief, Decennial Statistical Studies Division

Prepared by: Dianne Aubuchon
                      Community Address Updating Systems Branch
                      Decennial Statistical Studies Division

Subject: American Community Survey/Census Group Quarters Match

Attached is the final American Community Survey Research and Evaluation report for the American Community Survey/Census Group Quarters Match. The Decennial Group Quarters Validation operation resulted in a much more complete list of Group Quarters. There was some concern that the Group Quarters that the American Community Survey has added since the 2000 Census would now be duplicates of Group Quarters found in Group Quarters Validation, thus we performed a match between American Community Survey and Census Group Quarters lists to determine if the added Group Quarters should remain eligible.

If you have any questions about this report, please contact Jim Hartman (301) 763-1976 or Dianne Aubuchon (301) 763-7706.

Attachment

cc:

ACS Research and Evaluation Team   Matt Zimolzak (ACSO)
Tony Tersine (DSSD)   Judy Belton
Jim Hartman   Marcella Jones-Puthoff (POP)
Mark Asiala
Steve Hefter

American Community Survey/Census Group Quarters Match

FINAL REPORT

Dianne Aubuchon
Decennial Statistical Studies Division

1. **Executive Summary**

   The American Community Survey has been supplementing the Master Address File/Topographically Integrated Geographic Encoding and Referencing database with new Group Quarters as there has been no continuous updating of Group Quarters since the 2000 Census. When 2010 Census operations began to update Group Quarters, we wanted to make sure that the American Community Survey did not continue to supplement Group Quarters that had now been added by these operations. We decided to match the American Community Survey and 2010 Census Group Quarters lists.

   Just over 77 percent of the American Community Survey eligible Group Quarters not on the Master Address File/Topologically Integrated Geographic Encoding and Referencing database were matched to records on the Master Address File. There were 34 unmatched Group Quarters that will continue to be eligible for the American Community Survey. These will be included due to their size and the impact on the sample if they were to no longer be represented.

   The remaining unmatched Group Quarters will no longer be eligible as we felt the risk of duplication for these Group Quarters was too high.

2. **Introduction**

   The post-2000 universe of group quarters (GQs) in the Master Address File (MAF)/Topologically Integrated Geographic Encoding and Referencing (TIGER) database (MTdb) is deficient because there are no ongoing or complete GQ updates to the MTdb. The American Community Survey (ACS) attempted to supplement MTdb GQs with new GQs from other sources[1] in an effort to improve coverage of GQs. Those new GQs were included in the ACS GQ frame but were not added to the MTdb. Migrant worker camps and some military facilities were updated, although not consistently. Only some prisons were updated consistently during this time period, although updating was limited to available staff time.

   The Decennial Group Quarters Validation (GQV) and Address Canvassing (AdCan) operations worked to update the MAF through field verification. Listers went nationwide, with the exception of a few areas in Alaska and Maine, to update addresses and GQs in order to have the best lists possible for the 2010 Census. (Program Management Branch 2010)

   As a result of GQV, the MTdb will have a much more complete list of GQs. Those GQs confirmed by GQV will comprise the 2010 GQ Enumeration (GQE) universe. Ideally, the GQE universe will be complete and the 2011 ACS GQ frame could be the same as the GQE universe.

---

[1] This includes Address Problem Referral, other files on state prisons, military GQs, migrant worker camps, and GQs closed on Census day. (U.S. Census Bureau 2009)

However, since the initial Census address lists did not contain GQs that ACS added throughout the decade, there is some concern that these GQs may have been missed by GQV. If GQs added by ACS were missed by GQV then using the GQE universe would result in undercoverage. However, adding all the ACS GQs to the GQE universe would likely result in some duplication since GQV probably picked up most of the GQs that ACS added. ACS must decide what to do with the GQs that were included in the ACS frame that do not exist in the MTdb after GQV.

We do not want to just include all of these GQs as they may have been found in GQV but not linked together. Our research will address the question of whether there are GQs currently eligible for ACS that are not included in the Census GQE universe. And if so, should they continue to be eligible for ACS?

To answer this question we matched the ACS-eligible GQs from the ACS GQ frame that were not on the MTdb, to the January 2010 ACS MTdb extract (this extract includes the GQs in the GQE universe). There were about 5,000 such GQs.

3. **Methodology**

The match consisted of three phases. There was an initial computer match, followed by a clerical match and then a final computer match. The data universes and procedures are described below. In addition there was a match of "large"[2] GQs that took place after the final computer matching as concerns were raised over these particular GQs; details are described below.

3.1 Data Universes

The universe of ACS records that we matched included all ACS GQs that were not on the MTdb as of the January 2010 ACS MAF extracts. The universe of MTdb records used for the initial computer matching contained records that were categorized as a GQ and as being in the GQE universe. The clerical matching universe contained any unmatched MTdb records from the computer matching universe, any other GQs not considered to be a special place (SP)[3] on the MTdb, and any other record on the MTdb with a GQV action code not already included[4]. Initially, any Census record in the same county was included as a potential match for each ACS record. This was then narrowed by various flags to any Census record that had a "keyword" in common with the ACS record in the GQ name, facility name, location description, or street name fields. For the final matching phase, we included all MTdb records that were visited by either AdCan or GQV (Housing Units (HUs) and GQs).

---

[2] A GQ was designated as large by a state-level threshold of expected population defined by the ACS Sample Design Branch, Decennial Statistical Studies Division.
[3] Special places are facility level records that were created for Census 2000 in order to link GQs that were within the same facility.
[4] This includes any housing units that were identified in GQV.

3.2 Preparation of the Matching Data

Prior to any matching, keywords were chosen from the ACS GQ name and street name to narrow down the number of potential matches for the clerical matching phase. Up to three main words from each of these were then matched to Census records to produce potential matches. The potential matches were then flagged for where the keyword was found, creating the four keyword flags mentioned above.

In addition, multiple scores were created to help determine which Census records were really potential matches, and which were not. This included scores comparing addresses, GQ names, SP names, and partial names. Scores were created using the compged function in SAS®[5]. Points were added to the score based on the number and type of differences found between the names being compared. So a lower score should indicate a better potential match.

A number of flags were also created to help determine potential matches. Flags to tell us if records had matched on house number, street name, with-in structure components, ZIP codes, and blocks were created. These flags, along with the scores and keywords were used to determine the final lists of potential matches.

Two partial names were created (GQ and SP) by finding the shortest name among the potential matches and truncating all names to that length. These were also used to whittle down the number of potential matches. This was done because we found that often the beginnings of names would match, but that there would be slight differences near the end of the names.

3.3 Initial Computer Matching

After the flags and scores were created we started the computer matching. The first pass of computer matching was based on address and GQ name. In the first step of pass one, if an ACS record had only one potential match that matched exactly on address and GQ name then it was considered a computer match. In the second step, an ACS record that only had one potential match that had an exact address match and an exact partial GQ name match was considered a computer match. These two steps were repeated for any still unmatched records using the SP name in place of the GQ name for the ACS record.

3.4 Clerical Matching

Once the computer matching was complete, we moved on to the clerical matching. We expanded the MTdb universe to also include GQs that were not in the enumeration universe, transitory locations, transitory units within transitory locations, and anything that

---

[5] For more information on the compged function:
http://support.sas.com/onlinedoc/913/getDoc/en/lrdict.hlp/a002206133.htm

was visited in GQV including records that GQV determined to be HUs. We used the various flags and scores to narrow down the number of potential matches, in order to list only the most promising potential matches. The first criteria required the potential match to have a partial GQ or SP name score less than 1000. The second criteria required one of the following to be true:

- The house number, ZIP code, block, or street name was flagged as being an exact match between the ACS and MTdb records
- The MTdb record was flagged as having a keyword in common with the ACS record

3.5 Final Computer Matching Phase

After clerical matching, for any ACS GQ records still unmatched, we looked to see whether the address had been visited by either ADCAN or GQV. We did this by creating a universe of all MTdb records visited by ADCAN and/or GQV. This was computer matched to the unmatched ACS records by county, house number, and street name. Any ACS record that matched exactly on these criteria to an MTdb record was flagged as having been visited. For these records, we will assume that an ADCAN or GQV lister would have found any GQ at that address therefore the ACS GQ no longer exists.

3.6 Large GQ matching

There was some concern that certain "large" GQs that were unmatched would no longer be in the ACS universe. If the unmatched large GQs do exist, and are not in the GQE universe, then a good number of people would no longer be represented in sample. Additionally, these GQs are in sample every year, and ACS interviews are obtained. As we are confident they do exist, we do not want to exclude them just because we could not find them in the Census GQE universe. But we also do not want to include them if Census did find them. So another more extensive phase of matching was done for these GQs.

Two rounds of clerical matching were included for 248 large ACS GQs that remained unmatched. MAF records were chosen as potential matches if:

- they were in the same tract as the ACS GQ; or

- the facility name, GQ name, location description, or street name contained strings that were also in the ACS record[6]

These rounds of matching also allowed for Internet research into finding alternative names for a GQ, or to find information indicating that the GQ had been demolished, closed, or no longer exists.

---

[6] These did not have to be in the same county.

4. **Limitations**

    - The quality of the data is not the best. Spelling errors could exclude a match in computer matching[7] or, if more severe, it could exclude a record from being considered a potential match for clerical matching as well. If the error only affected one of the matching criteria, then it would probably be resolved in clerical matching. However, if a spelling mistake affected a keyword there was a possibility that the record would never have been marked as a potential match.

        Shorthand differed between the different Field Representatives, which made matching more difficult. Much of this could be caught in clerical matching, but it could have excluded records from becoming potential matches.

        We also had many cases where descriptions and addresses were all similar, which limited our ability to determine which MTdb record matched to the ACS GQ.

    - Time – the matching needed to be completed by a certain date so that the results could be used to form the 2011 ACS GQ universe. This limited the number of potential matches we could look through for each record. In these cases we did not match, but felt that the GQ was on the Census list.

5. **Results**

    Of the records that went through the computer, clerical, and final matching steps, almost 75 percent were either computer matched, clerically matched, or flagged as having been visited by ADCAN or GQV. (Table 1)

    Table 1. Match Status of ACS GQs

Status	Frequency	Percent of Total
Computer Matched	217	4.5%
Clerically Matched	1,716	35.5%
Visited by AdCan or GQV	1,619	33.5%
Unmatched	1,283	26.5%
Total	4,835	100%

    Source: U.S. Census Bureau, American Community Survey 2010 Supplemental Phase Unit Frame Universe and 2010 American Community Survey Group Quarters Universe

    We were unable to match, or flag as visited, just over a quarter of the records. The question we are seeking to answer is whether any of those unmatched GQs should be eligible for ACS. To see what types of GQs the unmatched records were, we looked at them by GQ source and major GQ

    ---

    [7] We did not use a record linkage package, opting for a simpler method of matching.

type. GQs with a source of either "Closed on Census Day" or "Migrant Worker Camps" were almost 80 percent of the unmatched records. (Table 2) The distribution of GQ types is consistent with this as almost 65 percent of GQs have a major GQ type of 'Worker GQs and other'. (Table 3)

Table 2. Unmatched ACS GQs by GQ source

Source	Frequency	Percent of Total
Address Problem Referral Research	66	5.1%
Closed on Census Day	614	47.9%
Migrant Worker Camps	409	31.9%
Military GQs	169	13.2%
State Prisons	25	2.0%

Source: U.S. Census Bureau, American Community Survey 2010 Group Quarters Universe

As noted in Table 2, almost half of the unmatched ACS GQs had a "Closed on Census Day" source. Since Census had not picked them up due to their being closed on Census day there was some concern that GQV may have missed them as well. However, the GQV operation took place well in advance of Census day 2010 and had much less stringent procedures for categorizing something as a GQ. Because of this, we are fairly confident that GQV would have picked up the Closed on Census Day GQs.

Table 3. Unmatched ACS GQs by GQ type

Status	Frequency	Percent of Total
Correctional Institutions	52	4.0%
Juvenile Institutions	18	1.4%
Nursing Homes	20	1.6%
Hospitals, Wards, Hospices	15	1.2%
College/University Student Housing	16	1.3%
Military Quarters	164	12.8%
Shelters/Service Locations	156	12.2%
Group Homes	9	0.7%
Worker GQs and other	833	64.9%

Source: U.S. Census Bureau, American Community Survey 2010 Group Quarters Universe

For unmatched records that had been sent out for an ACS interview, we looked at the status at time of interview. Of the 1,283 unmatched cases, 182 had been selected for interview. Of those sent out, 88 percent had a valid status at time of interview. (Table 4)

Table 4. Unmatched GQs time of interview status

Status	Frequency	Percent of Total
Valid	161	88.5%
Invalid	21	11.5%

Source: U.S. Census Bureau, 2006-2009 American Community Survey Group Quarters Facility Questionnaire results files

Given that ACS had found and validated 88 percent of the unmatched records that had previously been in sample, we decided that Census operations should have found the majority of these cases as well. For many Census records, the GQ name recorded was not specific enough. For military quarters we often came upon records where the GQ name was simply 'military barrack', which made it virtually impossible to match to an ACS record. A similar situation was encountered with migrant worker camps where we would have GQ names of dorm 1, dorm 2, dorm C, etc. Because of these difficulties we believe that GQV captured these records and that we were unable to match to them.

After the large GQ matching, only 34 of the 248 large GQs remained unmatched.

6. Summary/Conclusions

Between all matching operations we were able to match just over 77 percent of the ACS-eligible GQs not on the MTdb to records on the MAF. Of the GQs that remained unmatched, we decided that those GQs that fell into the category of "large GQ" will continue to be eligible for ACS. These GQs have been in sample every year for ACS. We did a more exhaustive matching process for these GQs so we feel sure they are not on the MTdb, and we obtain interviews for them so we are confident that they do exist.

The remaining unmatched GQs are considered to be "small GQs". The lack in specificity of the Census GQ names is one reason that we were unable to match the small GQs. We decided that, if we included the unmatched small GQs the risk of duplicate GQs was too high so they will be ineligible.

7. References

Program Management Branch, 2010, "2010 Census Operational Plan" U.S. Census Bureau, June 23, 2010. Pages 67, 73.

U.S. Census Bureau, 2009, Design and Methodology American Community Survey, U.S. Government Printing Office, Washington, D.C. Chapter 3. (http://www.census.gov/acs/www/Downloads/survey_methodology/acs_design_methodology_ch03.pdf)