



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

March 1, 2012

2012 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT
MEMORANDUM SERIES #ACS12-RER-14

DSSD 2012 AMERICAN COMMUNITY SURVEY MEMO SERIES #ACS12-UC-02

MEMORANDUM FOR ACS Research and Evaluation Steering Committee

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

Prepared by: Star Ying
Community Address Updating System Branch
Decennial Statistical Studies Division

Subject: Identifying Excluded From Delivery Statistics Records That Elude
the American Community Survey Housing Unit Frame Filters

Attached is the final American Community Survey Research and Evaluation report for Identifying Excluded From Delivery Statistics Records that elude the American Community Survey Housing Unit Frame Filters. This is an analysis of 2009 Address Canvassing results in an attempt to address the increasing over-coverage in the American Community Survey housing unit frame.

If you have any questions about this report, please contact Jim Hartman (301) 763-1976 or Star Ying (301) 763-6473.

Attachment

cc:
ACS Research and Evaluation Team
Matthew Zimolzak (ACSO)
Michael Ikeda (CSRM)
Clifford Loudermilk (DSMD)
Omolola Anderson (DSSD)
Larry Bates
Jim Hartman
Steve Hefter
Tony Tersine

March 1, 2012

Identifying Excluded From Delivery Statistics Records That Elude The American Community Survey Housing Unit Frame Filters

FINAL REPORT

I. Executive Summary

The goal of this research was to identify certain classes of records on the Master Address File to exclude from the American Community Survey housing unit frame in the future. I focused primarily on a certain type of record on the Master Address File termed “Excluded from Delivery Statistic” by the United States Postal Service. The suspicion being the reason the Postal Service excluded these records from their delivery statistic would also be reason for the American Community Survey to exclude the same records from the housing unit frame. Prior research had shown that removing “Excluded from Delivery Statistic” records also removed over-coverage from the housing unit frame. We use 2009 Address Canvassing results as a way to gauge whether any record stays on the housing unit frame.

We find that while certain classes of “Excluded from Delivery Statistic” records should be excluded from the housing unit frame, the risk of under-coverage for identifying those records in the next decade is too high from using just information available on the Master Address File.

II. Introduction

The latest National Estimates of Coverage Report showed that net over-coverage in the American Community Survey (ACS) housing unit (HU) frame increased from about two percent in 2002 to more than five percent in 2009; however there is a slight and relatively steady net under-coverage when records flagged as Excluded from Delivery Statistics (EDS records) are excluded (Kephart, 2010). This informs us that a large number of EDS records potentially may fall outside the ACS HU frame that current filters still include in the frame. Couple this with the ambiguity in the definition of EDS records and it motivated us to attempt to identify an underlying classification of EDS records to filter out of the ACS HU frame.

The Master Address File (MAF) is the sole source of housing unit records for the ACS HU frame. Between Decennial Censuses, the Geography division uses data from the United States Postal Service’s (USPS) Delivery Sequence File (DSF) to update city-style addresses in the MAF. The DSF updates the MAF on a bi-annual basis, known as the Spring and Fall DSF refreshes. It contains a record for each mail delivery point in the U.S. The USPS classifies each delivery point as either residential or nonresidential¹. Some delivery points are Excluded from USPS Delivery Statistics (EDS records) while others are Included in Delivery Statistics (IDS records). Most EDS and IDS records added to the DSF after the last Decennial Census are included in the ACS HU frame.² This process continues throughout the decade between Decennial Censuses each time the Census Bureau receives a new DSF. From one DSF to the next, there is no inherent progression between how the USPS flags a delivery point. Delivery points flagged as EDS can remain EDS, change to IDS, or drop off the next DSF. This is because of the way the USPS defines “EDS” and “IDS”. The most common delivery points are IDS records and usually represent a living quarter where the USPS delivers mail. EDS records, on the other hand, represent everything else.

¹ ACS does not use records for nonresidential delivery points. All future references to EDS or IDS records refer to those classified as residential delivery points.

² A delivery point is included in the ACS HU frame if it passes all the rules specified in the frame creation (Bates, 2011).

EDS is a “catch-all” category meaning the USPS flags delivery points as EDS for various reasons; two examples are new construction sites that may potentially become a delivery point in the future and USPS service centers where mail is picked up instead of delivered. The current ACS filters would put both these types of records in the ACS universe but while ACS would want to include a construction site (because of the time lag between DSF delivery and ACS interview), ACS would want to exclude any USPS service centers (since they are not valid living quarters). The difficulty arises from trying to repurpose USPS data (the DSF) for ACS purposes. Construction sites, USPS service centers, as well as any other types of records flagged as EDS would appear on the DSF (and MAF) with no apparent difference. No flag on either the DSF or MAF exists that separates them in this way. The only way currently for ACS to differentiate between such cases is by sending an ACS field representative (FR) to observe the actual location of the delivery point and make the distinction.

ACS classifies EDS records in the ACS HU frame as primarily either:

- Decennial Census records,
- post-Census DSF add,
- or post-Census Demographic Area Address Listing (DAAL) addition

EDS records on the DSF that matched MAF records in the most recent Decennial Census are Decennial Census records. Every DSF update after a Decennial Census could add more delivery points to the ACS HU frame. ACS flags these new delivery points as post-Census DSF additions. The ACS HU filter decides which post-Census DSF additions are eligible for ACS. The ACS HU filter is a set of rules that tries to filter out delivery points that are not actually valid housing units. Currently, ACS treats new EDS records nearly identically to ones flagged as IDS. Both records flagged as EDS or IDS must pass through identical filters to be included in the ACS HU frame but EDS records must pass through additional filters (Bates, 2011) that exclude:

- multi-unit placeholders,
- ZIP code conversions,
- Locatable Address Conversion System conversions by the USPS,
- commercial or government units,
- potential college addresses,
- and post-Census additions to the MAF with a delivery point type flag of “X” (which indicates that no information was provided) from the DSF.

Delivery point type (Kephart, 2010) describes the physical location of the delivery point. Delivery point type is a flag on the DSF set by the USPS. Delivery point types relevant to this research are:

- R refers to a delivery point type of Residential Curblin (EDS)
 - These are traditional curblin mailbox delivery
- S refers to a delivery point type of Residential Neighborhood Cluster Box Delivery Unit (NDCBU) (EDS)
 - NDCBU are delivery units like those in town home communities
- T refers to a delivery point type of Residential Central (EDS)

- These are records in buildings with more than one ZIP+4 code assigned to a bank of boxes
- U refers to a delivery point type of Residential Other (EDS)
 - These are records where the delivery point is one serviced by other than curblines, NDCBU, or central, and include door-to-door and door slot
- X refers to a delivery point type of No Other Information Provided (EDS)

The 2009 Address Canvassing (AdCan) operation allowed the chance to differentiate between different types of EDS records on a national level. The goal of the 2009 AdCan operation was to list all living quarters in preparation for the 2010 Decennial Census. This would include a construction site of a future living quarter since the Decennial Census did not start until 2010 and the Census Bureau assumed construction would have been completed by 2010. In contrast, 2009 AdCan lists deleted postal service centers since no actual living quarter existed. We assume the guidelines for a 2009 AdCan lister to identify a living quarter coincide with an ACS FR's guidelines to identify a valid housing unit enough for us to use 2009 AdCan results to identify units that should be eligible for ACS. Our goal would be to identify additional rules to supplement the current ACS filters in place of sending out an ACS FR.

III. Data

Our stated goal is to be able to identify EDS records currently included in the ACS HU frame that ACS would exclude because they are not actual housing units. To gather relevant data for analysis, we take every record on the MAF that:

1. is eligible for ACS HU frame in 2010 prior to 2009 AdCan results,
2. EDS on the last DSF prior to AdCan,
3. and returned from AdCan with a valid action code.

These three criteria result in a universe of records classified as EDS at the time of the AdCan that are valid for ACS at the time the Geography Division applied AdCan results to the MAF. Only 15 million records fit these three criteria. Sixty-two percent or about 17 million records have a valid AdCan action out of the roughly 28 million records flagged as EDS prior to the 2009 AdCan operation. The remaining 38 percent of records either had no AdCan action code because AdCan never sent those records out (ungeocoded³) or returned them without an action code. Around 90 percent of the 17 million records were eligible for the ACS HU frame. The 90 percent, or roughly 15 million records, will be the universe for this study.

IV. Limitations

There are some inherent limitations to the data. These limitations may bias our results. As was mentioned before, 38 percent of all EDS records had no AdCan action. This is for one of two reasons. Either the Geography Division was unable to match ungeocoded records to an AdCan add or the record sent out had no action performed on it during AdCan.

³ Not enough information was present when being added to the MAF for the Geography division to place it within a census tabulation block.

The exclusion of unmatched ungeocoded records underreports the number of EDS records found in AdCan. This may inflate or deflate misclassification rates depending on how biased and how numerous unmatched ungeocoded EDS records are.

Likewise records with no AdCan action underreports both EDS records found and not found in AdCan. This may also bias our results should these untouched records be significantly non-random.

Also, similar to unmatched ungeocoded records there are records that received a “delete/add” during AdCan. A “delete/add” is how AdCan approached moving records from one block to another. If the Geography Division was unable to match deletes to additions in AdCan then these “delete/add” records appear as deletes on the MAF and for our analysis. When in reality they existed somewhere else.

Lastly, the Census Bureau trained listers during AdCan to be conservative in their assessment of whether records actually exist. Since our goal is to classify whether EDS records actually exist, this conservatism may bias our results to be more conservative and classify non-existent EDS records as existing on the ground.

At the time of our analysis, there is no way to account for these limitations in the data.

V. Methodology

Our research goal is to identify EDS records in the ACS HU frame to correctly exclude. We do the following to classify EDS records as either to keep in or to exclude from the ACS HU frame. For EDS records that passed the criteria in Section III, we define the action the 2010 Census AdCan took as either:

- a negative action (AdCan action code) if it is a
 - Duplicate (“2”/“L”)
 - Delete (“D”)
 - HU to Group Quarters (GQ) conversion (“G”)
 - Convert from housing unit to group quarter facility
 - Nonresidential (“N”)
 - Business or other non living quarter
 - Transitory Location (“T”)

- or a positive action (AdCan action code) if it is a
 - Add (“A”)
 - Change (“C”)
 - Add and Verify (“E”)
 - GQ to HU conversion (“H”)
 - Convert from group quarter to housing unit
 - Change not involving address (“K”)
 - Block move (“M”)
 - Reinstate (“R”)

- Uninhabitable (“U”)
- Verify (“V”)
- Wrong County (“W”)
- Other Living Quarter (“Z”)

Our assumption is that ACS would exclude any record with a negative action from the ACS HU frame and include any with a positive action. This replaces an ACS FR observation of a MAF record. We exclude records that have no AdCan action because there is no data to identify if a record actually exists and is a valid housing unit.

We used historic DSF variables and metrics to predict whether AdCan took a negative or positive action. They fall in the following four categories:

1. implementable rules (rules that can be easily added to the current ACS HU filter):

- the number of times a delivery point has not been on a DSF update,
- the number of times a delivery point has been IDS on a DSF update,
- the number of times a delivery point has been non-residential on a DSF update,
- the number of times a delivery point has been EDS on a DSF update,
- the number of times a delivery point has been in its most common state (not on DSF, IDS, non-residential, or EDS),
- the length of time a delivery point has been updated by the DSF
- the number of times a delivery point is EDS on consecutive DSF updates leading to the 2008 Spring DSF update⁴,
- the maximum number of times a delivery point is EDS on consecutive DSF updates,
- and 2008 DSF Spring refresh delivery point type.

2. non-implementable rules (rules that cannot be easily added to the current ACS HU filter):

- the most common state (not on DSF, IDS, non-residential, or EDS) for a delivery point⁵,
- the expected state (not on DSF, IDS, non-residential, or EDS) of a delivery point,
- the median state (not on DSF, IDS, non-residential, or EDS) of a delivery point,
- the proportion of time a delivery point is EDS on consecutive DSF updates leading to the 2008 Spring DSF update out of the length of time a delivery point has been updated by the DSF,
- the maximum proportion of time a delivery point is EDS on consecutive DSF updates out of the length of time a delivery point has been updated by the DSF,
- and the variance in a delivery point’s DSF refresh state.

⁴ The 2008 Spring DSF update is the last DSF update to the MAF before the 2009 AdCan operation.

⁵ We include because excluding a delivery point without attaching the severity to which a delivery point is in its most common state is hard to justify.

3. DSF update status

- 1997 DSF Fall update status (not on DSF, IDS, non-residential, or EDS) to 2008 DSF Spring update status

4. MAF rules:

- the address type⁶ for the delivery point in the ACS HU frame for the 2009 main sample,
- ACS HU frame eligibility type for the 2009 main sample,
- location address within structure descriptor,
- and mailing address within structure descriptor.

We included all the variables in the four categories for several reasons, the main reason being they are available on a national level across our data. The current ACS HU frame rule sets take into account several national level flags. We assume that the ACS HU frame eligibility type would account for most of the explanative power of those rule sets. The existing rule sets, however, do not consider neither the DSF history of a record nor the within structure descriptor. Those are the only other national level variables available to ACS on a continuous basis for augmenting the ACS HU frame rule sets.

We split the variables above into primarily “implementable rules” and “non-implementable rules”. What we term “implementable rules” are rules that ACS can easily add to the current ACS HU frame filter. The hope was that they contain enough predictive power to use in the future. Non-implementable rules, in contrast, are any rules that ACS cannot easily add to the current ACS HU frame filter. Non-implementable rules were included to see if they would have higher predictive power and perhaps suggest the ACS HU frame filter needs a bigger overhaul. The past decades worth of DSF update statuses are included in case a particular update is significant. The thought being that a particular update reflected current events that influenced delivery points. MAF rules such as address type, frame eligibility type, and within structure descriptor were included in case flags on the MAF already had the predictive power to filter out EDS records. In addition, these same variables and metrics will be available in the future and future research can use them to predict whether a future EDS record should be excluded or included from the ACS HU frame.

To do this, we randomly sampled the data (due to computational limitations). The simple random sampling introduces no biases that were not already inherent limitations to the data. We then used a decision tree algorithm to define different classes of records that predict either a positive or a negative resolution. Any classes that do so with a low misclassification rate may potentially update the current ACS filters. We define the misclassification rate as the observed percentage of EDS records that ACS would improperly exclude from the ACS HU frame if ACS excludes all

⁶ The address type of a delivery point is either:

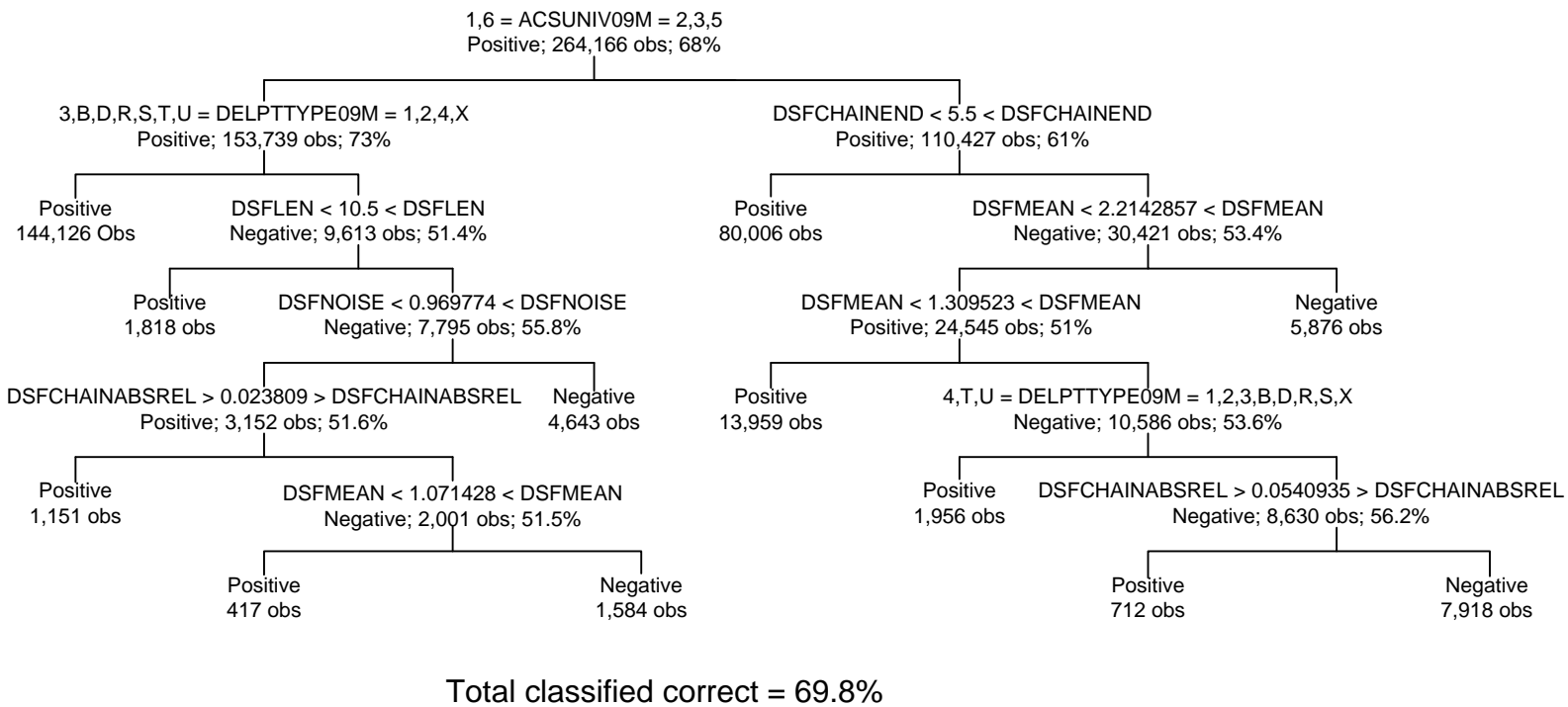
- City style
- Rural style (refers specifically to rural route addresses)
- Non-city style
- P.O. Box style

records.

VI. Results

We initially sampled from all eligible EDS MAF records. The initial analysis on the whole sample resulted in an overall misclassification rate of 30.2 percent. We define expected misclassification rate as the number of records incorrectly classified by the rule sets. The tree grown (Figure 1) overall shows the majority of the accuracy is derived from classifying by ACS HU frame eligibility type and delivery point type. The rule sets resulting from these were good at identifying records that actually existed (classifying as “Positive”) but not ones that were nonexistent (classifying as “Negative”).

Figure 1: Tree for Overall Eligible Delivery Points⁷



From the overall decision tree, we see that listers found the majority of all delivery points in AdCan. The majority of the classification leaves are positive meaning the majority of each leaf

⁷ The variable names correspond with the following variables:

- ACSUNIV09M: ACS HU frame eligibility type for the 2009 main sample
- DELPTTYPE09M: 2008 DSF Spring update delivery point type
- DSFLEN: the length of time a delivery point has been updated by the DSF
- DSFNOISE: the variance in a delivery point’s DSF update state
- DSFCHAINABSREL: the maximum proportion of time a delivery point is EDS on consecutive DSF updates out of the length of time a delivery point has been updated by the DSF
- DSFMEAN: the expected state (not on DSF, IDS, non-residential, or EDS) of a delivery point
- DSFCHAINEND: the number of times a delivery point is EDS on consecutive DSF updates leading to the 2008 Spring DSF update

was positive in AdCan. Each branch of the tree represents the variable it split upon, the estimated classification prior to split, the number of observations prior to the split, and the overall percentage of observations correctly classified by this split. For the first branch, ACSUNIV09M is the variable being split; it classifies all 264,166 observations as a positive action in AdCan, of which 68 percent were actually positive. For observations that are termed “1” and “6” in ACSUNIV09M, it then classifies those 153,739 observations as positive and only 73 percent actually were positive. A value of “1” refers to a 2000 Decennial Census addition and “6” refers to a special census addition. See footnote 7 for clarification on other variable names. Its apparent that the same few variables are used repeatedly. DSF mean or the expected state (not on DSF, IDS, non-residential, or EDS) of a delivery point is used to split three times.

Attempting to get a lower misclassification rate for excluding delivery points from the DSF, we ran the analysis on different restricted universes of delivery points: specific ACS frame eligibility types or specific delivery point types. The rule sets for identifying nonexistent delivery points with misclassification rates below 30 percent we list in Table 1.

Table 1: Lowest Expected Misclassification Rates from Decision Trees

Rule Set	Total	Living Quarters	Nonexistent	Expected Misclassification Rate
Neighborhood Cluster Box Post-Census DSF eligible add that has been EDS on the last one or two DSF updates, has been EDS for six percent or more of the time its been on the DSF, and has not been on the DSF for 19 or more DSF updates	656	120	536	18.3 percent
Neighborhood Cluster Box Post-Census DSF eligible add that has been EDS on the last nine or more DSF updates and the expected state is nonresidential	28,330	8,409	19,921	29.7 percent
Residential Central Post-Census DSF eligible add that has been EDS on 18 or more DSF updates	22,555	3,038	19,517	13.5 percent
Residential Central Post-Census DSF eligible add that has been EDS on 14 or more consecutive DSF updates	34,378	7,274	27,104	21.2 percent
Residential Central Post-Census DSF eligible add that has been EDS on 13 or less consecutive DSF updates, has been on the last two or three DSF updates, and has variance in DSF states above 1.45	1,119	312	807	27.9 percent
No other information provided Post-Census DAAL eligible add that has been EDS on the last 10 or more DSF updates	5,584	803	4,781	14.4 percent

The rule sets point to delivery points that ACS could have eliminated from the 2010 ACS HU frame at less than a 30 percent misclassification rate⁸. It is important to note that while the overall misclassification rate is around 30 percent, most rule sets for excluding delivery points were not. The rule sets above were the best performing for excluding records. What we observe is that, besides specifying the type of delivery point and type of ACS HU frame eligibility, the following four variables account for most of the rules in Table 1:

- the number of times a delivery point has been EDS on a DSF update,
- the number of times a delivery point has not been on a DSF update,
- the number of times a delivery point is EDS on consecutive DSF updates leading to the 2008 Spring DSF update,
- and the maximum number of times a delivery point is EDS on consecutive DSF updates.

It is interesting to note that non-implementable rules play almost no part in the rule sets above. This implies that non-implementable rules do not have as much explanative power as the implementable rules in Table 1 in terms of minimizing misclassification. Likewise, the decision tree algorithm discarded DSF status updates in favor of implementable rules.

The issue that arises now is that, while these rule sets are able to identify nonexistent EDS delivery points in the ACS HU frame, they require a significant history of DSF updates to do so. ACS redefines what records are Decennial Census records and what qualifies as a post-Census record after each Decennial Census. After 2011, ACS will base their HU frame on the 2010 Decennial Census results instead of the 2000 Decennial Census that was used for this analysis. Any response in the 2010 Decennial Census becomes a Decennial Census eligible record. New additions to the MAF after that primarily come from two sources; the majority comes from DSF updates while DAAL operations add others. For the purposes of this research, post-Census additions are the only new additions since the 2010 Decennial Census, meaning that they have no extended DSF update history. Additions to the MAF after the 2010 Decennial Census would have been on at most six DSF updates⁹. This makes the rule sets identified above effectively impossible to implement early on in the decade after a Decennial Census. DSF updates occur twice a year, once in the Spring and once in the Fall. This happens every year, so for example, if a rule requires eight DSF updates worth of information it would take four years before that level of data is attained.

To find rule sets implementable earlier in the decade, we looked at the four variables mentioned above across a broad range of values for different delivery point types and ACS HU frame eligibility types. The hope is to be able to find potential rules earlier in the decade without increasing the misclassification rate by a considerable amount. The trade off between implementable time frame and rising misclassification rate is shown below.

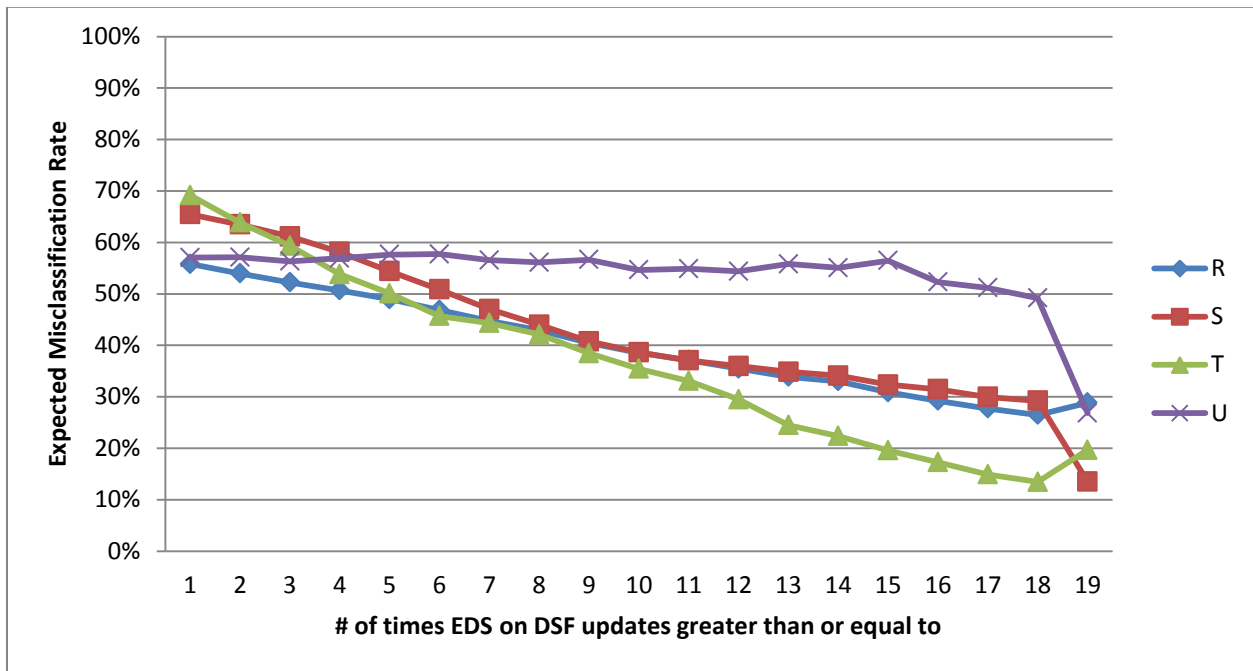
In Figure 2 we can see a decline in the misclassification rate when the records excluded are more consistently EDS in past DSF updates. The number of times a record is EDS on past DSF updates is a good indicator of whether an actual living quarter exists, specifically for delivery

⁸ The vast majority of records found during the 2009 Address Canvassing were positive, even among EDS delivery points where we know issues lay.

⁹ This is as of September 2011.

points of type T. Only about 30 percent of type T delivery points are misclassified when they have been flagged as EDS on 12 or more DSF updates (six or more years). This decreases to about 13 percent misclassified if we exclude delivery points flagged as EDS on 18 or more DSF updates (nine or more years). With ten DSF updates on the MAF to work with, we would expect about a 40 percent misclassification rate assuming we exclude only delivery points that have been EDS on all ten DSF updates.

Figure 2: Expected Misclassification Rate by # of Times EDS on DSF Updates



There is, however, a tradeoff that occurs. The fewer DSF updates we use, the earlier in the decade between Decennial Censuses we can implement a new rule. This allows more records to be classified by the rule but the misclassification rate increases. Figure 3 and Figure 4 below give an idea of how many delivery points ACS would exclude from the ACS HU frame. Using ten DSF updates as a benchmark, if ACS implements a rule for delivery points of type T halfway through the decade, ACS can expect about 120,000 delivery points (Figure 4) to be excluded from the ACS HU frame. At about a 35 percent misclassification rate, we would expect about 42,000 living quarters to be excluded from the ACS HU frame when in fact they exist. Other delivery point types fare worse than delivery point type T as can be seen by comparing Figure 2 and Figure 3.

Figure 3: # of Delivery Points by # of times EDS on DSF Updates

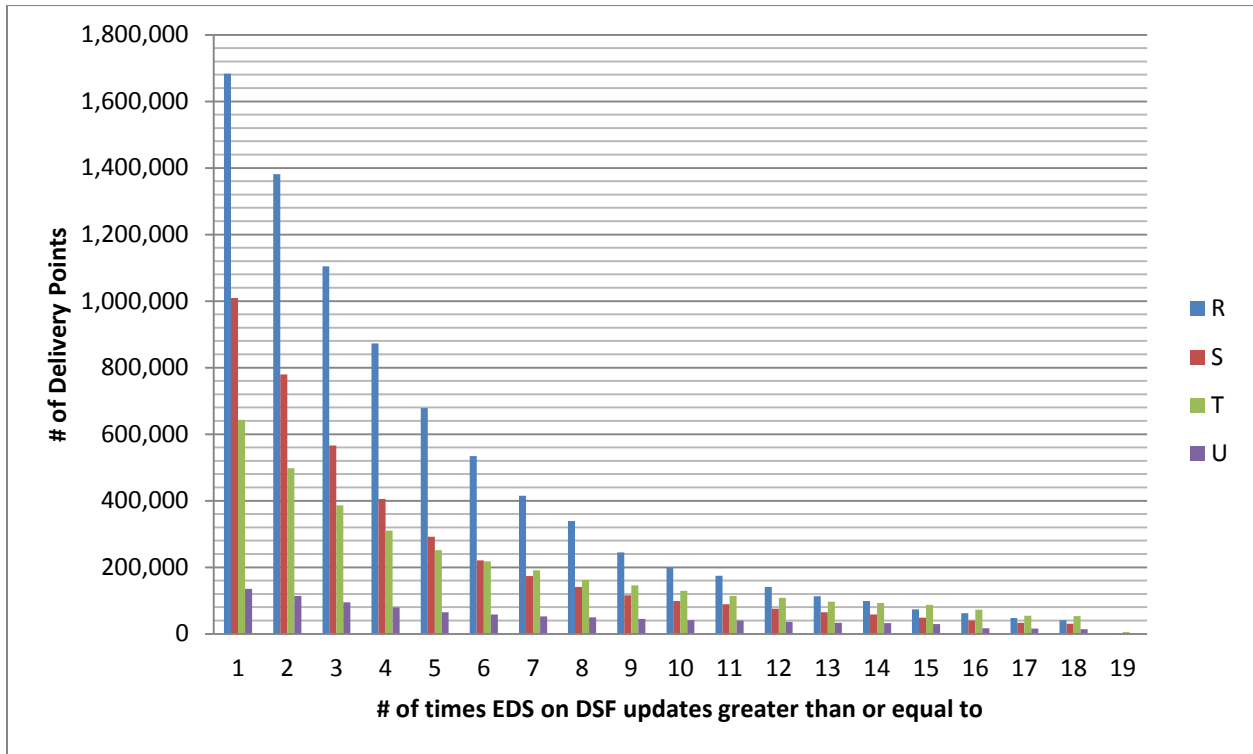
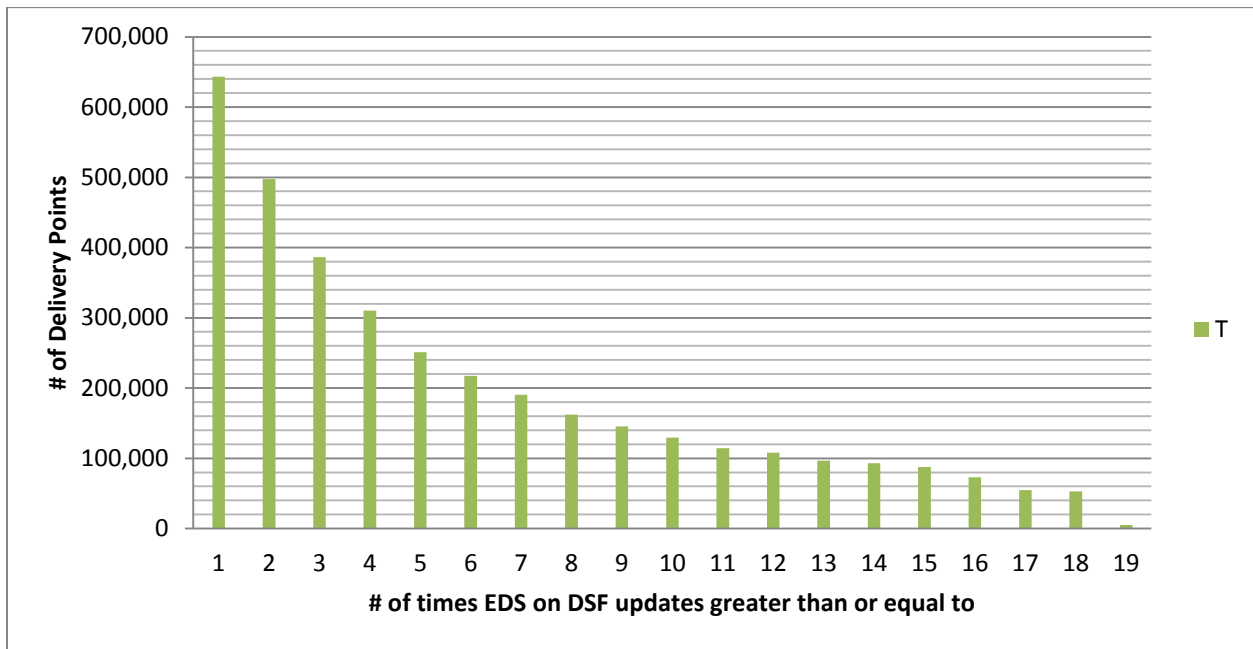


Figure 4: # of Delivery Points for Delivery Point Type T by # of times EDS on DSF Updates



The same analysis for the variables:

- the number of times a delivery point is EDS on consecutive DSF updates leading to the 2008 Spring DSF update,
- and the maximum number of times a delivery point is EDS on consecutive DSF updates,

produces similar results. Any rule set derived using only information available early in the decade would have a high expected misclassification rate. The figures (Figure 5 - Figure 8) for these variables are below.

Figure 5: Expected Misclassification Rate by Max Consecutive times EDS

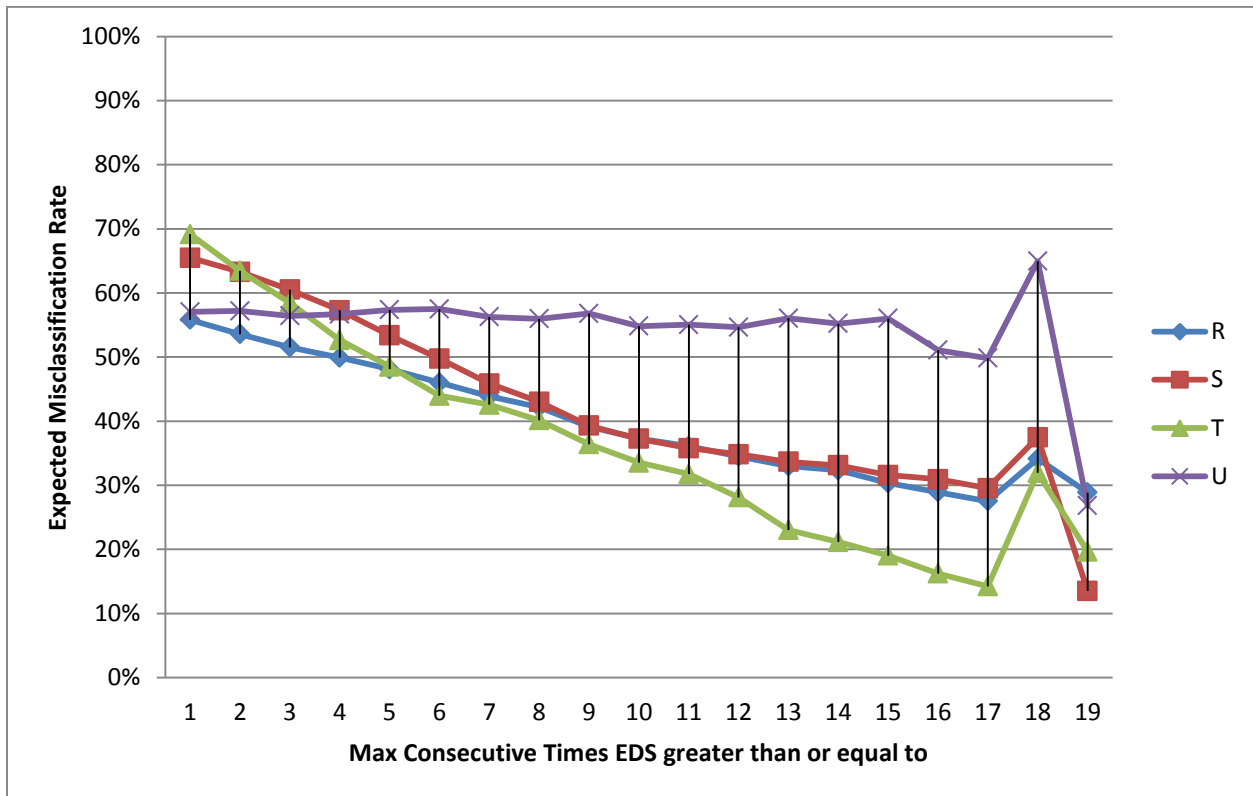
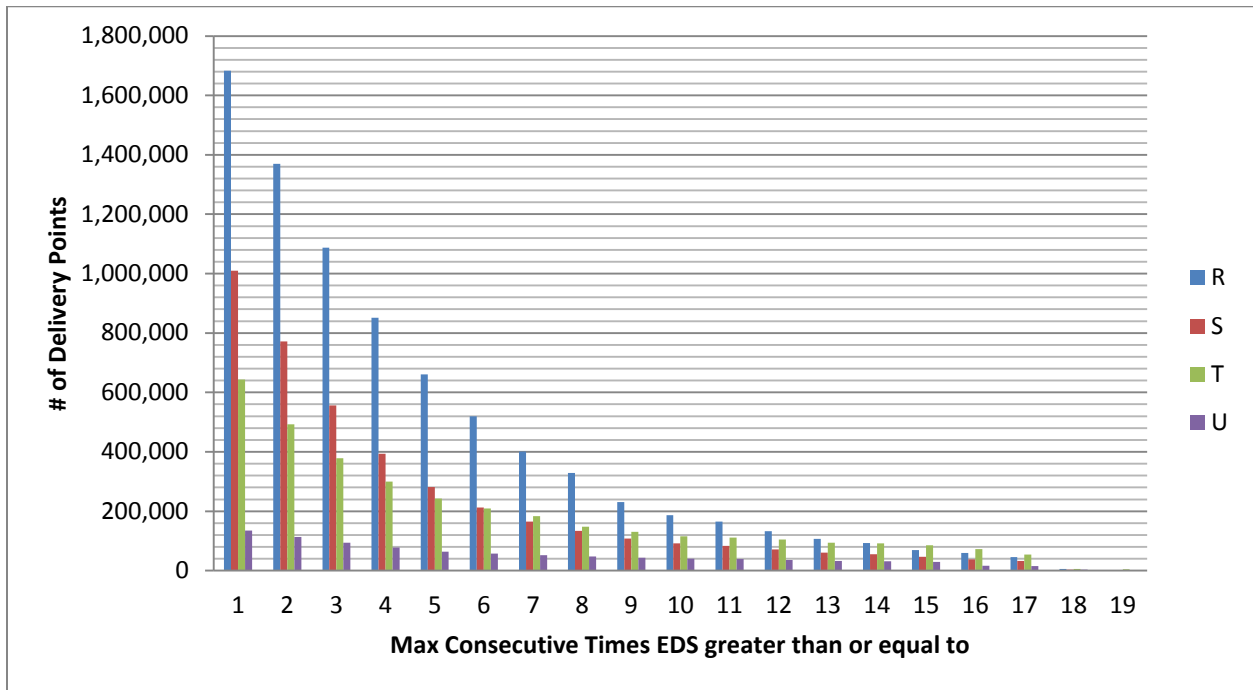


Figure 6: # of Delivery Points by Max Consecutive times EDS



In Figure 5, we see the same tradeoff occurring between the maximum consecutive times a delivery point is EDS and the expected misclassification rate. As more information comes in through the decade, the expected misclassification rate decreases. After ten DSF updates, we expect about 35 percent of delivery points to be incorrectly excluded if ACS decides to exclude all delivery points of type T that have been EDS on ten consecutive DSFs. From Figure 6, we would also expect ACS to exclude around 120,000 records from the ACS HU frame.

Figure 7: Expected Misclassification Rate by # of times EDS on recent DSF update

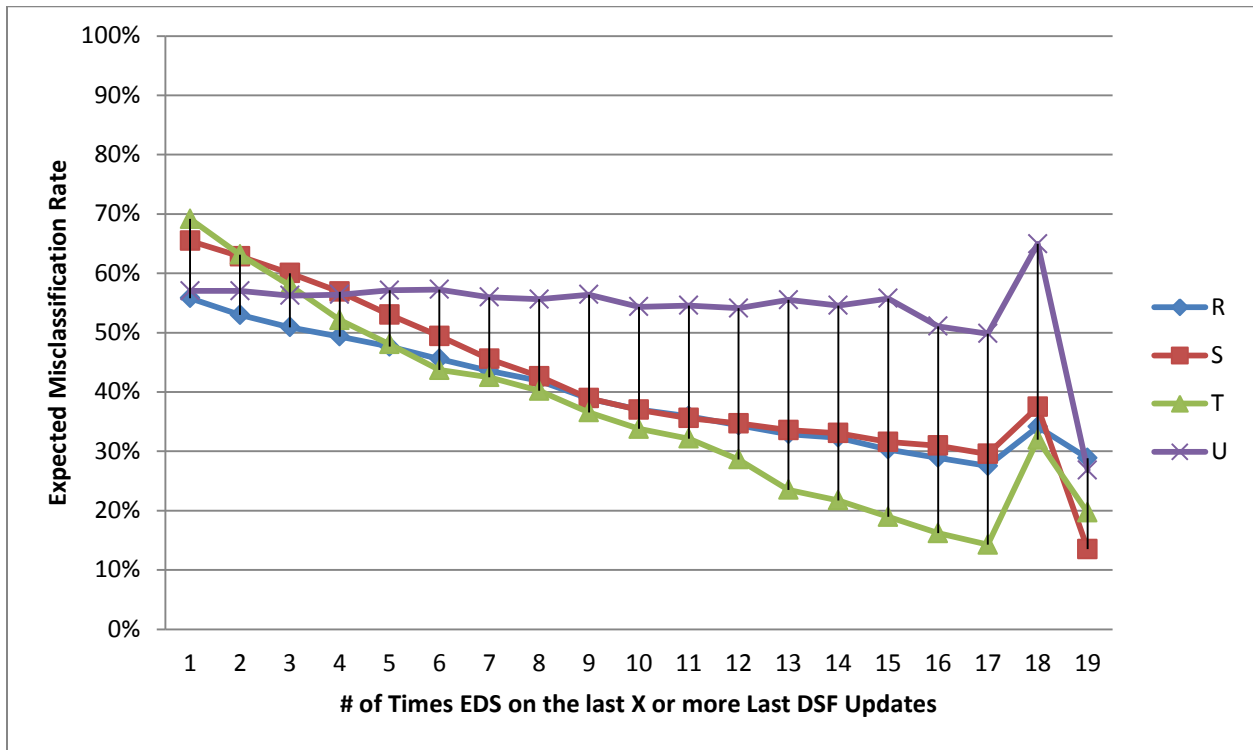
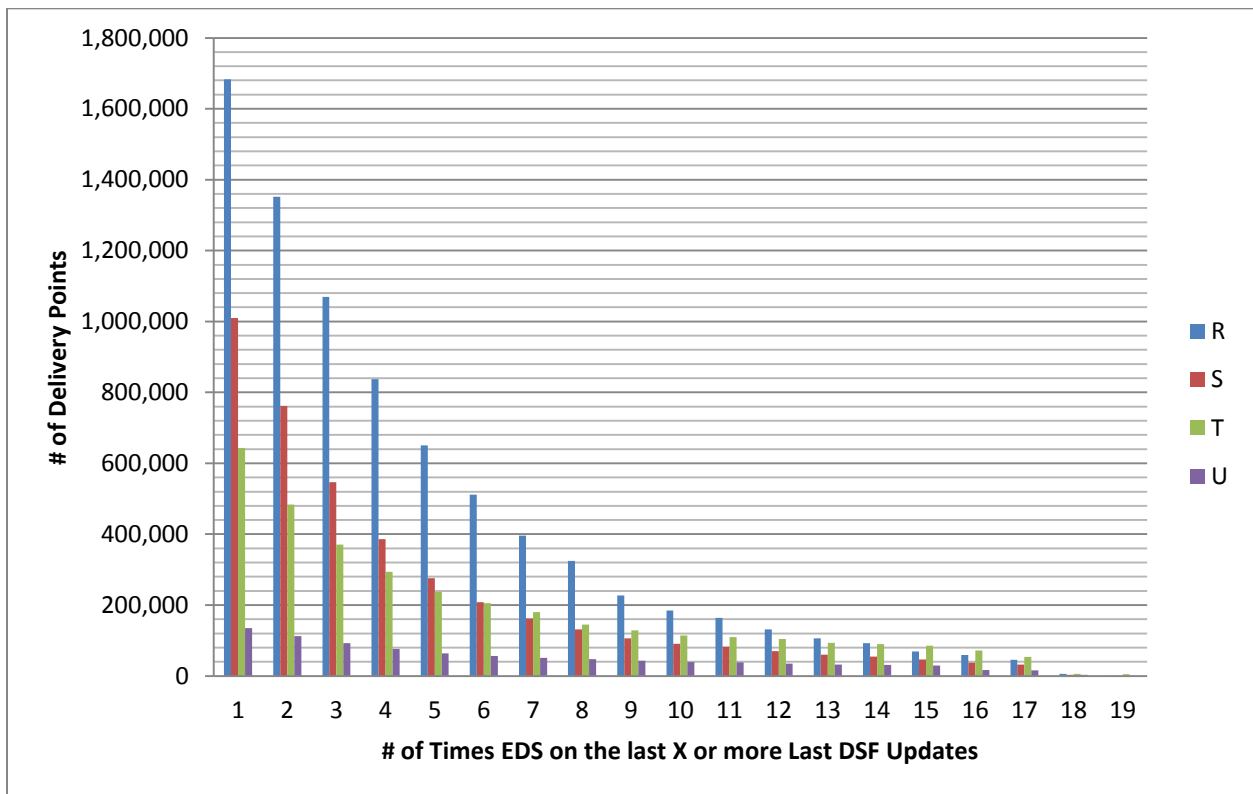


Figure 8: # of Delivery Points by # of times as EDS on recent DSF updates



What we see in Figure 7 is the same tradeoff again. Expected misclassification decreases as the Geography Division adds more DSF updates to the MAF. The analysis is nearly identical to before. Again, at ten DSF updates, we expect about a 35 percent misclassification and around 120,000 records to exclude from the frame as shown in Figure 8 for delivery point type T EDS records.

For the number of times a delivery point has not been on a DSF update, the trade-off reverses as shown in Figure 9 for all delivery point types except T.

Figure 9: Expected Misclassification Rate by # of times not on a DSF Update

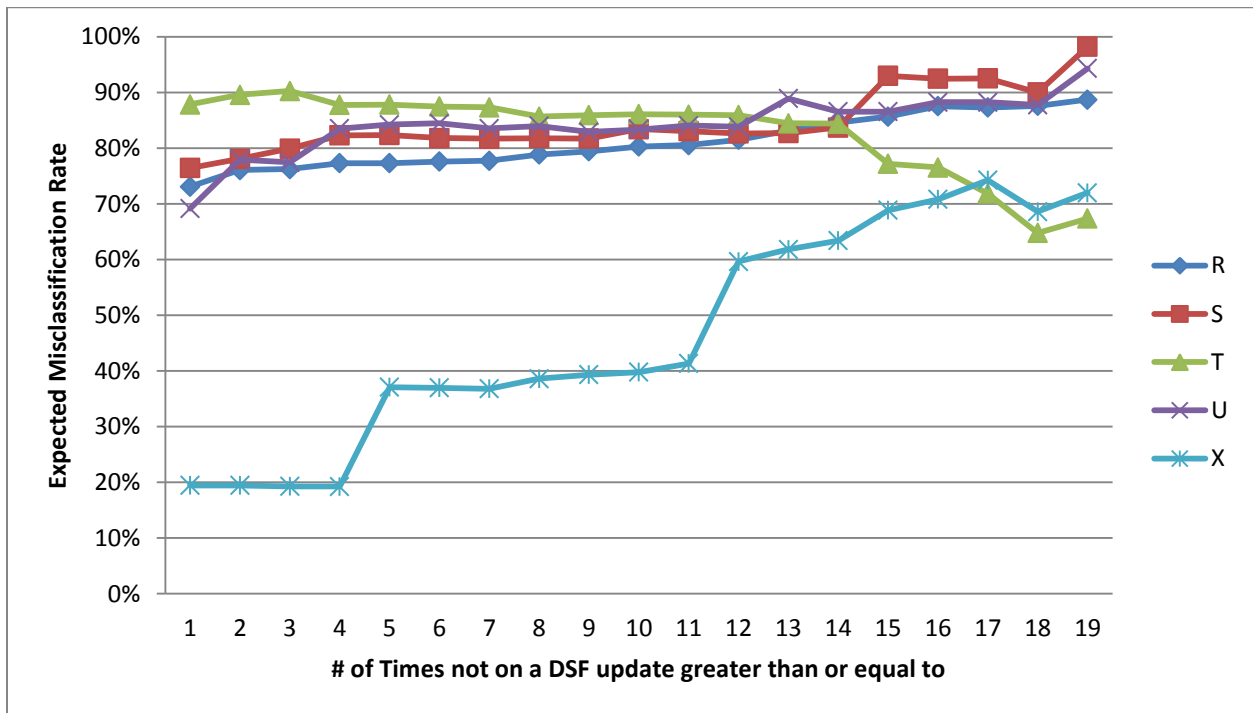
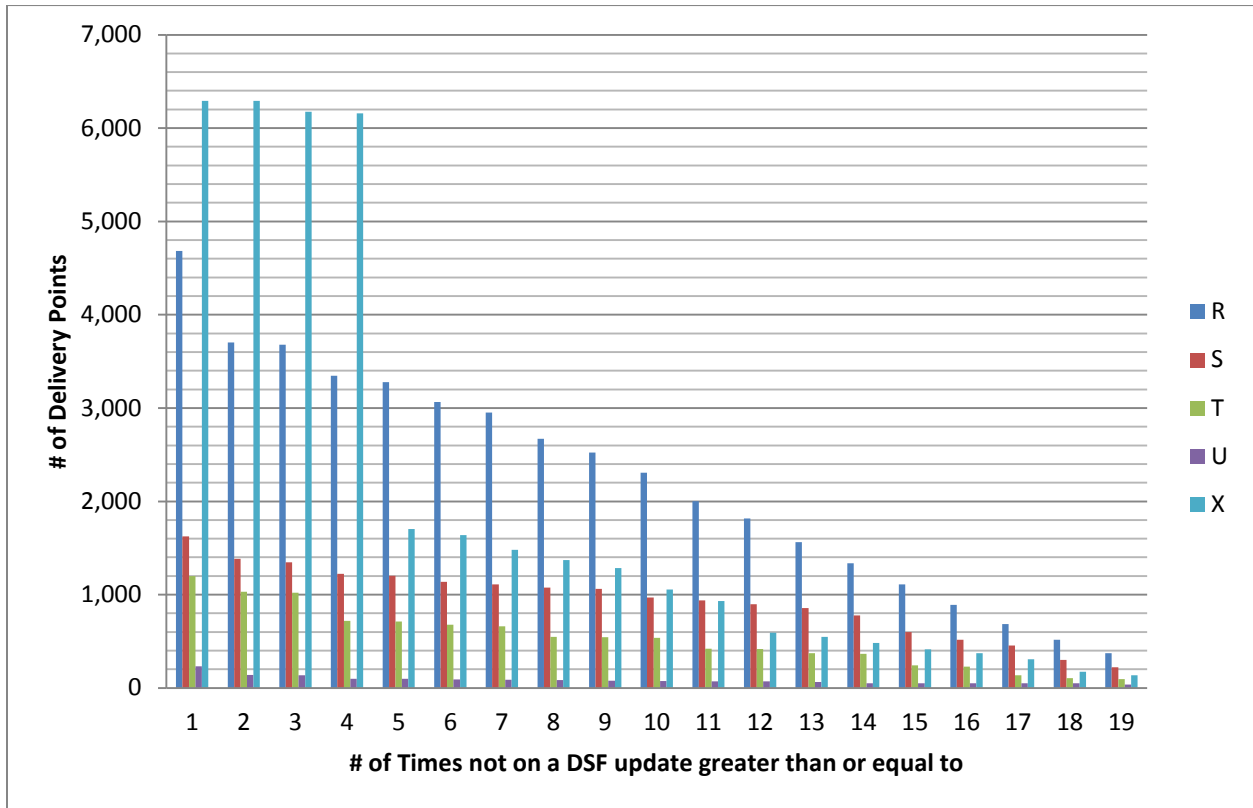


Figure 10: # of Delivery Points by # of times not on a DSF Update



The number of times a delivery point is not on a DSF is actually a measure of how recent a delivery point was added to MAF. It can be seen as how up-to-date the information on the MAF is. Delivery points not on eighteen or more DSF updates are mostly new delivery points on the DSF that were added immediately before the 2009 AdCan. These delivery points, as suspected, are more likely to be actual living quarters. This suggests for delivery point type X that excluding any records that were not recent additions from a DSF update is a viable option. These delivery point type X records are currently in the ACS HU frame. ACS only excludes delivery point type X records that are added to the MAF after a Decennial Census. For type X records found in a Decennial Census, the ACS HU frame currently includes them. This result suggests that this should not be the case.

The issue that arises again is that this is based on a decades worth of DSF updates to the MAF. In Figure 10, there is a sharp drop off in the number of delivery points that are also delivery point type X while the misclassification rate jumps up. What it tells us is that delivery points with four or less times not on a DSF are very likely to be nonexistent. Early in the decade, a post-Census DSF or DAAL eligible add will be a recent addition to the MAF. This makes the number of units we would want to exclude negligibly small. What we see is that to identify nonexistent delivery points by the predictive variables found requires a significant number of DSF updates. If ACS were to implement some rule in 2012, the misclassification rate would be quite high.

VII. Conclusion

Addressing potential net over-coverage in the ACS HU frame without causing significant net under-coverage would require a substantial amount of DSF update history. For the 2012 ACS HU frame, our recommendation is to make no changes to the existing ACS HU frame rule set because the potential decrease in over-coverage of the ACS HU frame is less than the potential increase in under-coverage from excluding any type of EDS record. We find that the number of DSF updates or years required to minimize over-coverage without incurring substantial under-coverage is more than the information available in the first couple of years following a Decennial Census can support. However, towards the end of the decade however, when net over-coverage is at its highest, ACS can safely exclude certain EDS records to address the increasing over-coverage in the ACS HU frame.

There are several possible extensions to this analysis that could better identify EDS records contributing to ACS HU frame over-coverage earlier in the decade. First, we ignore potential geographical identifiers due to the level of complexity that ACS would have to add to the current filter rule set. Incorporating geography may allow for identification earlier in the decade for certain locations, but will require substantial changes to the current ACS HU filter rule set. The current ACS HU frame rules do not consider geography (other than processing Puerto Rico differently). We could do this by clustering at some geographical level and then classifying units within those clusters. Second, we could take a causal predictive approach instead of classification. This would also require substantial changes to the current ACS HU filter rule set to implement. This new approach coupled with incorporating geography may give better results.

We could also use final 2010 Decennial Census status to decide whether EDS records exist, and therefore should be in the ACS HU frame, in lieu of the 2009 AdCan action. The 2009 AdCan data was used because at the time of this analysis, the final 2010 Decennial Census status was not yet available.

Time restrictions also blocked us from testing whether the classes we identified were reflected in field observations. Future ACS Time of Interview results can allow us to evaluate the actual misclassification rate of these rule sets and if they match the results found.

Possible extensions aside, our recommendation from our analysis is to make no changes to the way the present ACS HU frame addresses EDS records. The potential decrease in over-coverage is not enough to justify the resulting increase in potential under-coverage. We believe, however, that towards the end of the decade ACS can safely exclude certain EDS records to address the increasing over-coverage in the ACS HU frame without inducing a substantial increase in potential under-coverage.

VIII. References

1. Kephart, Kathleen. 2010. "National Estimate of Coverage of the Master Address File for 2009", *DSSD 2010 DECENNIAL CENSUS MEMORANDUM SERIES #G-23*: 11/30/10.
2. Bates, Lawrence. 2011 "Editing the MAF Extracts and Creating the Unit Frame Universe for the American Community Survey (2012 Main Phase)", *DSSD 2012 AMERICAN COMMUNITY SURVEY MEMORANDUM SERIES #ACS12-UC-1*: 08/23/11.
3. Martin, J. and Loudermilk, C., 2008. "Frame Assessment for Current Household Surveys (FACHS) Filter Rules Research: 2008 Analytical Report", *Demographic Statistical Methods Division 2010 Sample Redesign Document # 2010-4.0-G-8*:10/09/08.