

Salvo – 10/23/2015 CNSTAT 2020 Seminar (revised 10-28-2015)

- (SLIDE 2) Introduction – My goal is to examine some of the points on non-response follow-up (NRFU) that you just heard, through the lens of experience in New York City
- We have heard a lot about how the various stages of the decennial census are projected to work, complete with estimates of households that get enumerated at each point.
- Talk with anyone who has worked or observed a census in the field and you will get an earful about just how tough the final stages of an enumeration can be. It is messy and pressure-packed. That final ten percent takes an awful lot of effort. Crew leaders, who are under tremendous pressure, exert pressure on enumerators to finish and frustration rises. That frustration turns into desperation as deadlines loom. It is in these “last resort” circumstances where data collection can be compromised, in the form of proxy responses.
- (SLIDE 3) Between 2000 and 2010, New York City’s housing stock increased by some 170,000 units, a number which matched our local administrative data on housing unit changes. In Brooklyn, New York City’s borough with the largest number of units, the increase was more than 69,000 or 7 percent. When we first saw these numbers we were happy and relieved that the enumeration looked good.
- Then, we started to dig deeper into the occupancy status of housing units and found something strange – (SLIDE 4) Vacant housing units were reported to have increased by 46 percent over the decade, with Brooklyn alone accounting for an increase of 33,000, or 66 percent.
- Then we looked at the changes by census tract and this is what appeared. (SLIDE 5 - MAP).

- Most of the change in Brooklyn appeared in a single census office, outlined on the map, where vacant housing units increased dramatically. I work in a planning agency where the level of neighborhood abandonment implied by huge increases in vacant units would come to our attention via tax records, changes in housing value, lists of foreclosures, and other sources.
- Within hours of the census release, we went over to City Hall to explain. And, once the media got a hold of this, apartment seekers and real estate agents started calling us wanting to know the location of these vacant properties. You see Bay Ridge, Dyker Heights, Bensonhurst, and Sheepshead Bay are among the most desirable places to live in the city, where the housing is in great demand and, in some cases, over-occupied.
- Further research determined that something must have gone awry, probably in a desperate attempt to close out the enumeration.
- Now, what would have happened if this pattern appeared in real time? What if local census officials had administrative data that could have been used to resolve these cases – starting with instructions to enumerators on when best to visit housing units and perhaps ending with determinations based on vacancy records from the Postal Service?
- Perhaps if a flag had gone up, a call could have been placed to the local census coordinator to ask questions about these neighborhoods, much like what happens when mail response rates are lagging in some neighborhoods.

(SLIDE 6 – MAP)

- Let's now turn to another map of the 2010 Census results, which shows the percentage of households that were “substituted” in

the 2010 Census using New York City's Neighborhood Tabulation Areas (NTAs). This map depicts those housing units that literally had all of their information substituted, using a donor household from neighboring areas. These can be viewed as casualties of the enumeration, in that they existed but their characteristics were cloned from other households. While the upper interval of five percent may seem modest, the concentration of areas with high levels is quite marked. The cluster in Brooklyn is a mix of neighborhoods that are heavily black and Hispanic, which run the gamut economically. This cluster in the highest category contains more than 500,000 persons, where 25 to 30 thousand were substituted.

- Substitution adds more of what we already know, leaving out what we don't know about the missing households. A similar story can be told about the cluster in the southeast Bronx. Now, let's ask:
 1. How can administrative records help make this situation better by providing missing information?
 2. Do administrative records of high enough quality exist for these poorer residents of the city that can be used as a basis for enumerating them and providing data on their characteristics?
 3. Is the marginal gain from administrative records truly valued-added to the enumeration?

(SLIDE 7 – MAP)

- This next map provides data on households that were enumerated, but where at least one piece of information was missing and had to be imputed. While imputation exists in all surveys, what is noteworthy is that imputation here refers to just the census short-form characteristics – age, sex, race-Hispanic origin, relationship and tenure. Again, sizable numbers of persons are affected by this problem, including the large majority of Bronx neighborhoods. We need to ask whether administrative records can help alleviate this information deficit.

(SLIDE 8)

- The only way to answer these questions is through census testing of the older methods – substitution and imputation – against the results from the new methods that incorporate administrative records.
- Thus far, the Census Bureau has conducted analyses against the 2010 Census and as part of a 2014 Census test in the DC area, where occupancy status was evaluated using postal service undeliverable-as-addressed data. The results seemed promising for sizable numbers of households when it comes to vacancy. However, these tests are hardly generalizable, as the Bureau readily admits and they tell us very little about whether data on the characteristics of households derived from administrative records does justice to the concepts they are supposed to represent.
- The 2015 test in Maricopa and Savannah is still being analyzed, but a just-released Government Accountability Office (GAO)

report provides some results indicating that administrative records were successful in curbing the NRFU workload by identifying vacant units. Much less is known about the ability of administrative records to assign characteristics, however.

- Most important will be the results of the 2016 test in Los Angeles. Like New York, LA has a number of address nomenclature issues that may compromise the capacity of administrative records to determine occupancy status, along with their capacity to curb substitution and imputation.
- The question that the Census Bureau should ask with data from the test censuses is whether administrative records make the enumeration of missing households and missing attributes better than what would be obtained from these traditional substitution and imputation methods. It is possible that costs can be lowered through the use of administrative records because fewer cases will be left in those “last resort” final stages of the enumeration; but, what about the quality of the records being used and the potential biases introduced in the application of these records?
- The Bureau’s own Census Advisory Committee established a subcommittee on the use of administrative records. Their final report expressed concern about these biases because the existence and quality of those records can vary greatly by race and Hispanic origin, and by geographic area.

- There is a huge issue involving the types of data available in administrative records as a useful representation of race and Hispanic groups and subgroups, household types, relationships, and even the representation of who is actually in the household. What about the well-documented undercount of children and their coverage in administrative records? While the KIDLINK database from the Social Security Administration links parents and children by their social security numbers, the Census Bureau has yet to negotiate for the exchange of that information, much less using it in an actual census test.
- In the south Bronx or eastern Brooklyn the situation is ironic; the very people who are missed and subject to high levels of substitution and/or imputation, are those where administrative records, for example from the IRS, may be the most tenuous. And, records that would help fill in the blanks for these populations (i.e. SNAP, TANF, and WIC) are state-based and are not available for most states. Even if these records were available universally, differences issues involving data formatting and data quality vary and impose barriers to their usefulness. Finally, any reasonable timeframe for actual tests of these data goes well beyond the Census Bureau's timeline for finalizing 2020 methods.
- The Bureau needs to use its testing to shed light on this issues by contrasting the test census information for Los Angeles with that from other test census areas, or for areas within the LA test area itself. Essential to this process is an analysis of enumerated households and persons, which attempts to replicate their characteristics using administrative records. In cases where there

is a large divergence, the opportunity to revisit households in the 2016 test may prove to be very useful in evaluating these sources.

- Costs will decline proportionally to the size of the NRFU universe, and administrative records have the capacity to curb the number of NRFU cases. Only further testing, however, will reveal the right formula for applying administrative records in the census, in what is likely to be a hybrid approach where they are selectively applied in instances where they have proven to be both effective and cost-saving. Without such testing, administrative records may actually end up introducing errors. Thus, the appropriation for census testing is likely as essential to the success of the decennial census as the appropriation for the census itself.
- Figure 2 on page 10 of the recent GAO Report provides a list of administrative data sets, with information on their use, availability and status.

Reference: U.S. Government Accountability Office. “2020 CENSUS: Additional Actions Would Help the Bureau Realize Potential Administrative Records Cost Savings.” GAO-16-48. October 20, 2015.