UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001

September 12, 2012

2010 CENSUS PLANNING MEMORANDA SERIES

No. 237

MEMORANDUM FOR The Distribution List

From: Burton Reist *[signed]*
Acting Chief, Decennial Management Division

Subject: 2010 Census Global Positioning System (GPS) Evaluation Report

Attached is the 2010 Census Global Positioning System (GPS) Evaluation Report. The Quality Process for the 2010 Census Test Evaluations, Experiments, and Assessments was applied to the methodology development and review process. The report is sound and appropriate for completeness and accuracy.

If you have any questions about this document, please contact Ryan Cecchi at (301) 763-0301 or RJ Marquette at (301) 763-2987.

Attachment

2010 Census Program for Evaluations and Experiments
September 12, 2012

2010 Census: Global Positioning System Evaluation

U.S. Census Bureau standards and quality process procedures were applied throughout the creation of this report.

FINAL REPORT

Ryan Cecchi and RJ Marquette
Decennial Statistical Studies Division

This page intentionally left blank.

TABLE OF CONTENTS

LIST OF TABLES

Executive Summary

Introduction

The 2010 Census Address Canvassing operation was implemented to improve the completeness and accuracy of the initial census address list. The U.S. Census Bureau then used the updated addresses to mail or deliver census questionnaires. To conduct Address Canvassing, the universe was divided into collection blocks, which were then grouped into Assignment Areas for operational convenience.

The original listers used hand-held computers to conduct the verification of the address list and census maps. The hand-held computers offered several advantages over paper listing operations, one being the ability to collect Global Positioning System coordinates for living quarters.

The Address Canvassing Dependent Quality Control check was designed to detect significant errors and provide an evaluation of each work assignment. Every Assignment Area was checked in the Dependent Quality Control. The quality control lister's task during the Dependent Quality Control check was to compare what was found on the ground to what was on the address list.

For each Assignment Area, a quality control lister was assigned to check a sample of addresses, beginning with an address randomly selected by computer software. The quality control lister would begin with the randomly selected address and work from the ground to the address list, ensuring what was found on the ground was reflected on the address list. The sample size was based on the number of addresses in the Assignment Area, and the software notified the quality control lister when enough addresses had been checked. It is important to understand that the quality control sample was designed to be representative of the entire Assignment Area. If the quality control sample contained errors, then it was assumed that there was about the same rate of errors in the remainder of the Assignment Area. Consequently, if the quality control sample failed, then the entire Assignment Area failed and was recanvassed.

These Dependent Quality Control check procedures are used to identify, deter, and correct errors and data falsification; however the quality control lister can only identify poor performance on the part of the original lister if the sampled group of structures coincides with an area where the original lister made a faulty update to the address list. In past listing operations, the quality control focused on the validation of listing data to identify procedural errors and/or data falsification, however Global Positioning System data can be used as a tool to detect and deter procedural violations, which may or may not lead to poor data.

Global Positioning System technology was first incorporated into the Address Canvassing operation with the introduction of hand-held computers during the 2006 Census Test. As listers manually recorded map spots for structures, the Global Positioning System technology simultaneously collected longitude and latitude coordinates for nearly all structures visited by field listers. It should be noted that the Global Positioning System coordinates were only available where/when the Global Positioning System signal could be received. In addition, Address Canvassing procedures specified that listers were to only record map spots for one unit in multi-unit structures due to limitations in the software capabilities. Software on the hand-held computer stored the Global Positioning System coordinates of the physical location as well as the manual map spot identified by the lister. Thus, the two sets of coordinates provided a simple, but previously unexplored, method to check the manual map spot coordinates recorded by listers in the field.

A review of data using hand-held computers during both the 2006 Census Test and 2008 Census Dress Rehearsal has shown instances where the Global Positioning System coordinates did not correspond well with the coordinates of the manual map spots recorded by listers. In many cases, a series of these discrepancies were clustered together in a small area, which suggests that listers were recording manual map spots for structures without actually canvassing the entire area and visiting every structure. Current quality control methods do not compare the Global Positioning System and manual map spot coordinates and have a limited probability of identifying these types of procedural violations[1]. Both the distance between a manual map spot and Global Positioning System map spot (referred to as a "strand length") and clustering of Global Positioning System map spots in a relatively small area (referred to as a "curbstoning cluster") could potentially be used to identify procedural violations that could negatively impact the quality of Address Canvassing data. If clustered discrepancies between the two types of map spots were the result of listers who recorded map spots without canvassing an entire area, then that would represent a violation of canvassing procedures and/or data falsification. Thus, comparison of Global Positioning System and manual coordinates could represent an important new technique for the quality control of Address Canvassing field work.

<u>Results</u>

Assignment Areas containing one or more curbstoning clusters failed the Dependent Quality Control check at a rate of 13.70 percent, compared to the overall Dependent Quality Control check fail rate of 8.43 percent[2]. This indicates that a curbstoning cluster can be used as a predictor of falsification and/or procedural violations. As long as the automated instrument can compute the Global Positioning System data and flag Assignment Areas with one or more curbstoning clusters when completed by the listers, this method could be used for detection of procedural violations and prediction of data falsification.

Assignment Areas flagged as having an average strand length significantly greater than the overall average strand failed the Dependent Quality Control check at a rate of 14.62 percent, compared to 8.29 percent for Assignment Areas not flagged.

Similarly, Assignment Areas flagged as having an average strand length significantly greater than the 90th percentile of overall strand lengths had a Dependent Quality Control check fail rate of 12.56 percent, compared to 8.19 percent for Assignment Areas not flagged. Both of these

[1] The quality control procedures for Address Canvassing involve confirming the *manual* location of the mapspot and recording a noncritical error if the unit was incorrectly placed. There was no review of the Global Positioning System coordinates, so there was no way for the quality control listers to determine whether the lister actually visited each house and collected coordinates per the procedures.

[2] Again, it is important to keep in mind that not all Assignment Areas that failed Dependent Quality Control check were a result of falsification and/or procedural violations. Assignment Areas could also fail the Dependent Quality Control check because of honest unintentional mistakes by listers.

tests indicate that average strand length can be used as a predictor of falsification and/or procedural violations.

As long as the automated instrument can compute the Global Positioning System data and flag Assignment Areas that have a significantly higher average strand length than some parameter when completed by the listers, strand length can be used for automated detection of data falsification and/or procedural violations.

It does not appear that one approach is better than the other. Both approaches reveal an increase in the Dependent Quality Control check fail rate for Assignment Areas flagged during each approach. The combination of the curbstoning cluster and strand length tests show that the two tests were not flagging a large portion of the same Assignment Areas. For example, the average overall strand length test (35.68 meters) also caught 30.30 percent of the Assignment Areas with curbstoning clusters and the 90^{th} percentile overall strand length test (13.68 meters) caught 44.98 percent of the Assignment Areas with curbstoning clusters.

Recommendations

The following recommendations would improve future listing quality control operations:

Global Positioning System data for manual map spot coordinates and actual lister location should be collected for each housing unit added or verified by the lister in order to be used as a tool for quality control.

Determine if (and how) the use of Global Positioning System coordinates could be used for units deleted by the lister.

The listing instrument should flag Assignment Areas that contain one or more curbstoning clusters, along with Assignment Areas that have an average strand length significantly greater than some pre-defined parameter.

The value for the strand length test parameter should be studied and determined for the 2020 Census. Several options include using 2010 Address Canvassing overall average strand length, 2010 Address Canvassing 90^{th} percentile overall strand length, or the process average during live production.

The Assignment Areas flagged by either method should undergo a more stringent quality control to determine if data falsification and/or procedural violations occurred. This can be done by a variety of options, including increasing the size of the quality control sample string or verifying multiple quality control sample strings within the Assignment Area.

Since the curbstoning cluster method and strand length method did not flag a large portion of the same Assignment Areas, both methods should be considered/implemented, rather than one or the other.

The listers responsible for the flagged Assignment Areas should be monitored in some way by the managers. Counts of flagged assignments may help managers correct procedural violations, whether by additional training or termination.

Global Positioning System coordinates should be used as a preventable measure during the listing operations, such as a pop-up warning on the instrument if the lister appears to be too far from the unit the lister is mapspotting.

Global Positioning System coordinates comparisons need to be real-time in order for the pop-up warning to be a viable measure. If Global Positioning System data are only used for quality control purposes, in order to be beneficial, the coordinates comparisons need to be completed before being worked in quality control.

Any quality control methods using the Global Positioning System data should work as a supplement to the traditional quality control methods.

1. INTRODUCTION

1.1 *Scope*

This report presents the results of the evaluation of the Global Positioning System (GPS) and how it could be used for future listing quality control (QC) operations. The main objective of this study is to evaluate the potential of incorporating GPS data collected from hand-held computers into the quality control of field work.

To support this objective, the evaluation will analyze curbstoning clusters and strand lengths (defined in the Methods section) to see how well these serve as predictors of data falsification or procedural violation.

1.2 *Intended Audience*

The intended audience of this report includes program managers and staff responsible for planning future listing QC operations.

2. BACKGROUND

2.1 *Address Canvassing Operation*

The 2010 Address Canvassing (AC) operation was implemented to improve the completeness and accuracy of the initial census address list. The U.S. Census Bureau then used the updated addresses to mail or deliver census questionnaires. To conduct AC, the AC universe was divided into collection blocks, which were then grouped into Assignment Areas (AAs) for operational convenience.

The original listers used hand-held computers (HHCs) to conduct the verification of the address list and decennial census maps. The HHCs offered several advantages over paper listing operations, one being the ability to collect GPS coordinates for living quarters.

2.2 *Dependent Quality Control Check*

The AC Dependent Quality Control (DQC) check was designed to detect significant errors and provide an evaluation of each work assignment. Every AA was checked in the DQC. The QC lister's task during the DQC was to compare what was found on the ground to what was on the address list.

For each AA, a QC lister was assigned to check a sample of addresses beginning with an address randomly selected by computer software. The QC lister would begin with the randomly selected address and work from the ground to the address list, ensuring what was found on the ground was reflected on the address list. The sample size was based on the number of addresses in the AA, and the software notified the QC lister when enough addresses had been checked. It is important to understand that the QC sample was designed to be representative of the entire AA. If the QC sample contained errors, then it was assumed that there was about the same rate of errors in the remainder of the AA. Consequently, if the QC sample failed, then the entire AA failed and was recanvassed.

For detailed information about the AC QA program, please see 2010 Census Planning Memoranda Series No. 184, "2010 Census: Address Canvassing Quality Profile" (Marquette, 2012).

These DQC procedures are used to identify, deter, and correct errors and data falsification; however the QC lister can only identify poor performance on the part of the original lister if the sampled group of structures coincides with an area where the original lister made a faulty update to the address list. In past listing operations, the QC focused on the validation of listing data to identify procedural errors and/or data falsification, however GPS can be used as a tool to detect and deter procedural violations, which may or may not lead to poor data.

2.3 *Global Positioning System Technology in Address Canvassing Operations*

GPS technology was first incorporated into the AC operation with the introduction of HHCs during the 2006 Census Test. As listers manually recorded map spots for structures, the GPS technology simultaneously collected longitude and latitude coordinates for nearly all structures visited by field listers. It should be noted that the GPS coordinates were only available where/when the GPS signal could be received. In addition, AC procedures specified that listers were to only record map spots for one unit in multi-unit structures due to limitations in the software capabilities. Software on the HHC stored the GPS coordinates of the physical location as well as the manual map spot identified by the lister. Thus, the two sets of coordinates provided a simple, but previously unexplored, method to check the manual map spot coordinates recorded by listers in the field.

A review of data using HHCs during both the 2006 Census Test and 2008 Census Dress Rehearsal has shown instances where the GPS coordinates did not correspond well with the coordinates of the manual map spots recorded by listers. In many cases, a series of these discrepancies were clustered together in a small area, which suggests that listers were recording manual map spots for structures without actually canvassing the entire area and visiting every structure. Current QC methods do not compare the GPS and manual map spot coordinates and have a limited probability of identifying these types of procedural violations[3]. Both the distance between a manual map spot and GPS map spot (referred to as a "strand length") and clustering of GPS map spots in a relatively small area (referred to as a "curbstoning cluster") could potentially be used to identify procedural violations that could negatively impact the quality of AC data. If clustered discrepancies between the two types of map spots were the result of listers who recorded map spots without canvassing an entire area, then that would represent a violation of

[3] The QC procedures for AC involve confirming the *manual* location of the mapspot and recording a noncritical error if the unit was incorrectly placed. There was no review of the GPS coordinates, so there was no way for the QC listers to determine whether the lister actually visited each house and collected coordinates per the procedures.

canvassing procedures and/or data falsification. Thus, comparison of GPS and manual coordinates could represent an important new technique for the QC of AC field work.

3. METHODOLOGY

The research questions for the evaluation of the GPS will be answered through analysis of two files delivered to the Decennial Statistical Studies Division (DSSD):

- A file, provided by the Geography Division (GEO), containing 1) all housing units with a GPS strand length greater than 50 meters or 2) all housing units within a defined curbstoning cluster (see section 3.2.1, below).
- A file, provided by the Field Data Collection Automation (FDCA) contractor, containing address information for all housing units in the AC operation.
- A file, provided by the FDCA contractor, containing AA information, which includes DQC results.

3.1 *Defining a Curbstoning Cluster*

Clustered GPS map spots (i.e., curbstoning clusters) could be a predictor of poor lister performance due to procedural violations or data falsification. The main goal of this evaluation is to explore the potential of curbstoning clusters as a quality indicator.

Ideally, the manual map spots recorded by listers during address canvassing should directly, or very closely, coincide with the GPS map spots recorded by the hand-held computer. Significant discrepancies between the two types of map spots could indicate violations of canvassing procedures or even data falsification. However, exact definitions must be established to distinguish between divergences that occur as a result of GPS resolution issues[4] and those due to poor lister performance.

The following requirements were set to establish a working definition for curbstoning clusters:

- Multi-unit structures were ignored.

  The distances between manual and GPS map spots in multi-unit structures would not exceed the minimum cluster threshold in most instances. Clustering of GPS map spots for multi-unit structures could be an indicator of poor lister performance. However, although listers were instructed to only collect one map spot for a multi-unit structure, the software did not prevent them from collecting multiple points for the same multi-unit structure. It would be more difficult to distinguish between legitimate and suspicious curbstoning clusters for multi-unit compared with single-

[4] Due to atmospheric interference and other factors, the GPS coordinates collected by the HHC can be several meters or more from the correct location on the ground. In addition, confusing situations can arise for the listers; for example, if a road is not placed correctly on the map, the GPS may appear to be giving an incorrect location when in fact it is the map that is wrong.

unit structures. Thus, the focus is on single-unit structure curbstoning clusters for this study; examination of multi-unit curbstoning clusters could be a topic for a follow-up study.

- A minimum distance between GPS map spots was defined for a cluster to qualify as a curbstoning cluster.

- Townhouses, row houses, brownstones, etc. would be the densest structures listers would typically encounter without being multi-unit structures.

- In addition, a cluster size, or minimum number of structures must be defined for a curbstoning cluster. For the purposes of this evaluation, this number was set to a square with 0.0001 degrees on each side that contained six or more housing units. Based on these specifications, the following procedure was used to detect curbstoning clusters in the 2010 AC data:

    1. Divided the country into squares of length 0.0001 degrees on each side (the exact distance varied, but each side was approximately 40 feet or 12.2 meters).
    2. Counted the number of GPS map spots within each square.
    3. If a square had 6 or more GPS map spots, it was considered a curbstoning cluster.
    4. In addition, any adjacent square with six or more GPS map spots was considered part of the same curbstoning cluster.

3.2 *Curbstoning Cluster Analysis*

Once all address areas were checked for curbstoning clusters, analysis of the curbstoning cluster data began in earnest. The usefulness of the curbstoning cluster data as a predictor of poor lister performance was evaluated by comparing AAs containing curbstoning clusters to the pass/fail decisions of the AC DQC.

This analysis compared AAs containing curbstoning clusters with those that did not contain curbstoning clusters. If the AAs containing curbstoning clusters failed more often, then this would imply the presence of curbstoning clusters within a block is an indicator of poor lister performance.

For example, consider a situation where an AA that has not had many new housing structures built or demolished in recent years. A lister could have sat at a local coffee shop and marked each structure as “Verified” and the AA would have probably passed the current DQC.   However, the curbstoning cluster analysis could detect that the lister was not following procedures and may not have checked all the structures thoroughly.

A sample of 621 curbstoning clusters was selected and plotted using Google Earth. Each curbstoning cluster was assigned a code based on what was able to be determined from the overhead imagery. Each code had an associated decision whether the curbstoning cluster was “Suspicious” (lister appeared to not follow procedures), “Not suspicious” (lister seemed to follow proper procedures), or "Unable to Determine." The codes and decisions used are below:

Table 1: Curbstoning Cluster Classification List

Code	Description	Decision*
A	Marina: mapspots refer to boats; but GPS is on land in the marina	Not Suspicious
B	Insufficient imagery (out of date or poor resolution)	Unable to Determine
C	GPS coordinates are significantly removed from the manual coordinates for no obvious reason	Suspicious
D	Potential traffic issues: may have been unsafe to stop along main road to collect mapspots	Not Suspicious
E	Appears to be a multi-unit structure (lister collected GPS for each unit)	Not Suspicious
F	Cul-de-Sac; GPS coordinates near center of cul-de-sac	Not Suspicious
G	Gated neighborhood	Not Suspicious
H	Appears to be a zooming problem (i.e., lister was too far zoomed in or out)	Not Suspicious
I	GPS coordinates are clustered for houses on a given street or streets	Suspicious
J	GPS coordinates are centered on one house (or other structure) far from the manual coordinates of the units	Suspicious
K	Further research required	Unable to Determine
L	GPS coordinates clustered on one house; manual coordinates are nearby (i.e., incorrectly zoomed while falsifying data)	Suspicious
M	Parking lot of a townhouse (not apartment) community	Suspicious
N	Middle of the ocean	Not Suspicious
O	Middle of a field with no houses around	Unable to Determine
P	GPS coordinates are clustered on a road and manual coordinates are not at houses (i.e., in woods or forest)	Suspicious

* "Not Suspicious" decisions were assigned to curbstoning clusters where the lister could have been following proper procedures and "Suspicious" decisions were assigned to curbstoning clusters where the lister did not appear to be following proper procedures.

Data Source: DSSD Quality Assurance Branch

3.3 *Strand Length Analysis*

The distance between manual map spot and GPS map spot coordinates, or "strand length," was also considered as an alternative indicator of poor lister performance.

Use of strand lengths represented a more general approach for detecting poor lister performance than curbstoning clusters, since discrepancies between manual and GPS map spots were considered in isolation rather than as a clustered group. For each map spot, the distance between the GPS and manual coordinates was calculated. For each AA, the average strand length was calculated then tested against statistical formulas to indicate AAs that had an unusually large average strand length. The tests were as follows:

1. Comparing AA's average strand length to the overall average strand length of 35.68 meters:

   The AA was flagged if the average strand length was greater than 35.68 meters + 3δ, where δ= standard deviation based on the number of strands.

2. Comparing AA's average strand length to the 90th percentile overall strand length of 13.68 meters:

   The AA was flagged if the average strand length greater than 13.68 meters + 3δ, where δ= standard deviation based on the number of total strands.

The DQC fail rate for AAs that were flagged by one of these tests was then compared to AAs that were not flagged.

3.4 *Role of Overhead Imagery*

Overhead imagery was thought to be very important in validation of the algorithms used for the automated curbstoning cluster detection. Not all detected curbstoning clusters were the result of poor lister performance; in some cases structures were actually close enough to trigger the detection algorithm. Overhead imagery for a particular AA was used to classify detected curbstoning clusters into two categories: suspicious or not suspicious. Algorithms should be developed so that "not suspicious" curbstoning clusters are a relatively rare occurrence and detected curbstoning clusters accurately predict falsification. The Google Earth software was used to compare manual and GPS coordinates and categorize each cluster.

3.5 *Research Questions*

This evaluation will address the following questions:

1. Can a curbstoning cluster be used as a predictor of falsification?
2. Could curbstoning clusters be used for automated detection of falsification?
3. Can the strand length be used as a predictor of falsification?
4. Could strand lengths be used for automated detection of falsification?
5. Is one approach (curbstoning cluster or strand length) better than the other?

4. LIMITATIONS

For this report, it was assumed that the GPS in the HHCs used in the field achieved a high level of accuracy relative to the true physical location. Without a proper level of fidelity with the true physical coordinates, GPS technology would be unsuitable for application in the quality control of field work.

It was assumed that there will be relatively few instances where a square with 0.0001 degrees on each side contains six or more housing units. Part of the analysis used overhead imagery to verify this assumption. This was the parameter used to define a curbstoning cluster.

AAs could have been split in the field because of their size. These split AAs were not flagged as AAs with curbstoning clusters and, therefore, were included in the non-curbstoning AAs count.

5. RESULTS

In this section, the results of both curbstoning clusters and strand length analyses is presented.

5.1 *Can a curbstoning cluster be used as a predictor of falsification?*

Overall, 2.67 percent of the total AAs (733,636) were flagged as containing one or more curbstoning clusters. The AAs with one or more curbstoning clusters had a DQC fail rate of 13.70 percent, while the overall DQC fail rate was 8.43 percent[5]. This indicates that a curbstoning cluster can be used as a predictor of falsification and/or procedural violations. The AAs with no curbstoning cluster had a fail rate of 8.29 percent, which was very similar to the overall fail rate. The table below illustrates the different AAs and DQC results.

Table 2: Assignment Areas with Curbstoning Cluster(s) by Dependent Quality Control Check Result

Type of AA	Pass		Fail		Total		
	Count	Row Percent	Count	Row Percent	Count	Row Percent	Percent of Total AAs
Curbstoning AAs	16,876	86.30	2,680	13.70	**19,556**	100.00	2.67
Non-curbstoning AAs	654,917	91.71	59,163	8.29	**714,080**	100.00	97.33
Total AAs	671,793	91.57	61,843	8.43	**733,636**	100.00	100.00

Data Source: Address Canvassing Assignment Area File and Curbstoning Cluster File

A sample of 621 curbstoning clusters was selected in order to take a closer look at the situations surrounding some of these AAs:

- 33.17 percent were situations where the GPS coordinates were centered on one house (or other structure) far from the manual coordinates of the units,
- 32.69 percent were situations where the GPS coordinates were clustered for houses on a given street or streets, and
- 10.63 percent had GPS coordinates that were clustered on a road and manual coordinates were not at housing units (i.e., in the woods or forest).

The complete distribution of curbstoning cluster coding is shown in Table 3, below, and the cubstoning cluster classification list can be seen in section 3.2, Table 1:

[5] Again, it is important to keep in mind that not all AAs that failed DQC were a result of falsification and/or procedural violations. AAs could also fail the DQC because of honest unintentional mistakes by listers.

Table 3: Curbstoning Cluster Classification Results		
Code	Count	Percent
A	-	-
B	2	0.32
C	7	1.13
D	-	-
E	6	0.97
F	5	0.81
G	1	0.16
H	52	8.37
I	203	32.69
J	206	33.17
K	4	0.64
L	47	7.57
M	18	2.90
N	-	-
O	4	0.64
P	66	10.63
Total	621	100.00
Note: The shaded codes indicate suspicious curbstoning clusters		

Data Source: Curbstoning Cluster File

The 621 sample curbstoning clusters were distributed among 988 AAs. The AAs that were assigned a "Suspicious" classification had a weighted DQC fail rate of 13.82 percent, and the AAs that were assigned a "Not Suspicious" classification had a weighted DQC fail rate of 15.14 percent[6]. Both of these rates are higher than the overall DQC fail rate for AC (8.43 percent), so the presence of a curbstoning cluster gives an indication that the listing is of lesser quality than the norm. The table below shows the curbstoning cluster classification and DQC result.

Table 4: Sample Assignment Areas with Curbstoning Cluster Classification By Dependent Quality Control Check Result

Curbstoning Cluster Classification	Dependent Quality Control Result						Total			
	Pass			Fail						
	Count	Row Percent	Weighted Row Percent	Count	Row Percent	Weighted Row Percent	Count	Row Percent	Weighted Row Percent	Column Percent
Suspicious	779	85.42	86.18	133	14.58	13.82	**912**	100.00	100.00	92.31
Not Suspicious	55	84.62	84.86	10	15.38	15.14	**65**	100.00	100.00	6.58
Unknown	10	90.91	86.83	1	9.09	13.17	**11**	100.00	100.00	1.11
Total	844	85.43	85.99	144	14.57	14.01	**988**	100.00	100.00	100.00

Data Source: Address Canvassing Assignment Area File and Curbstoning Cluster File

[6] It is not surprising the DQC fail rate for "Not Suspicious" curbstoning clusters is higher than it is for "Suspicious," because these situations represent situations where the lister may have been unable to complete the assignment as intended, due to locked gates or other obstructions.

The results show that the AAs with sample curbstoning clusters considered "not suspicious" had a higher DQC fail rate than those with curbstoning clusters considered "suspicious." As a result of this, for the remainder of this evaluation all curbstoning clusters, instead of just the sample of 621, will be analyzed.

5.2 *Could curbstoning clusters be used for automated detection of falsification?*

Curbstoning clusters could be used for automated detection of falsification if the automated instrument can compute the GPS data and flag AAs with one or more curbstoning clusters when the AAs are completed by the listers.

5.3 *Can the strand length be used as a predictor of falsification?*

The first test conducted flagged AAs with an average strand length that was significantly greater than the overall average strand length (35.68 meters). This first test flagged 2.20 percent of the total AAs. The flagged AAs had a DQC fail rate of 14.62 percent, compared to 8.29 percent for AAs that were not flagged by this test.

A second test performed flagged AAs with an average strand length that was significantly greater than the 90th percentile of overall strand length (13.68 meters). This second test flagged 5.59 percent of the total AAs. The flagged AAs had a DQC fail rate of 12.56 percent, compared to 8.19 percent for AAs that were not flagged by this test. See Table 5, below, for the strand distance test results.

Both of these tests indicate that average strand length can be used as a predictor of falsification and/or procedural violations.

Table 5: Assignment Area Strand Distance Test Result By Dependent Quality Control Check Result

Strand Distance Test		Dependent Quality Control Result				Total		
		Pass		Fail				
Parameter	Test Result	Count	Row Percent	Count	Row Percent	Count	Row Percent	Column Percent
Average (35.68 meters)	Not Flagged	658,036	91.71	59,487	8.29	**717,523**	100.00	97.80
	Flagged	13,757	85.38	2,356	14.62	**16,113**	100.00	2.20
	Total	671,793	91.57	61,843	8.43	**733,636**	100.00	100.00
90th Percentile (13.68 meters)	Not Flagged	635,952	91.81	56,693	8.19	**692,645**	100.00	94.41
	Flagged	35,841	87.44	5,150	12.56	**40,991**	100.00	5.59
	Total	671,793	91.57	61,843	8.43	**733,636**	100.00	100.00

Data Source: Address Canvassing Address File and Curbstoning Cluster File

5.4 *Could strand lengths be used for automated detection of falsification?*

Strand length can be used for automated detection of falsification if the automated instrument can compute the GPS data and flag AAs that have a significantly higher average length than some parameter when the AAs are completed by the listers.

5.5 *Is one approach (curbstoning cluster or strand length) better than the other?*

Only 0.81 percent of the AAs had both one or more curbstoning clusters and were flagged for having an average strand length significantly greater than the overall average strand length. These AAs had a DQC fail rate of 15.70 percent.

Similarly, only 1.20 percent of the total AAs had both one or more curbstoning clusters and were flagged for having an average strand length significantly greater than the 90th Percentile of overall strand lengths. These AAs had a DQC fail rate of 15.21 percent.

Table 6: Assignment Areas with Curbstoning Cluster(s) and Flagged During Strand Distance Test By Dependent Quality Control Check Result

Strand Distance Test ***and*** Curbstoning Cluster(s)		Dependent Quality Control Result				Total				
		Pass		Fail						
Parameter	Test Result	Count	Row Percent	Count	Row Percent	Count	Row Percent	Percent of Total AAs (733,636)	Percent of Total AAs with Curbstoning Cluster(s) (19,556)	Percent of Total Flagged Strand Length AAs
Average (35.68 meters)	Flagged	4,995	84.30	930	15.70	**5,925**	100.00	0.81	30.30	36.77 (of 16,113)
	Not Flagged	666,798	91.63	60,913	8.37	**727,711**	100.00	99.19		
	Total	671,793	91.57	61,843	8.43	**733,636**	100.00	100.00		
90th Percentile (13.68 meters)	Flagged	7,459	84.79	1,338	15.21	**8,797**	100.00	1.20	44.98	21.46 (of 40,991)
	Not Flagged	664,334	91.65	60,505	8.35	**724,839**	100.00	98.80		
	Total	671,793	91.57	61,843	8.43	**733,636**	100.00	100.0		

Data Source: Address Canvassing Address File, Assignment Area File, and Curbstoning Cluster File

If the combination of the curbstoning cluster test and average overall strand length test were used to indicate AAs with possible procedure violations, only 30.30 percent of the AAs with curbstoning clusters would be caught and 36.77 percent of the AAs flagged for long strands would be caught.

Similarly, if the combination of the curbstoning cluster test and 90th percentile overall strand length test were used to indicate AAs with possible procedure violations, 44.98 percent of the AAs with curbstoning clusters would be caught and 21.46 percent of the AAs flagged for long strands would be caught.

It does not appear that one approach is better than the other. Both approaches reveal an increased DQC fail rate for AAs flagged during each approach. The combination of the curbstoning cluster and strand length tests show that the two tests were not flagging a large portion of the same AAs.

6. RELATED EVALUATIONS, EXPERIMENTS, AND/OR ASSESSMENTS

- 2010 Census: Address Canvassing Quality Profile
- 2010 Census Operational Assessment for Address Canvassing

7. KEY LESSONS LEARNED, CONCLUSIONS, AND RECOMMENDATIONS

In this section, the conclusions from the GPS evaluation are summarized and recommendations for improving future listing QC operations are provided.

7.1 *Conclusions*

AAs containing one or more curbstoning clusters failed the DQC at a rate of 13.70 percent, compared to the overall DQC fail rate of 8.43 percent[7]. This indicates that a curbstoning cluster can be used as a predictor of falsification and/or procedural violations. As long as the automated instrument can compute the GPS data and flag AAs with one or more curbstoning clusters when completed by the listers, this method could be used for detection ofprocedural violations and prediction of data falsification.

AAs flagged as having an average strand length significantly greater than the overall average strand failed the DQC at a rate of 14.62 percent, compared to 8.29 percent for AAs not flagged. Similarly, AAs flagged as having an average strand length significantly greater than the 90th percentile of overall strand lengths had a DQC fail rate of 12.56 percent, compared to 8.19 percent for AAs not flagged. Both of these tests indicate that average strand length can be used as a predictor of falsification and/or procedural violations.

As long as the automated instrument can compute the GPS data and flag AAs that have a curbstoning cluster or significantly higher average length than some parameter when completed by the listers, curbstoning clusters strand length can be used for automated detection of data falsification and/or procedural violations.

It does not appear that one approach is better than the other. Both approaches reveal an increase in the DQC fail rate for AAs flagged during each approach. The combination of the curbstoning cluster and strand length tests show that the two tests were not flagging a large portion of the same AAs. For example, the average overall strand length test (35.68 meters) also caught 30.30 percent of the AAs with curbstoning clusters and the 90th percentile overall strand length test (13.68 meters) caught 44.98 percent of the AAs with curbstoning clusters.

[7] Again, it is important to keep in mind that not all AAs that failed DQC were a result of falsification and/or procedural violations. AAs could also fail the DQC because of honest unintentional mistakes by listers.

7.2 *Recommendations*

The following recommendations would improve the listing QC programs:

- GPS data for manual map spot coordinates and actual lister location should be collected for each housing unit added or verified by the lister in order to be used as a tool for QC.

- Determine if and how the use of GPS coordinates could be used for units deleted by the lister.

- The listing instrument should flag AAs that contain one or more curbstoning clusters, along with AAs that have an average strand length significantly greater than some pre-defined parameter.

- The value for the strand length test parameter should be studied and determined for the 2020 Census. Several options include using 2010 AC overall average strand length, 2010 AC 90^{th} percentile overall strand length, or the process average during live production. Having different parameter values for different geographic areas (i.e., urban and rural, LCO types) would be ideal. For example, a longer strand length for a unit in a rural area would be expected compared to that of a unit in an urban area.

- The AAs flagged by either method should undergo a more stringent QC to determine if data falsification and/or procedural violations occurred. This can be done by a variety of options, including increasing the size of the QC sample string or verifying multiple QC sample strings within the AA.

- Since the curbstoning cluster method and strand length method did not flag a large portion of the same AAs, both methods should be considered/implemented, rather than one or the other.

- The listers responsible for the flagged AAs should be monitored in some way by managers. Counts of flagged assignments may help managers correct procedural violations, whether by additional training or termination.

- GPS coordinates should be used as a preventable measure during the listing operations, such as a pop-up warning on the instrument if the lister appears to be too far from the unit the lister is mapspotting.

- GPS coordinates comparisons need to be real-time in order for the pop-up warning to be a viable measure. If GPS data are only used for QC purposes, in order to be beneficial, the coordinates comparisons need to be completed before being worked in QC.

- Any QC methods using GPS data should work as a supplement to the traditional QC methods.

8. ACKNOWLEDGEMENTS

Special thanks to RJ Marquette, Heather Parks, and Tracey McNally for their assistance with and feedback on this report, and to all of those who reviewed it and provided comments along the way.

9. REFERENCES

Marquette, RJ and Ronia Chaar (2012), "2010 Census: Address Canvassing Quality Profile," 2010 Census Planning Memoranda Series No. 184, U.S. Census Bureau, April 4, 2012.

Schneider, Glenn, et al., (2012), "2010 Census Operational Assessment for Address Canvassing," 2010 Census Planning Memoranda Series No. 168, U.S. Census Bureau, January 17, 2012.