

Summary of the 2000 NORC National Frame

Steven Pedlow
December 3, 2008

Executive Summary

The 2000 NORC National Sampling Frame takes advantage of the availability of USPS postal address lists for much of the United States. Traditional listing (sending out field employees to list every housing unit in certain selected census blocks) is only used in areas for which USPS address lists are unavailable. NORC has researched and used the USPS lists in urban areas for the Casey Foundation project “Making Connections” and the 2001 General Social Survey (GSS 2001) project [1]. That work compared the USPS lists with listings done traditionally by field personnel, and showed that the USPS lists outperform traditional listings, with better coverage at a lower cost.

The 2000 NORC National Sampling Frame was updated based on the 2000 Decennial Census. The 2000 NORC National Sampling Frame consists of 79 National Frame Areas (NFAs) in which each of the 55 non-certainty NFAs represent 1% of the U.S. population (55% total); the 24 certainty NFAs represent the remaining 45% of the U.S. population. Approximately 50% of the U.S. population lives within the 79 National Frame Areas selected.

2000 NORC National Sampling Frame NFAs (First-Stage Selections)

The first stage of sampling for the 2000 NORC National Sampling Frame consists of National Frame Areas (NFAs), which are counties, pairs of counties, or metropolitan statistical areas (MSAs), as defined by the Census Bureau prior to the 2000 Census. These NFAs were divided into three categories. Each of these is described below.

Category 1 NFAs are the largest MSAs and are selected with certainty. Most of the 24 Category 1 NFAs are single MSAs, but four combination NFAs were formed: These four Category 1 combinations were labeled Central California, South Florida, North Carolina and Central Texas. The MSAs involved are:

NFA Combinations

Central California
South Florida
North Carolina
Central Texas

MSAs Combined

Sacramento and San Francisco MSAs
Palm Beach, Broward, Miami-Dade MSAs
Charlotte and Winston-Salem MSAs
San Antonio and Austin MSAs

The smallest Category 1 NFA is Cleveland (with 799,988 housing units or HUs, as of the 2000 Decennial Census). Taken together, these 24 Category 1 NFAs contained approximately 45% of the U.S. population at the time of the 2000 Census.

Category 2 NFAs are urban MSAs that have at least 30% of their population in “Type A” tracts, as well at least 15,000 people in these “Type A” tracts. “Type A” tracts are census tracts in which at least 95% of the population lives in blocks whose addresses are in the USPS listings database. Category 2 NFAs contained approximately 40% of the U.S. population. We actually split Category 2 NFAs into two parts: all Type A tracts comprise an NFA in Category 2a, while the remaining tracts comprise an NFA in Category 2b. Category 2b NFAs were combined with Category 3 NFAs for selection and traditional listing. Category 2a NFAs contained approximately 30% of the U.S. population, and NORC selected 30 Category 2a NFAs.

Category 3 NFAs are rural MSAs or counties that have less than 30% of their population in Type A tracts or not more than 15,000 people in these Type A tracts. The selection of Category 2b and Category 3 NFAs were combined. Any NFAs with less than 10,000 in population were combined with their nearest neighbor (whether the NFAs were in Category 2b or Category 3 was ignored during the combining). Category 3 NFAs contained approximately 15% of the U.S. population, while Category 2b NFAs contained approximately 10% of the U.S. population (25% combined), and NORC selected 25 Category 2b/3 NFAs.

Table 1 below provides a summary of the 2000 NORC National Frame NFAs and Table 2, at the end of this report lists all 24 certainty NFAs:

Table 1. Summary of NFA numbering for the 2000 NORC National Frame

| NFA Category | Numbering | Segments |
|----------------------------|----------------|------------|
| 1 (Certainty Urban) | 200001-200024 | 439 |
| 2a (Non-Certainty Urban) | 200201-200260* | 289 |
| 2b/3 (Non-Certainty Rural) | 200301-200325 | 149 |
| TOTAL | | 877 |

*Among Category 2a NFAs, NORC uses only odd-numbered (200201, 200203, etc.) NFAs.

National Frame Segments (Second-Stage Selections)

The area probability sample second-stage units are referred to as “segments.” For tracts with USPS listings available, the entire census tract (usually 1,000-3,000 housing units) is considered the segment. For rural areas without USPS listings available, traditional segments were formed with contiguous block numbers to reach a minimum size of 300 housing units. In this section, we describe how the NORC National Frame segments were selected (most will be reselected to best meet the ICP Evaluation needs).

NORC set a goal of 10 segments per non-certainty NFA. Having more segments reduces the intraclass correlations, which can be significant for some variables (especially financial variables), which results in higher effective sample sizes for analyses. If achieved, this would have resulted in 1000 total segments. This goal was achieved for NFAs with USPS listings, but not for traditionally listed segments due to cost constraints. Nevertheless, NORC still achieved six segments per NFA among those with traditional listings (except for one NFA that has only five listed segments). The total number of 2000 NORC National Frame segments is 877.

The segments were selected in four strata that depend upon the NFA Category. For clarity, the strata are presented out-of-order.

Stratum 2 is defined to contain the segments from Category 2a NFAs. Category 2a NFAs consist entirely of Type A census tracts. Since these tracts have USPS address listings, a segment was defined as an entire census tract. Within each of the 30 Category 2a NFAs, 10 census tracts were selected with probability proportional to the number of housing units. For three NFAs, there were fewer than 10 census tracts in the NFA (two have only five census tracts, while the third has nine). In these cases, all tracts were selected, yielding a total of $30 \times 10 = 289$ segments.

Stratum 3 is defined as the segments from Category 2b and Category 3 NFAs. These NFAs were rural areas that needed traditional listing methods because of the limited availability of USPS address lists. In order to limit the amount of listing necessary, segments were created by combining consecutively-numbered census blocks until the number of housing units (according to the 2000 Decennial Census) was at least 300. For twenty-five segments, however, USPS listings were available for every block. For these segments, we acquired the USPS listings rather than traditionally listing the blocks. Since we were unsure of how many segments we could acquire USPS listings for, we originally selected nine segments in each of the 25 NFAs. A random sub-sample of six segments per NFA was then selected for use in the 2000 NORC National Sampling Frame. However, one NFA has only five segments. Therefore, there are $25 \times 6 = 149$ Stratum 3 segments (and 124 of these were traditionally listed). These segments will be used for the ICP Evaluation unless the MAF is used to re-select the rural segments.

For Category 1 NFAs, segment selection was divided into two parts: Stratum 1 consists of all Type A census tracts. Since these tracts have USPS address listings, a segment was defined as an entire census tract (just like Stratum 2). Within the 2000 NORC National Frame, Stratum 1 tracts represented 94% of the Category 1 population (which represent approximately 45% of the U.S. population in the 2000 NORC National Frame), so 94% of $45 \times 10 = 423$ tracts were drawn proportional to the number of housing units. The number of tracts drawn per NFA varied from 70 (NFA 200001 New York) to only 7 (NFA 200023 Denver).

Stratum 4 consists of all non-Type A census tracts in Category 1 NFAs. Segments were selected from these tracts in the same way as for Stratum 3. 6% of $45 \times 9 = 25$ segments were originally selected, but then a subsample of 6% of $45 \times 6 = 16$ was taken in the

same way as Stratum 3. Of these 16, we were able to acquire USPS listings for 5, and traditionally listed the other 11.

The Certainty NFA segments are summarized in Table 2. For each NFA, the total number of segments is given, along with how they break down into entire census tracts, segments for which we have USPS addresses, and segments for which we used traditional listing.

Selection of Area Probability Housing Units

The NORC National Frame is designed to achieve equal-probability national samples. To achieve an equal-probability national sample of housing units, we determined that the conditional sampling rate for housing unit k in NFA i segment j (π_{ijk}) should be:

$$\pi_{ijk} = \frac{f_o}{\pi_i \pi_{ij}},$$

where π_i is the selection probability for NFA i , π_{ij} is the selection probability for segment j (given that NFA i was selected), $f_o = q_o / Q$ is the overall (national) sampling rate for housing units, q_o is the sample size, and Q is an estimate of the nation's total housing stock. This choice of π_{ijk} means that the unconditional selection probability for each and every household will be

$$\text{Prob}(\text{housing unit } k) = \pi_i \pi_{ij} \pi_{ijk} = f_o,$$

This results in an equal-probability sample of housing units. The within-segment sampling rate is constant across all housing units (subscripted by k) within the segment. This method of sampling will cause the final segment sizes to be roughly constant (within Stratum), varying only due to any discrepancies between the 2000 Census figures (used to select segments) and the actual number of housing units found. Within a segment, housing units are selected by systematic sampling to spread out the addresses as much as possible.

It is important to note that the address frames used are reasonably up-to-date. Traditional listings were collected in the summer of 2003 while the USPS listings are currently up-to-date as of March 2007. The USPS listings are updated at least once every two years. NORC does not currently use any procedure to detect and sample "missed" or newly-built housing units. Our experience is that such a procedure results in very few additional housing units (adding less than 1%), but does result in substantial costs, both in training and interviewer burden. More importantly, such a procedure seriously sidetracks interviewers from their main job of collecting interviews.

References

[1] O’Muircheartaigh, C., Eckman, S., and Weiss, C. (2002). Traditional and Enhanced Field Listing for Probability Sampling. Proceedings of the American Statistical Association Social Statistics Section, pp. 2563-7.

Table 2. The 24 Certainty NFAs and the Number of Entire Tracts and Segments by Type

| NFA | Description | Total segments | Entire Tracts | Segments with USPS | Segments listed |
|------------------------------|--------------------|-----------------------|----------------------|---------------------------|------------------------|
| 200001 | New York, NY | 70 | 68 | 1 | 1 |
| 200002 | Los Angeles, CA | 52 | 51 | 0 | 1 |
| 200003 | Chicago, IL | 31 | 31 | 0 | |
| 200004 | Central California | 32 | 30 | 1 | 1 |
| 200005 | Philadelphia, PA | 23 | 22 | 1 | |
| 200006 | South Florida | 20 | 19 | 0 | 1 |
| 200007 | Detroit, MI | 18 | 18 | 0 | |
| 200008 | Dallas, TX | 17 | 16 | 0 | 1 |
| 200009 | Washington, DC | 16 | 16 | 0 | |
| 200010 | Houston, TX | 17 | 15 | 2 | |
| 200011 | Boston, MA | 15 | 15 | 0 | |
| 200012 | Atlanta, GA | 14 | 13 | 0 | 1 |
| 200013 | Phoenix, AZ | 11 | 10 | 0 | 1 |
| 200014 | Tampa, FL | 11 | 11 | 0 | |
| 200015 | Minneapolis, MN | 11 | 10 | 0 | 1 |
| 200016 | North Carolina | 11 | 9 | 0 | 2 |
| 200017 | St. Louis, MO | 9 | 9 | 0 | |
| 200018 | Central Texas | 9 | 9 | 0 | |
| 200019 | Baltimore, MD | 10 | 10 | 0 | |
| 200020 | San Diego, CA | 10 | 10 | 0 | |
| 200021 | Seattle, WA | 9 | 9 | 0 | |
| 200022 | Pittsburgh, PA | 8 | 7 | 0 | 1 |
| 200023 | Denver, CO | 7 | 7 | 0 | |
| 200024 | Cleveland, OH | 8 | 8 | 0 | |
| TOTALS FOR CATEGORY 1 | | 439 | 423 | 5 | 11 |