History of Sampling Methodology in the HUD Assisted Housing Household Rent Quality Control Studies

Pedro Saavedra[1], Francine Barrington[2], Sophia Zanakos[3]
Mary Kay Sistik[4]
[1]ICF International, 11785 Beltsville Drive, Suite 300, Calverton, MD 20705
[2] ICF International, 11785 Beltsville Drive, Suite 300, Calverton, MD 20705
[3] ICF International, 11785 Beltsville Drive, Suite 300, Calverton, MD 20705
[4]ICF International, 11785 Beltsville Drive, Suite 300, Calverton, MD 20705

Abstract

For over three decades HUD has conducted a series of national surveys examining the eligibility and rent determination processes in assisted-housing programs. While the sample designs for these surveys have been similar, there have been significant modifications over the three decades. The HUD Quality Control Studies have always involved three phases or stages -- geographically defined PSUs, assisted-housing projects (AHPs) of different types, and the selection of households within AHPs. However, the exact population, the estimates required, and the information contained in the frame files have changed over time. We explore some of the modifications that have been made over the years, including the change in PSUs from congressional districts to county-based clusters, the diverse structure of AHPs, the shift from record abstraction to tenant interviews and third-party verification, and the types of estimates for different assisted-housing programs over the years.

Key Words: Cluster Sampling, PPS, Measure of Size, County-Based Clustering

1. Introduction

The Improper Payment Information Act (IPIA) passed by Congress in 2001 states that any program or activity of an agency whose improper payments exceed $10 million provide an estimate of improper payments and actions to reduce them. The HUD Assisted Housing Quality Control studies (HUDQC) provides information which HUD uses in answering to IPIA. With over $10 billion in assisted housing subsidies, based on their size alone, HUD is required to provide estimates of improper payments.

Long before the IPIA, though, HUD has worked to identify improper payments in household rental assistance payments. One of the first HUD Assisted Housing QC Studies began over three decades ago and was completed in 1982. While the methodology and focus has changed over the years, the general purpose of the QC studies has been: to determine whether projects calculated the rent correctly, to determine whether projects documented and obtained information provided by the households, and to determine the amount associated with those errors.

There are several reasons to revisit the sampling designs of the past three decades. First, it is interesting to see the evolution and reasoning behind the current/modern design. Second, it is good to revisit all of the different challenges, how they were approached and resolved. . Finally, as new methods are suggested for the improvement of the current design, it is important that one be aware of the approaches which were discarded and why the current design is as it is.

Since 1980, there have been three sets of HUD QC studies closely corresponding to each of the past three decades. Each set of studies did not necessarily consist of one report or one data collection per set, however, the emphasis of the study and the sampling methodology was fairly uniform throughout each set.

The three studies are described as follows:

- The 1980s: The first set consisted only of record abstractions from assisted housing projects.
- The 1990s: While the second set began in the late 1980's, the study was primarily conducted in the 1990's. It included household interviews in addition to record abstractions.
- The 2000's to present: The third set began in the late 1990's and continues today. It consists of both record abstractions and household interviews.

The three sampling methodologies all involved a multi-stage sampling design where: 1) the primary sampling units (PSUs) were sampled first, 2) projects were sampled from PSUs, and 3) household records were selected from projects. The general differences in the three studies represent changes in objectives on the client side, changes in the population of assisted housing, and changes in the data available to pull the sample. Each change provided its own challenges. These changes fueled changes in the sampling methodology including: differences in determining the probabilities of selection, differences in sample sizes, in the data collected, in the definition of PSUs, and in the program types defined.

The paper will discuss each of the three sets of studies, with a focus on the third and most recent study, and end with a discussion of some of the alternative methodologies that have been experimented with over the years before concluding.

2. Study 1: The 1980's

The objective of the first Study was to identify the degree to which projects calculated the rent correctly and documented the information provided by households. The main focus was on projects. While there was some interest in national totals, the major objectives were defined in terms of separate estimates and models for each program type. In addition, the sample was designed to permit estimates regarding projects, and thus the number of households per project was larger than for the other two. One additional difference between *Study 1* and the other two studies is that households were not interviewed in the first study. Study 1 included three program types: Section 8, Public Housing, and Section 236. Section 8 included existing housing, new construction, and substantial rehabilitation programs. Section 8 was also assigned a larger sample size because there were several kinds of management models and a larger sample size was needed to obtain estimates and conduct analysis.

The primary sampling units for *Study 1* (*PSU*s) were congressional districts and combinations of congressional districts. Study 1 was the only study to use congressional districts. This was primarily because the frame contained that information. However, they were also appealing because each district had a similar population, and they made good sets of boundaries around which to arrange the PSUs. Overall, a total of 242 PSUs were created, and 33 PSUs were selected. The ten largest metropolitan areas formed one PSU each. PSUs were grouped into 6 strata, with the first stratum consisting of the 10 largest MSAs. The remaining strata were defined by the number of projects found in the strata. Unlike the two later studies, the first one did not use probabilities proportional to size. PSUs were randomly selected from strata and projects were randomly selected from PSUs. The sample had 188 Section 8 projects, 161 Public Housing projects, and 147 Section 236 projects. The number of household files was determined by the desire to obtain project level estimates with a 95% Confidence Interval of 10%. This entailed taking a census for projects fewer than 50 households and 50 to 78 households for larger projects. All in all 25,000 files were sampled. However, a number of files could not be used. Most were Section 236 households that were also subsidized by other programs.

All of the programs in scope for the first study involved assisted housing projects that could be associated with one particular address. Subsequent studies included tenant-based Section 8 programs (vouchers and certificates) which would have been more difficult to handle if congressional districts had been used as PSUs. This motivated the move to a county-based definition of PSUs.

3. Study 2: The 1990's

Starting in Study 2, the building blocks for the modern design begin to appear. The second study started interviewing tenants to identify sources of income the project may have missed, and it focused more on households than the first. This study did abstract records and record documentation in the files, but also interviewed households and obtained documents from employers.

Study 2 called for estimates for 2 kinds of projects:

- PHA-administered projects including
    - *Public Housing*
    - *Turnkey III Homeownership*
    - *Section 8 Mod Rehab, Certificates, Finders Keepers and Vouchers* and
- Multifamily projects including
    - *Section 236, projects without household-based subsidy, rental assistance payments, Section 8 subsidy.*
    - *Section 8 New Construction/Substantial Rehabilitation*
    - *Section 8 Loan Management/Property Disposition*

Study estimates were required to be made separately for the two kinds of projects. Therefore sampled projects and households were to be approximately equal for the 2 domains with one exception. In study 2 an additional study looking at the Analysis of Automated System for Household Eligibility Certification (ASTEC) was combined with the main study. Hence projects in this program were oversampled. Additional all-ASTEC PSUs were created and sampled and the ASTEC projects in these PSUs were

combined with the ASTEC projects in the main sample. The sample was selected to accommodate two waves of data collection so that each wave would yield a representative national sample, but with lower precision than the two waves combined. This ended up being a good thing because due to budget considerations, only one wave was collected and analyzed. This paper presents the original design, even though only half the PSUs were actually sampled.

Study 2 also introduced vouchers into the frame. Section 8 Vouchers are household-based, thus do not correspond to buildings and have no address associated with them. This resulted in needing to change the building blocks for the PSUs. The Section 8 Voucher had records included for each contract. It was determined that contracts had no significance to the study, since PHAs did not differentiate vouchers under different contracts and some of the contracts had a very low number of participants associated with them. The universe lists contained the county in which the project was located so projects were defined as a PHA/County combination. When county information was missing, it was imputed proportionately to the number of households funded by the PHA that resided in each county.

The main objective of the study also changed. The new objective was the measurement of gross dollar error, defined as the absolute value of the difference between the actual rent paid by the household, and the rent the tenant should have paid (the Quality Control (QC) rent. The target variance resulted in estimates that met a 95% confidence interval of 4%. The final design resulted in 114 PSUs with 6 projects per PSU and 7 households per project. A minimum number of tenants per project to be in scope were set at 14 tenants per project. This was derived from selecting 7 households per project, and having 7 additional potential replacements. Only projects in the 48 states and DC were considered.

When the PSUs were created, conditions were imposed. To meet the requirements to be a PSU, four PHA-administered and four Multifamily projects were required in each PSU. In addition, each PSU needed to contain projects administered by at least two PHAs, and needed to be geographically compact so that the field staff could collect data economically. Finally, the design called for sampling PSUs with Probabilities Proportional to Size (PPS). This was the first study to introduce this method of sampling PSUs. The measure of size was the average of the proportion of households residing in the PSU for each of the two programs. Overall, 652 PSUs were created from the frame, consisting of single or adjacent counties. To select the PSUs, the probabilities of the PSUs were added and rounded to obtain the allocation for each region. Within each region, each PSU was ordered and a systematic random sample with probability minimum replacement was drawn. If the probability was greater than one, the cluster could have been selected more than once. A total of 114 PSUs were selected. However, some were selected multiple times resulting in 107 distinct PSUs. Six projects were selected per PSU, of which 2 to 6 were PHA-administered and the rest were Multifamily. Projects were also selected using PPS, where households served as the measure of size. In total, 684 projects were selected.

4. Study 3 (1999 – Present)

Study three represents the modern day sampling design and is represented by multiple cycles. There are many similarities to study number 2; however, there are several differences. While the basic sampling design remains mostly the same, each cycle has contributed slight refinements to the design. Some of these changes have been due to circumstances or changes in the frame. While some changes were experimental and were dropped, others were considered improvements and have been retained.

The first thing to change was the population. The population now included all 50 states plus the District of Columbia and Puerto Rico. The program types included

- *Public Housing (administered by a local PHA, project is defined as a building)*
- *Section 8 Vouchers/Moderate Rehabilitation/Certificates (administered by a local PHA, household based, project is defined as PHA/county combination) and*
- *Multifamily (administered by a management company)*

The Section 236 program, which was part of the previous two studies, was no longer included in the population.

The next thing to change was the estimates. In addition to overall estimates, error rates were to be estimated for each of the three types of programs that met the following criteria. The overall estimates were required to meet the precision of a 95% confidence interval within 2 percentage points of the true population rent calculation errors defined as the mean total gross rents. The individual estimates were based on the case error rate and not dollar error. Their criteria were a 95% confidence interval plus or minus 5%.

The initial sample design called for 60 PSUs, 6 projects per PSUs, and 8 households per project. However, this design was never implemented. Because the three program types are not evenly distributed among geographical clusters, a slightly larger design effect was expected for the entire sample. Expecting a larger intraclass correlation at the project level, the number of projects was augmented and the number of households per project was reduced. The ultimate sample design resulted in 60 PSUS, 10 Projects per PSU, and 4 households per project for a total of 2,400 households. This basic sampling design continues to be used today.

As with the previous studies, there are frame issues that need to be dealt with. Projects with less than 14 tenants were dropped from the frame. Originally, 14 tenants was set as the minimum number of tenants to preserve comparability in the frame to avoid selecting projects with not enough replacements. Fiscal Year 2011 was the first year this number changed, dropping to 10 tenants per project. Another issue that arose was the introduction of Asset Management Programs (AMP). Public Housing reorganized their projects into AMPs, which ultimately resulted in larger projects potentially with multiple buildings. Public Housing also had some projects called ‘scattered sites’ which were dispersed in location and small. Previously, these were aggregated by PHA if their physical locations were close to each other. However, the introduction of AMP numbers and larger projects resolved these issues. For Section 8 Vouchers with missing county info, they were imputed by county as in Study #2.

When creating the PSUs, initial cycles classified urban Section 8 Vouchers by MSA, so that PSUs could not break up MSAs. Subsequent cycles classified projects by counties. PSUs were constrained to balance the ease of replacements with logistics and the costs of data collection. Generally the constraints were:

- Every county must be contained in a PSU
- Counties within a PSU must be adjacent
- PSUs must be entirely contained within a state
    - *Each PSU must contain a certain number of projects*
    - *Each PSU must contain a certain number of projects of each program type*
    - *Each PSU must have a certain number of households*

The above constraints have been modified from cycle to cycle based on frame issues. The number of PSUs in the frame has changed from year to year. In addition, PSUs with a total absence of one of the three program types have been allowed some years, and the minimum number of households or projects have changed as well. Often these changes are in response to: a situation found the previous cycle, or a special program that renders an entire PHA out of scope.

Similar to study #2, the measure of size was the average of the three proportions of the total US number of households that were in the PSU for each of the three program types for which estimates were required. PSUs were sorted by region, state and measure of size, with serpentine sorting at the last stage. PPS systematic sampling was used with probability minimum replacement. After selection the PSUs are examined by hand and a determination is made as to whether it is sensible to break down a PSU. In certain cases, a cluster may be too large logistically, or there are natural features that aren't taken into account by our clustering program such as a mountain or a body of water separating two adjacent counties. Each sub-PSU is required to meet most of the conditions of a PSU. A slight departure (e.g., a slightly lower number of households) is allowed. Finally, one sub-PSU is selected with PPS.

Projects are clustered in PSUs and households are clustered in projects. Ten projects for each time the PSU is selected are allocated to each PSU. Fractional allocations are made for each project type, so they add up to 10 for each PSU hit and approximately 200 for each program type. Iterative rounding is used to obtain integer allocations that add up to 10 per PSU and 200 per program type. Ten projects are allocated to each PSU.

PPS is used based on the number of households in the project. Systematic PPS has been used most cycles. Public Housing and Multifamily are sampled without replacement. Some of the Section 8 Voucher projects are very large (e.g., the Los Angeles County Housing Authority), so vouchers are sampled with probability minimum replacement.

5. Alternative Methodologies

The current approach involves the sub-stratification of clusters in a multi-Stage design. PSUs are selected with a composite measure of size and projects are stratified by program type within each PSU. While the current design has proven to be versatile and effective, it can have certain drawbacks. The first drawback is that the design effect due to weighting can be high. In addition, replacement projects sometimes need to be drawn from other PSUs and sometimes an entire PHA is found to have changed to an out-of-scope program, requiring partial re-sampling.

As mentioned before, while the basic sampling design remains mostly the same, each cycle has contributed slight refinements to the design. Some changes were attempts to continually improve the design and were experimental. While some of these methodologies were dropped, they are worth discussing. We discuss four methodologies that were examined and ultimately disregarded.

The first alternative methodology involved Permanent Random Numbers (PRNs). The sample was treated as three independent studies; however, the sample was selected using PRNs and Pareto Sampling to maximize overlap of the PSUs, where some PSUs could be selected for only one project type.

The desired result was to have three samples where the PSUs overlapped and were each selected with PPS using each set of units as a size measure. One approach was to use PRNs and select 60 PSUs for each program type with each size measure, hoping that we select the same PSUs for each. However, when we selected 60 PSUs for each program type, we found that the overlap for the three samples was smaller than expected. The end result would have been around 100 PSUs instead of 60. After running these simulations it was evident that this methodology would not be cost-effective and was ultimately discarded.

The second alternative approach treated program types as subdomains with no oversampling. In this approach, the program type was disregarded except to achieve a proportional representation, while conducting domain estimates with what sample each yields. If the combined program type estimates were more important than the individual program type estimates, this would be the best approach. However, since the individual program type estimates were as important as the overall estimate, this approach was rejected as well.

A variation of the second alternative approach was explored as a third alternative approach. Like the second alternative approach, the third approach treated program types as subdomains, however, it allowed for oversampling through the adjustment of size measures. The program type was again disregarded in the sampling design, except for the inflation of the measure of size for two of the programs to achieve an equal expected sample size for each program. This was implemented in early cycles; however, it was eventually discarded because of the importance of having a fixed number of projects per program type. Using this approach would by chance lead to having more of one program type than another, leading to a greater expected variance in the estimates.

The last approach involved a multi-phase design where the number of projects and households were allowed to vary from PSU to PSU. Under this design, the number of projects per PSU varied considerably. This led to some PSUs (with only one hit) having as few as 7 projects selected and some as many as 14 projects. This created a problem in keeping the work load per data collector as equal as possible. This design was executed one time and was discarded due to the logistical problems it created.

6. Conclusion

While the sample design has been relatively stable over the past decade, the nature of the population and the structure of the frame are ever changing. This requires continual examination of the sampling design to ensure that the most effective design is used.

Simulations remain a central aspect of the redesign process. PHAs which cut across county lines and conversions of units to different programs present special difficulties. Large burden can fall on small PHAs which are a large component of a PSU. Lastly, there is an ongoing discussion as to the role that PHAs should play in the sample design.