Unpacking the DSF in an Attempt to Better Reach the Drop Point Population

Katie Dekker, Ashley Amaya, Felicia LeClere, and Ned English

NORC at the University of Chicago, 55 E. Monroe Street, Chicago, IL 60603

Abstract
Drop points are a challenging component of the United States Postal Service's Delivery Sequence File (DSF) with respect to survey research. Drop points occur when there is a single mail receptacle serving multiple units, resulting in no designation between units. This lack of unit identifier impedes mail or phone contact. NORC conducted a limited field study of drop points in 2011, which showed that they are often geographically clustered. Additionally, buildings containing drop point units tend to be similar in structure to neighboring buildings without drop points. At question is whether or not individuals living in drop points are different from residents at non-drop point addresses. This paper attempts to determine if the presence of drop points on the DSF creates coverage bias and whom we would be missing if we eliminate drop points from the sampling frame.

Key Words: Drop point, drop stop, Delivery Sequence File (DSF), frame coverage, address-based sample

1. Introduction

In recent years, address-based sampling (ABS) has become a common method of building sampling frames (O'Muircheartaigh, Eckman, and Weiss 2003, Link et al. 2008). Studies have shown that the U.S. Postal Service Delivery Sequence File (DSF), which is the basis for ABS frames, has sufficient population coverage, particularly in urban areas, to lower the cost of creating sampling frames (Iannacchione, Staab, and Redden 2003, Link et al. 2008). The DSF is a compilation of all housing units in the United States that receive mail delivery, which is continually updated by mail carriers and is organized by zip code, carrier route, and walk sequence number (O'Muircheartaigh, Eckman, and Weiss 2003). Most addresses on the DSF are reported as a city-style address, making them straightforward to sample. The DSF, however, has a few idiosyncrasies that undercut its value as a sampling frame. One of these complications is an address style called the drop point or drop stop.

A drop point as defined on the DSF as an address associated with multiple housing units, but which contains no unit designations. Although there is information regarding the number of units at the address, there is no information on the particular unit numbering scheme. For example, instead of "1 N. Main St. #1" and "1 N. Main St. #2", there would be two records on the DSF with no unit designation. Drop point residents receive mail in a common receptacle by name not unit number. Unlike the USPS, survey researchers often treat the housing unit not the person as the object of interest. Thus, addresses that do not include a unit number prevent surveys from reaching the sampled households. Address-based sampling, by definition, is then based on selecting an address, not an

individual, which means that drop point addresses are incomplete for the purpose of ABS designs. At a national level, drop points represent a small portion of all addresses yet they are not randomly distributed but clustered in cities. Because drop point residences represent less than three percent of all addresses, some survey organizations choose to eliminate them from sampling frames. This poses a substantial problem for surveys that either focus on local geographies that include cities or on populations that are more likely to live in cities.

In this paper, we take a closer look at the characteristics of drop point addresses and the population who inhabit them to assess whether dropping these units introduces substantial coverage bias in a frame constructed from the DSF. We use information from the DSF merged to additional information from the American Community Survey and the 2010 US Census for the city of Chicago to understand the characteristics of both the units that house drop point addresses and the population in these units. The ultimate goal is to understand whether eliminating drop points from a DSF-based sampling frame will introduce coverage bias that has the potential to bias the resulting survey statistics.

2. Literature Review and Background

Although not a great deal of research has been conducted on drop points in survey research, some existing work suggests that drop point addresses may be problematic particularly for matching telephone numbers to addresses and for mailing to specific households with certainty (Link et al. 2008). Iannacchione (2011), in fact, has suggested that drop point addresses can only feasibly be included in field surveys and should be dropped for sampling frames used for surveys delivered in other modes.

ABS sampling frames also lend themselves to multimode delivery of the questionnaire. Multiple modes, such as in-person, telephone, internet, and mail, are often used to lower costs and increase response rates (Couper 2011). Drop point addresses present different challenges for different modes, which then complicate a multimode design. For an in-person study, drop points can be problematic, but the solutions are simple. Interviewers can be systematically trained to list all housing units at a drop point building and select the n^{th} unit for interview. It does however introduce the extra cost of field verification. Mail surveys are, of course, more difficult as all mail delivery at a drop point address is delivered to a central location and residents are typically responsible for sorting and distributing the mail. Similarly telephone surveys conducted as part of an ABS design at a drop point address cannot be matched to a telephone number without knowing a unit designation. Telephone matching can then only be inexact, where the phone number is associated with the building address but not the exact unit.

As there is very little research on the specific characteristics of drop points, in this paper we will begin by first assessing whether drop points differ from other multi-unit housing. Multi-unit housing is distributed in recognizable patterns in American cities (Ford 1986). High density, multi-unit buildings are frequently in the core of cities, while single family structures are often associated with suburban neighborhoods surrounding the central city. Multi-unit buildings are often created by the conversion of single family units into multiple dwellings. Previous field research suggests that drop points are frequently located in this type of structure (Amaya et al., 2011). Duncan and Hauser (1960) note during the 1950s, housing unit conversion in Chicago resulted in six to seven substandard housing units for every ten standard housing units. Additionally, frame houses in Chicago were frequently moved to the back of lots in the early 1900s. This movement

created coach houses, or alley houses (Bigott 2005), which are also frequently listed as drop points on the DSF. The continuous change in residential building structures and demographics in the US (Conzen 2001) has created drop point residences and influenced the population who lives in them.

The second step in this analysis will explore the characteristics of persons who reside in drop points. Drop point addresses are often in substandard housing and, thus, likely to be inhabited by persons with lower incomes, ethnic minorities, and immigrants. We examine the degree to which this is true in the city of Chicago.

Finally, we examine the consequences of excluding drop points from a sampling design by investigating key demographic characteristics for the city of Chicago with and without drop point addresses. We will use a simulation to test our assumption that drop points can undermine effectiveness of ABS designs that depend on the accuracy of the address to complete the delivery of survey instruments.

3. Methods

Nationwide, 2.9 percent of residential addresses are drop point deliveries. As this research is part of a larger project, which includes field validation, we limit this study of drop points to the addresses within the city boundaries of Chicago, IL. Chicago has a concentrated drop point population, with 15% of residential addresses in Chicago receiving mail at a drop delivery point. Other cities with high drop point densities are New York City (20%) and Boston (10%).

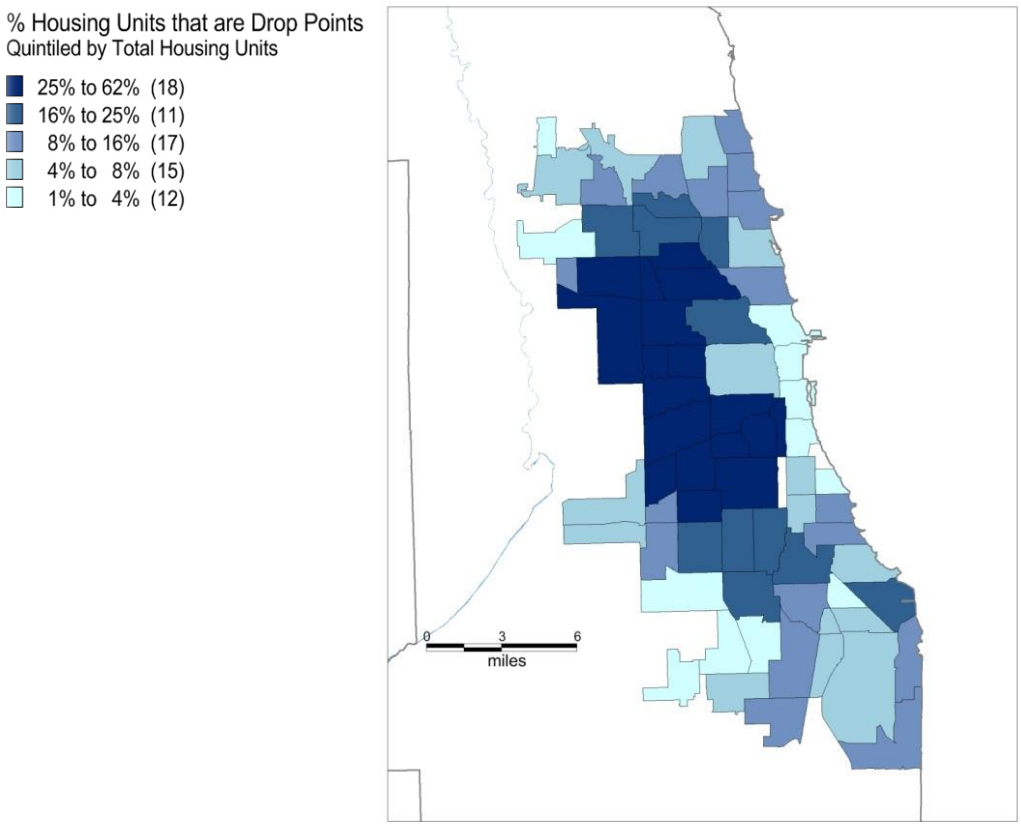
We created a subset of the universe of DSF addresses for the city of Chicago, which includes 1.1 million housing units. The addresses were geocoded to appropriate Census blocks. The geocoded address file was then converted to a dataset organized by Census block with the total number of DSF housing units, total drop point units, and summary statistics, such as proportion of drop point units (of total units) added to the file. Nearly 50,000 Census blocks in the city of Chicago were summarized in this manner, to which we added a variety of variables from the Census 2010 Summary File 1 and 2010 5-Year American Community Survey (ACS) files. The population variables added include total population, median income, owner occupancy, race/ethnicity, home vacancy, residential mobility, and year of housing unit construction. For summary purposes, we also aggregate these characteristics at the Census block group and Chicago community area levels. The block, block group, and community level datasets were then used to characterize Chicago communities with high proportions of drop points and, then, to simulate demographic outcomes if a study were unable to reach the drop point population. We describe our methods for these simulations below.

4. Results and Discussion

We calculated, at the Census block, Census block group, and community area levels, the proportion of residential drop delivery points over total residential delivery points, which include single and multiunit housing units. This calculation is for drop units not buildings. Figure 1 illustrates the geographic distribution of Chicago's drop points, which are concentrated on the west side of the city. The drop point delivery concentration at the community area level ranges from 1% to 62%. The communities to

the west and south of the central city, which is located along Lake Michigan in the central part of the map, have the highest concentration of drop point units.

Figure 1: Drop Point Density by Chicago Community Area[1]



Figures 2 through 5 illustrate the relationship between drop point density and key Census variables at the Chicago community area level. In each chart, the 77 Chicago community areas are sorted in ascending order by the percentage of drop point units along the x-axis. The corresponding Census variable appears on the graph in the form of a trend line by community area. The trend lines are smoothed across the 77 communities.

These figures demonstrate clear trends in the relationship between the characteristics of Chicago community areas and concentration of housing units that are drop points. Community areas with lower proportions of drop points have higher median household income than those communities where drop points are concentrated (Figure 2). In Figure 3, we see that higher proportions of drop points are often located in community areas with lower owner occupancy. Drop points are usually associated with multi-unit housing structures and home-ownership is usually associated with single-family dwelling units so this trend is not surprising. Community areas with high proportions of drop points are also more likely to have larger Latino populations (Figure 4). Finally, drop points occur in higher concentrations in communities with older housing stock, specifically with homes built before 1940 (Figure 5). Multi-unit housing (between 2 and 8 units)

[1] From Valassis ADVO file, March 2012

accounted for nearly 50% of all new construction in Chicago between 1900 and 1930 (Hubka and Kenny 2006).

Figure 2: Drop Points and Income by Chicago Community

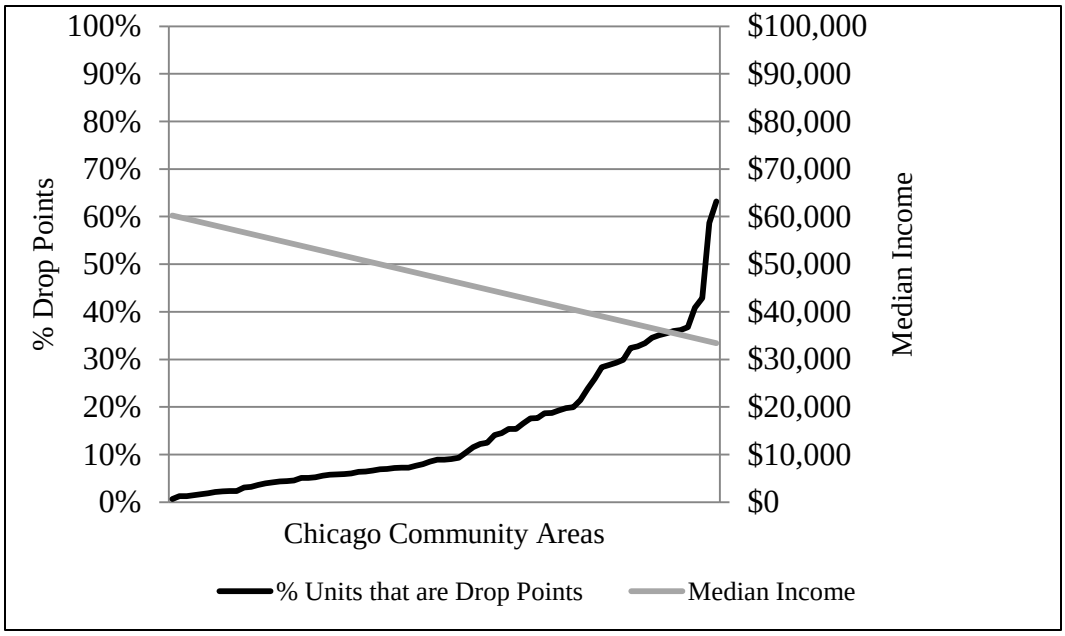


Figure 3: Drop Points and Owner Occupancy by Chicago Community

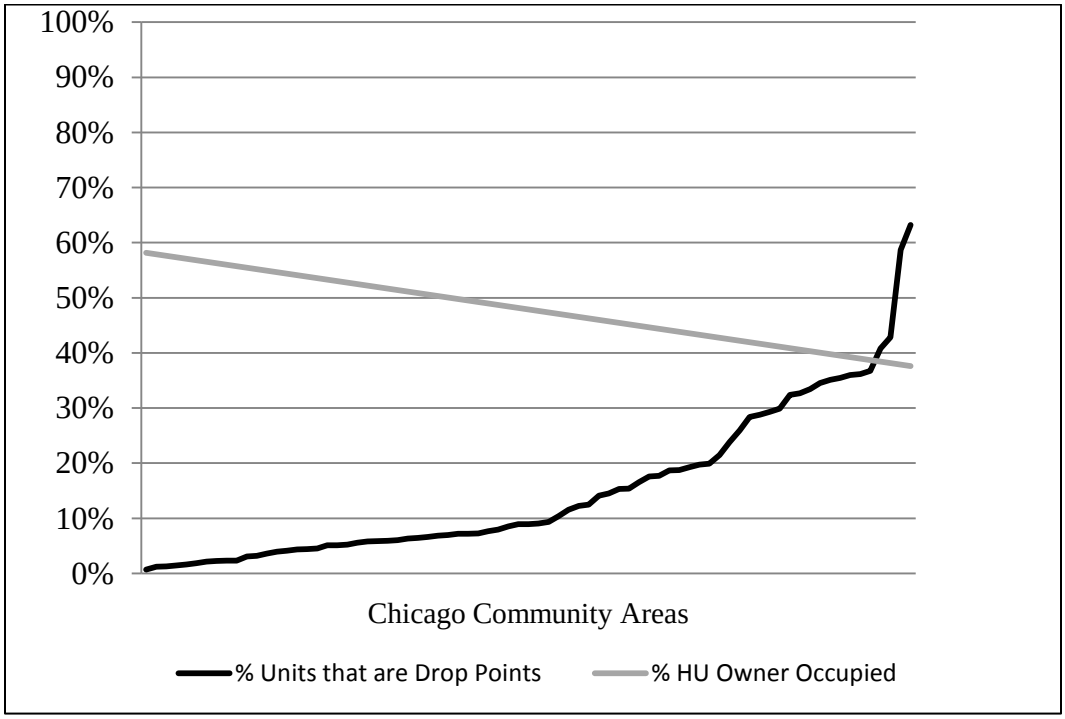


Figure 4: Drop Points and % Latino by Chicago Community

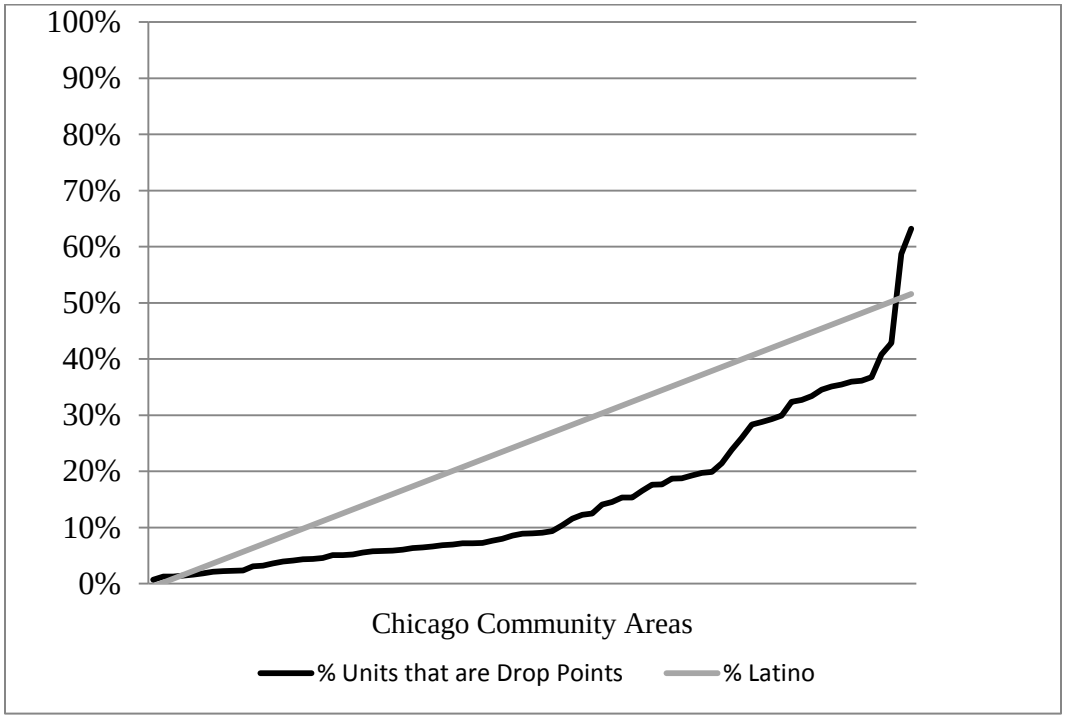
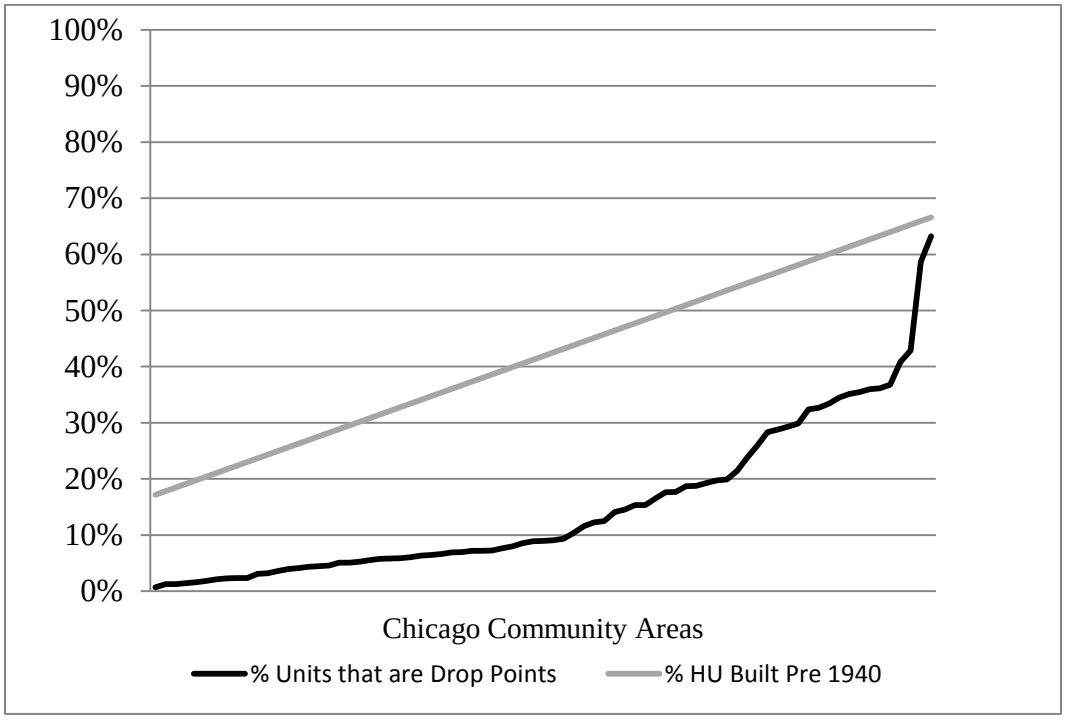


Figure 5: Drop Points and Age of Housing by Chicago Community



In summary, Chicago community areas with high densities of drop points have lower income, lower rates of home ownership, larger Latino populations, older housing stock, higher vacancy rates, and lower African American populations (data are not provided here). In this analysis, it is important to note that these figures are generated with data from the community area not from the drop point units themselves. We can only infer these trends are present in community areas with greater drop point density. We do not know the characteristics of persons that reside in the drop point units themselves.

The analysis above suggests that drop point addresses may contain populations who are considerably different than the general population. As discussed earlier many survey organizations choose to eliminate drop points from the frame because they are too difficult to mail to or telephone. We cannot determine the direct impact of eliminating drop points from the frame as we do not have demographic data from the drop point units. We can, however, simulate an approximation. To do so, we calculate median income, percent owner occupied, percent housing units built pre-1940, and percent Latino at the city and community level and weight them by DSF housing unit counts at the Census block level. After obtaining these summary figures, we recalculate the same rates, but adjust the weights to reflect the probability of selection for non-drop point housing units only. In other words, if a Census block has 100 units, but 50 of these are drop points, in the first calculation this block would receive a weight of 100, but in the second, this block would receive a weight of 50. A Census block with 100 units, none of which are drop points, would receive a weight of 100 in both simulations. With no information at the unit level, we assume the demographic characteristics are evenly distributed between drop points and non-drop points within the block.

These simulations suggest that there may be real consequences for eliminating drop points from a sampling frame. Table 1 shows the results for a subset of demographic characteristics for the city of Chicago. For the four rates we examine here, there are significant differences between the results when weights are trimmed by percent drop point at the block level. If drop points are excluded from a Chicago sampling frame, the resulting demographic distributions would portray a population who have higher incomes, higher rates of home ownership, fewer people living in homes built before 1940, and fewer Latinos than are actually present in the city.

Table 1: Chicago Drop Point Simulation

	Median Income	**% Owner Occupied**	**% HUs built pre 1940**	**% Latino**
	2010 5 year ACS	**2010 Census**	**2010 5 year ACS**	**2010 Census**
Without Drop Point Weighting Adjustment	$51,302	44.7%	44.0%	22.8%
With Drop Point Weighting Adjustment	$52,607***	45.5%***	41.5%***	19.9%***

* p<0.05; ** p<0.01; *** p<0.001

While we see significant differences at the city level, they are small. We expect to see more dramatic differences at the community area level because as demonstrated drop points are clustered. Eliminating all drop points from the sampling frame for surveys that are focused on specific neighborhoods with higher drop point densities could substantially compromise the results.

5. Conclusions

The analyses in this paper show that, in our limited study of the city of Chicago, neighborhoods with higher concentrations of drop point residences tend to have lower incomes, lower rates of owner occupancy, older housing stock, and more dense Latino populations. A limitation to this study is that one cannot directly compare drop point residents with non-drop point residents, so we approximate these differences by comparing neighborhoods of different drop point densities. Nevertheless, this research demonstrates that efforts should be made to include drop points in all sampling frames, whether in-person, mail, or telephone. One method for doing this is already in practice. Inexact assignment of a synthetic numbering scheme to units in a drop point is done to attempt mail delivery. Another proposed method of handling drop points is to create a separate drop point stratum in the sample and include all units associated with drop point building selections, although the inherent bias may not make this a viable solution (Link 2008). From both our descriptive analyses and simulations for cities such as Chicago, it appears that the elimination of drop point addresses from the frame will have clear consequences.

6. References

Amaya, Ashley, Felicia LeClere, Lee Fiorio, and Ned English. Forthcoming. Improving the Utility of the DSF Address-Based Frame through Ancillary Information. *Field Methods*.

Bigott, Joseph C. 2005. Housing Types. *The Electronic Encyclopedia of Chicago*. Chicago Historical Society. <http://www.encyclopedia.chicagohistory.org/pages/609.html>.

Conzen, Michael P. 2001. The Study of Urban Form in the United States. *Urban Morphology* (2001) 5(1), 3-14.

Couper, Mick P. The Future of Modes of Data Collection. *Public Opinion Quarterly* 75(5) 889-908.

Duncan, Beverly and Philip M. Hauser. *Housing A Metropolis – Chicago*. Glencoe, IL: The Free Press of Glencoe, Illinois, 1960.

Ford, Larry R. 1986. Multiunit Housing in the American City. *Geographical Review* 76(4) 390-407.

Hubka, Thomas C. and Judith T. Kenny. 2006. Examining the American Dream: Housing Standards and the Emergence of a National Housing Culture. Perspectives in Vernacular Architecture 13(1) 49-69.

Iannacchione, Vincent G. 2011. The Changing Role of Address-Based Sampling in Survey Research. *Public Opinion Quarterly* 75(3) 556-575.

Iannacchione, V. G., J. M. Staab, and D. T. Redden. 2003. Evaluating the Use of Residential Mailing Addresses in a Metropolitan Household Survey. *Public Opinion Quarterly* 67(2) 202-210.

Link, Michael W., Michael P. Battaglia, Martin R. Frankel, Larry Osborn, and Ali H. Mokdad. 2008. A Comparison of Address-Based Sampling (ABS) Versus Random-Digit Dialing (RDD) for General Population Surveys. *Public Opinion Quarterly* 72(1) 6-27.

O'Muircheartaigh, C. A., S. A. Eckman, and C. Weiss. 2003. Traditional and Enhanced Field Listing for Probability Sampling. *Proceedings of the Survey Research Methods Section, American Statistical Association*.