URBAN SAMPLING FRAMES USING CITY DIRECTORIES

David W. Smith, Jennifer J. Thompson, University of Oklahoma
David W. Smith, College of Public Health, University of Oklahoma, PO Box 26901, Oklahoma City, OK 73190

Key Words: Sampling Frames, Coverage, Urban Samples

Our goal was to sample and interview teenagers and their parents living in the inner cities of Oklahoma City and Tulsa where telephone coverage is likely to be low. It is widely accepted that the ability to contact respondents using the telephone has greatly advanced health care research (Freeman, Kiecolt, Nicholls, and Shanks, 1982; Marcus and Crane, 1986). Telephone interviewing has reduced overall costs as well as offering the administrative convenience of a central locality for interviewers (Groves and Kahn, 1979). In general, non-telephone owners comprise a small proportion (5%) of national households and the impact of probable differences is slight for surveys of the general population (Robert Wood Johnson Foundation, 1987).

However, Marcus and Crane (1986) have recommended caution when the target population has low telephone coverage. Lower telephone coverage is associated with lower income and lower educational level; such sub-populations have been shown to have higher rates of nonresponse as well (Weaver and Glenn, 1975; Marcus and Crane, 1986). Non-telephone owners have been shown to differ with respect to certain characteristics or behaviors. For example, Corey and Freeman (1990) found that those persons living in households without telephones have a lower rate of health insurance (62%) than the national rate (87%) though the number of ambulatory visits and hospital experience did not differ.

Race is also related to varying telephone service rates. According to the Current Population Survey for May, 1999 (Census Bureau web site) telephone service rates are highest among Whites (95%) and Asian or Pacific Islanders (95%), lower among Blacks (89%), and lowest among American Indians (83%).

Sampling households in the inner city by random digit dialing, where telephone service rates are both low and concentrated in clusters, is also problematic. In the inner city areas of Oklahoma City (four zip codes) and Tulsa (six zip codes) targeted by our planned survey, telephone service rates range from 75% to 94% by zip code (United States Census, 1990). The telephone ownership rates in these inner city areas of Oklahoma City, 82%, and of Tulsa, 88%, are considerably lower than their respective Metropolitan Statistical Areas, both at 93%, Oklahoma, at 91%, and the nation, at 95%. Closer examination in Oklahoma City of telephone coverage by the 32 census tracts reveals wider variation. Examining the 27 tracts with over 150 households, four coterminous, concentrated tracts have telephone service rates under 67%, the lowest at 55%. Seven census tracts have 90% or higher telephone coverage, with the highest at 99%. The other 16 census tracts had rates between 67% and 90%. Not only is telephone service low in the inner city, it is highly variable by geography.

We evaluated city directories on CD-ROM, which have become available in the last few years, as possible sampling frames for these inner city areas, since they include households without telephone service or with unlisted telephones as well as households with telephones listed in the telephone directory. We evaluated their ease of use, since our goal was to substitute modern electronic technology as far as possible for enumeration of blocks and buildings by physically examining them. We also evaluated the coverage of the area by the CD-ROM directory listings, the proportion of dwelling units that are excluded from a directory. Finally, to evaluate efficiency, we measured the proportion of non-telephone listings that were probably unusable or potentially unusable.

City directories on CD-ROM could permit computerized random sampling of the listings they include, usually individual households and businesses. Telephone directories, on the other hand, include only the name, telephone number, and, usually, the address of individual people who have telephones and choose to have them listed in the directory. Historically, city directories had been available only in printed editions and they provided more extensive information, such as home ownership. These directories often included multiple indexes, that is, a listing by telephone number and a listing by street address as well as a listing by name, permitting users to cross-reference these items of information. Printed versions of these directories have traditionally been quite expensive, particularly compared with the telephone directories supplied free by telephone companies. Moreover, city directories were not updated every year, a significant barrier to their use in sampling. In recent years, a few city directories have provided listings on CD-ROM that include addresses for both telephone and non-telephone owners and even vacant addresses. These CD-ROM directories are updated frequently, sometimes twice a year or four times a year. Therefore, a city directory that includes non-telephone residential addresses could be used to sample inner city

areas without visiting the areas and listing residential blocks.

METHODS

We reviewed material on several directories supplied by producers, such as advertising brochures, and obtained two for further examination. This initial screening was based on the apparent ease of use, that is, whether we could extract or download all the listings in our target areas for processing with other software, particularly spreadsheets and statistical packages. This step was judged essential to drawing random samples, either simple random samples or stratified random samples.

The city directories usually show a telephone number where the residents have chosen to be listed in the telephone directory; listings which do not show telephone numbers are people who either do not have a telephone or have chosen to remain unlisted in the telephone directory or are addresses with no resident. Our further evaluation of the city directories focused on the listings without a telephone number.

To assess the eligibility of listings without a telephone number, we randomly sampled 1400 such listings in each of Oklahoma City and Tulsa for visual examination of the addresses, either by ourselves or by our research assistants. The sampled listings were examined for evidence of a building and, where a building existed, evidence of residential occupancy. Inhabitable buildings were classified in three categories: 1) probably occupied 2) for sale, for rent, or otherwise vacant, and 3) unable to determine if occupied. A wide range of evidence was accepted to classify a house as probably occupied, including children's toys on a porch, people in or around the building, or a car in the driveway. Buildings for sale or for rent can be occupied but may well be vacant and we combined them with a small fraction of clearly vacant buildings. All the rest were classified as unable to determine if occupied, though these could also be occupied.

Categories for the uninhabitable listings were developed upon visual inspection. These included 1) vacant lots, 2) no lot, with no space for a lot, 3) boarded up buildings, 4) businesses, and 5) condemned buildings. A listing was coded 'vacant lot-no house' when there was clearly a lot with space for a house but the entire lot was empty or had been converted into a parking lot. Buildings that were boarded up could not be inhabited, nor could those that were condemned as unfit for human habitation. A few buildings housed businesses and should not have been included as residential addresses in the directory. In a number of cases there was no room for an address, it would have to be sandwiched between two existing adjacent buildings or be behind another listed address, or it was at the end of a block with no space for a building. Also, listings were coded 'no lot-no space' when the house number from the directory was "off the block," that is the number was higher than the last house on the end of the block. Side entrances of corner houses were checked for such numbers.

RESULTS

Ease of Use

We obtained information on several CD-ROM directories. Only two of these appeared to have the necessary information for sample selection and were advertised as providing extensive capacity to download listings with names, addresses, and telephone numbers for processing in external software. Since we needed to sort and organize all the listings in order to sample from them, we rejected all but two directories as unusable. When we attempted to download the listings from one directory, we found that it did not permit more than 50 listings to be extracted at a time. We judged this restriction impractical for sampling and did not evaluate it further. The second directory permitted the unlimited extraction of listings for the geographic area included in our license and we examined its listings in further detail.

Total Residential Listings

The remaining selected CD-ROM city directory (Cole Digital Directory) provided 32,471 residential listings in the targeted area in Oklahoma City, somewhat larger than the 26,895 reported by the 1990 Census, and 32,471 in the targeted area in Tulsa, similar to the 1990 Census figure of 32,549. Many of the listings did not have a name, 44% in Oklahoma City and 40% in Tulsa.

Completeness of Coverage

We evaluated whether our directory listings showed good coverage. In the Oklahoma City area, we drew a sample of 30 city blocks with 336 listings by placing the entire area on a coordinate grid, choosing random coordinates, and selecting the block nearest to each selected point. We listed all the addresses, residential and business, on both sides of each selected block. We examined each block visually, noting any discrepancies, particularly buildings or street addresses that were not shown in the listings.

We identified six unlisted buildings with seven apparent street numbers, about 2% of the original 336 listings. A postcard mailed to each confirmed that four receive mail, others were returned. Consequently, the noncoverage rate was slightly over 1%. We found 21 additional apartments in two buildings that showed a listed total of 7 apartments. Thus, while buildings were well-covered, some apartment buildings might have

substantially more dwelling units than are listed on the CD-ROM directory.

Efficiency of Non-telephone Listings

Our examination revealed that there were a substantial proportion of occupied dwelling units in the non-telephone listings, approximately 73% for both cities were found to be inhabitable dwelling units, as shown in Table 1. Of the sampled listings, 52% in Oklahoma City and 53% in Tulsa were probably occupied dwelling units. We were unable to determine the occupancy for 18% of the listings in Oklahoma City and 15% in Tulsa. A small proportion were probably not occupied, 2% in Oklahoma City and 6% in Tulsa, since they were for rent, for sale, or otherwise appeared vacant.

Vacant lots, 12% in Oklahoma City and 17% in Tulsa, accounted for the largest proportion of uninhabitable listings. Boarded-up and condemned buildings, 4% in Oklahoma City and 3% in Tulsa, represented a noticeable proportion of the listings. Businesses, 2% in Oklahoma City and 1% in Tulsa, represent mis-classifications in the directory.

Table 1. Categories of Non-Telephone Listings by City (in percent) for samples of size 1400 in each inner city area.

Category	Oklahoma City	Tulsa
INHABITABLE		
Probably Occupied	52	53
Undetermined if Occupied	18	15
For Sale, For Rent, or Vacant	3	6
Inhabitable Total	73	74
UNINHABITABLE		
Vacant Lot	12	17
Boarded Up	3	2
Condemned	1	1
Business	2	1
No Lot, No Space for a House	9	5
Uninhabitable Total	27	26

Some listings could never correspond to inhabitable dwellings, 9% in Oklahoma City and 5% in Tulsa, since there is no corresponding lot or space for a house. We pursued the identification of a few of these in several ways. Some useful information was obtained from the local electrical company which informed us that some of those listings were serving electricity to the street lights, often at the end of the block, thus accounting for street numbers that were higher than any building on the block.

Uninhabitable listings accounted for 27% of the sample in Oklahoma City and 26% in Tulsa. The non-telephone listings comprised 44% of Oklahoma City and 40% of Tulsa. Consequently, 12% of the listings in Oklahoma City are not inhabitable and 10% of the Tulsa listings.

Effectiveness of Coverage

If we assume that all the telephone listings are inhabited and have telephones, we can compute the telephone coverage rate for the inhabitable listings in each inner city area. In Oklahoma City, the telephone coverage rate is 64% and Tulsa it is 67%. The coverage of the directory listings can be similarly computed on these assumptions as well over 99%. A random digit dialed survey of these inner city areas would miss about one-third of the population, virtually all of whom would be included in a sample based on the CD-ROM directory.

CONCLUSIONS

The evaluated CD-ROM city directory is effective for sampling inner cities with low telephone service rates. The directory shows good coverage and efficient listings of non-telephone households, so it is essential for sampling inner cities. The coverage of a random telephone survey of the same area is much lower. The CD-ROM we evaluated is convenient for selecting simple random samples of dwelling units using standard software. Others we evaluated were difficult to use and were not evaluated further. Other directories that we did not identify in the early stages of our evaluation may also be effective for drawing random samples of inner city areas.

The similarity of the results for Oklahoma City and Tulsa support the generalizability of our conclusions, so city directories are likely to be effective sampling frames for other cities.

Historically, city directories have been produced locally, probably because much of the effort to maintain and update the directories had to be done using paper records such as building permits. The nature of the industry appears to be shifting to more centralized producers using more automated methods to update their directories. While our results pertain primarily to inner

city areas and may not apply as well to rapidly growing areas where new roads and houses are under construction, automated and more frequent methods of updating city directories will provide better coverage of these areas. As these methods and the sources of information become more standardized, the quality of the information in city directories will become more uniform across cities.

It has been difficult to obtain information about the methods of updating city directories that are used by the suppliers since they likely view these methods as proprietary and confidential. Consequently, empirical validation of the quality of the directories will remain essential.

The products available as city directories have changed in the last year. The selected CD-ROM city directory will not be available beyond December 31, 1999 because the software is not Year 2000 compatible and a replacement directory has been introduced. We plan to evaluate this new directory in a similar way using our samples of 1400 non-telephone listings and 30 randomly selected blocks.

Acknowledgment: This work was partially supported by Grant Number U88/CC612534 from the Centers for Disease Control and Prevention.

REFERENCES

Corey, C. R. & Freeman H. E. (1990). Use of telephone interviewing in health care research. *Health Services Research*, 25, 129-144.

Current Population Survey. (1999). [On-line]. Available at: http://venus.census.gov/cdrom/lookup.

Freeman, H. E., Kiecolt, K. J., Nicholls, W. L., & Shanks J. M. (1982). Telephone sampling bias in surveying disability, *Public Opinion Quarterly*, 46:392-407.

Groves, R. M. & Kahn R. L. (1979). *Surveys by Telephone and a National Comparison with Personal Interviews*. New York: Academic Press.

Marcus, A. C. & Crane, L. A. (1986). Telephone surveys in pubic health research. *Medical Care*, 24, 97-112.

Robert Wood Johnson Foundation. (1987). *Access to Health Care in the United States: Results of a 1986 Survey*. Special Report. Princeton, NJ: Author.

Weaver, C. N. & Glenn N. D. (1975). Some characteristics of inaccessible respondents in a telephone survey. *Journal of Applied Psychology*, 60, (2), 260-262.