The Accuracy of Small Area Sampling of Wireless Telephone Numbers

Martin Barron[1], **Felicia LeClere**[2], **Robert Montgomery**[3], **Staci Greby**[4], and **Erin D. Kennedy**[5]

[1]Institution: NORC at the University of Chicago

[2]Institution: NORC at the University of Chicago

[3]Institution: NORC at the University of Chicago

[4]Institution: Centers for Disease Control and Prevention

[5]Institution: Centers for Disease Control and Prevention

Abstract

The value of telephone surveys for assessing effects at geographic areas is impacted by increasing wireless telephone use. Highly accurate landline samples may be drawn for national, state, county, or even smaller areas; however, wireless samples have less geographic precision. This requires additional data collection effort and screening costs in order to ensure the appropriate geographic area is surveyed. In this paper, we examine the accuracy of wireless sample frame from the 2010–2011 National Flu Surveys. We illustrate differences and variations in wireless sampling accuracy for different geographic areas, focusing on variability by area in placement of wire centers related to residences. Our results suggest that the accuracy of wireless sampling may be dependent on differences in geographic areas with the accuracy of wireless sampling decreasing as the level of geographic aggregation gets more specific; landline accuracy remains relatively stable regardless of geographic specificity. To explain this phenomenon, we examine patterns of geographic dispersion of wireless telephone numbers related to telephone switch centers and geographic area. Based on the evidence from these surveys we present several options to estimate the geographic specificity of an area prior to sampling.

Introduction

Many surveys require estimates at the national, state, county, or local areas. To calculate these estimates, telephone surveys require a match of telephone numbers to the geographic area. But differences in how geography is assigned to landline and wireless telephone numbers can lead to very different levels of accuracy.

Landline and wireless sampling frames are, in principal, constructed in a similar manner. Switch centers are part of the telephone system's infrastructure to efficiently route calls from sender to receiver. Each telephone number is assigned to one switch center based on geographic location; the switch center remains assigned to the telephone number. Since the geographic location of each switch center is known, survey researchers assign an

approximate geographic location to telephone numbers associated with each switch center. However, landline numbers are assigned to a particular location that rarely changes and may be assigned to the switch center closest to the location of the home. The switch center location serves as a relatively accurate proxy for landline telephone location (Marketing Systems Group 2012). Wireless numbers are mobile and assigned to the switch nearest the store where they are purchased, which is not necessarily near the respondent's home (Marketing Systems Group 2012). This makes assigning a location to a wireless telephone less accurate than assigning a location to a landline. Additionally, there is variability by area in placement of wire centers vs. residences, which may affect also the accurate assignment of a location.

There is limited research on the consequences of including wireless phones in the construction of geographically specific sampling frames for random digit dialing (RDD) surveys (Christian et al. 2009; Dutwin et al. 2011; Skalland et al. 2012). The challenge of sampling small geographic areas for dual-frame RDD surveys (that is, surveys that randomly select sample from two sampling frames, in this case landline telephone numbers and wireless telephone numbers) using switch center assignment has not been addressed. We describe the consequences of using switch centers to make geographic assignment of wireless and landline sample lines in small areas on the 2010–2011 National Flu Surveys (NFS). We examined the proportion of telephone numbers sampled that actually belonged in the targeted geographic areas, showing the differences in the geographic accuracy between wireless and landline samples and of samples drawn at numerous levels of aggregation. We show how variation in state level switch assignment affected sub-state accuracy of assignment of sample lines to specific geographic areas.

Methods

This research uses data from the NFS sponsored by the Centers for Disease Control and Prevention. The NFS was a large (73,203 completed interviews) RDD survey targeting households with landline and wireless telephone service. Data were collected between November 1–14, 2010, and March 3–30, 2011, to provide in-season estimates of influenza vaccination coverage and influenza knowledge, attitudes, and behaviors for national and 20 selected local areas. The local areas were county clusters, individual counties, or sub-county areas (Appendix A). The data from both surveys were combined in this analysis.

Separate sampling frames were constructed by dividing the universe of telephone banks into mutually exclusive banks of landline and wireless numbers. A sample was drawn from each of the 20 local areas with the goal of completing 280 wireless and 1,120 landline interviews in each area. A 21st sampling area consisted of all U.S. areas other than the 20 local areas which, when combined and properly weighted with the local areas, allowed calculation of national vaccination coverage estimates.

All wireless sample lines were screened for the wireless-only/mainly status of the household. Wireless only households were households where the respondent reported that he or she only had wireless service. Wireless-mainly households were households where the respondent reported the presence of both wireless and landline service, and it was unlikely

that anyone in the household would pick up the landline if it rang. Wireless-only/wireless mainly households remained in the final sample. All other wireless households where respondents reported that someone was likely to pick up the landline if it rang were screened out of the final sample.

All respondents were asked their residential mailing address zip code. This was compared with the location of the switch center to calculate a geographic accuracy rate. We defined the geographic accuracy rate as the proportion of all respondents with a self-reported residential zip code that is within the original specified sampling area as determined by the switch location. This was used as a measure of the proportion of the sampled and interviewed households that were actually located in the geographic areas used for estimating survey statistics. For the purposes of determining geographic accuracy, we excluded the cases sampled in the sampling area outside the 20 local areas, as those cases were not selected in a way to make geographic comparisons meaningful. We recalculated the geographic accuracy rate at different levels (Table 1) of geographic aggregation for our analysis below. That is, we calculate the accuracy of a given piece of sample assuming a broader geography than originally specified. For example, we may have sampled a case at the county level, but we can ask how accurate our sampling would have been had we sampled at the state level. In order to attempt to explain the geographic patterns seen in the wireless phone results, we mapped the movement of respondents from county to county based on sampled and self-reported zip code data. Maps were reviewed for discernable patterns.

In the 2010–2011 NFS, 20,071 wireless cases completed the interview, of which 18,470 provided their residential mailing address zip code. A further 53,132 landline cases completed the interview, of which 49,830 provided their residential mailing address zip code. The American Association for Public Opinion Research (AAPOR) Response Rate 3[1] (RR3) for the landline sample was 34.8 percent (November) and 35.5 percent (March). The AAPOR RR3 for wireless sample was 19.2 percent (November) and 19.3 percent (March). For both the landline and wireless-only/mainly samples, our analyses used cases where a respondent reported residential zip code was available. Appendix A gives the number of cases for each of the local geographic areas.

Results

Overall, 95.5 percent of the landlines sampled and 40.5 percent of the wireless-only/mainly households sampled were located within the sampled estimation area. Table 1 presents accuracy rates at different levels of geographic aggregation (including the original sampled area, which contains different levels of geographic precision) for wireless-only/mainly and landlines samples.

The accuracy of sample location decreased as geographic areas were more finely defined (Table 1). The decline was greater among the wireless-only/mainly population where 93.4 percent of the wireless-only/mainly cases were in their sampled Census Region but only

---

[1]AAPOR RR3 was calculated assuming e, the eligibility rate among sample with unobserved eligibility, was equal to the eligibility rate among cases with observed eligibility. This is frequently referred to as "CASRO" assumptions since the RR3 is equal to the CASRO response rate.

36.5 percent were in the sub-county area where they were sampled. In contrast, 99.8 percent of the landline cases were in the sampled Census Region and 95.2 percent were in the sub-county where they were sampled.

There was variation in the accuracy rates between the selected local areas as well (Table 2). The geographic accuracy rate ranged from 9.5 percent in New Hampshire to 75.7 percent in Minnesota. With the exception of the District of Columbia, all areas had at least 75 percent of their sample located within the sample state (77.1 percent to 85.7 percent with an average of 87.57 percent).

Several patterns in the geographic distribution of sampled cases surrounding the sampling targets using the distribution of the switch centers were identified. We illustrate using data from four areas: Tennessee, Maine, New Hampshire, and Cook County, IL.

In Tennessee (Figure 1) cases not located in the sampled area were clustered in bordering counties or nearby, with the largest concentration of the out of area cases in three nearby counties (red counties). Tennessee contained a number of switches within the county of interest, but none in the surrounding counties. Individuals with wireless service in an adjoining county had a higher probability of being assigned to a switch in Davidson County. (Similar patterns were seen in New Mexico and Texas.)

In Maine (Figure 2), cases showed little geographic clustering, and cases out of the sampled area were found throughout the state. All the switches in Maine were clustered around Cumberland County; thus, there was a wide distribution of cases sampled for Cumberland actually located in other counties. (Similar patterns were seen around Philadelphia.)

New Hampshire (Figure 3) was a unique example. Three Northern counties were sampled (Belknap, Coos, and Grafton), but there were no switch centers located in these counties. To sample these areas, wireless numbers were drawn from switches anywhere in the state. This led to a lower accuracy rate (9.5 percent).

Sub-county sampling units also posed problems for drawing accurate samples. Though switches exist in Cook County, IL, there were few switches located in the City of Chicago (Figure 4), the sampling target for the 2010–2011 NFS. This meant that sampled switches covered a larger area than otherwise desired. A great deal (30.6 percent) of the sample was discarded because the households were within Cook County but outside the City of Chicago.

Conclusion

Our results support the conclusion that landline sampling has greater geographical accuracy than wireless sampling. We found that smaller geographic units used for sampling resulted in lower geographic accuracy. While there was a substantial amount of variation between local areas in the accuracy of sampled addresses, some of the variation could be explained by the location and density of the switch centers in the geographic area.

When switch locations were examined, several patterns related to geographic accuracy of wireless sampling were observed. When switch locations were distributed evenly throughout

the state (such as in Tennessee), the geography accuracy of the original sampling strategy was high as subscribers were more likely to be assigned to a switch close to their residence. When switches were geographically concentrated or unevenly distributed throughout the state (such as in Maine), geographical accuracy decreased (Maine's in area accuracy was 25.3 percent compared to an overall in area accuracy of 40.5 percent). Additionally, in county or sub-county areas where no switches exist (such as New Hampshire and Chicago), geographic accuracy also is decreased. Thus, we conclude that to achieve a relatively high accuracy rate, a targeted area needed a cluster of local switch centers and additional switch centers distributed across adjacent areas.

Sampling wireless numbers at a sub-state level is possible but poses unique constraints. A geographically targeted survey that includes wireless sample should screen for the respondent's actual location and not rely on sampling information. In the 2010–2011 NFS, when specific small areas were targeted for the wireless survey, it was necessary to draw large oversamples to reach the desired number of interviews in the local area. When our sample target was large (e.g., a state or the entire United States) the geographic accuracy rates were roughly comparable to landline rates. Future work on the specific switch locations may yield some empirical methods for maximizing the accuracy of wireless samples. In addition, other approaches to determining geographic location, such as using billing zip codes (Dutwin 2014), appear promising. Future research should focus on the impact of differential residential mobility on geographic accuracy as respondents move to other locations bringing with them the wireless phones assigned to the original switch location.

Acknowledgments

Appendix A

Area Name	Definition	Wireless completes	Landline completes
Arkansas	AR: Arkansas, Ashley, Bradley, Chicot, Cleveland, Desha, Drew, Jefferson, Lee, Lincoln, Monroe, Phillips, Prairie, and St. Francis counties	1,070	2,193
AZ-Maricopa	AZ: Maricopa County	763	2,278
CA-Fresno	CA: Fresno County	661	2,355
CA-Los Angeles	CA: Los Angeles County	491	2,198
Colorado	CO: Denver, Jefferson, Adams, Arapahoe and Douglas counties	864	2,412
Connecticut	CT: New Haven, Hartford, and Middlesex counties	566	2,381
District of Columbia	Washington DC (NIS Boundaries)	915	2,485
Georgia	GA: Gwinnett and Fulton counties	1,258	2,820
IL-City of Chicago	IL: Chicago (NIS Boundaries)	907	2,280
ME-Cumberland	ME: Cumberland County	1,328	2,397
MI-Washtenaw	MI: Washtenaw County	990	2,638

Area Name	Definition	Wireless completes	Landline completes
Minnesota	MN: Anoka, Carver, Dakota, Hennepin, Ramsey, Scott, and Washington counties	543	2,346
New Hampshire	NH: Belknap, Coos, and Grafton counties	2,218	2,670
New Mexico	NM: Sandoval, Santa Fe, Bernalillo, and Valencia counties	569	2,505
New York	New York City: Bronx, Kings, New York County, Queens, Richmond	419	2,211
PA-Philadelphia	PA: Philadelphia (NIS boundaries)	720	2,222
TN-Davidson	TN: Davidson County	1,437	2,338
TX-Bexar	TX: Bexar County	688	2,418
TX-City of Houston	TX: Houston (NIS Boundaries)	887	2,409
WA-King	WA: King County	762	2,315

References

Christian, L., Dimock, L., Keeter, S. Accurately locating where wireless respondents live requires more than a phone number. Pew Research Center; Washingtion, DC: 2009.

Dutwin, D., Call, K., McAlpine, D., Beebe, T., Rapoport, R., Buttermore, N. Stratification of cell phones: implications for research. 66th Annual American Association for Public Opinion Research Conference; Phoenix AZ. 2011.

Dutwin D. Billing zip codes in cellular telephone sampling. Survey Practice. 2014; 7(4)

Marketing Systems Group. Construction of Cellur RDD Sampling Frames Based on Switch Location. 2012. Available at: http://www.m-s-g.com/CMS/ServerGallery/MSGWebNew/Documents/GENESYS/whitepapers/Cellular-RDD-Frame-Construction.pdf

Skalland, B., Khare, M., Furlow, C. Geographical accuracy of cell phone samples and the effect on telephone survey bias, variance, and cost. American Association for Public Opinion Research 67th Annual Conference; Orlando, FL. 2012.
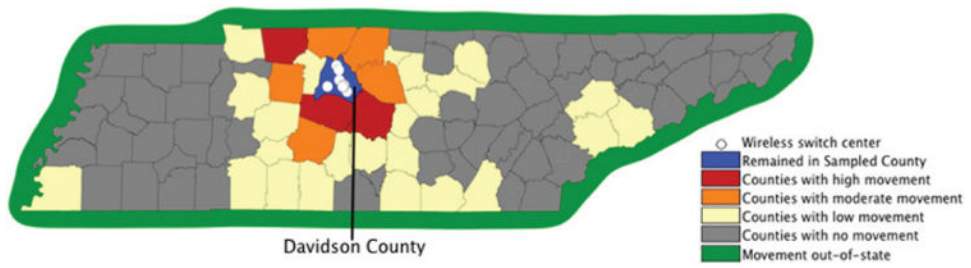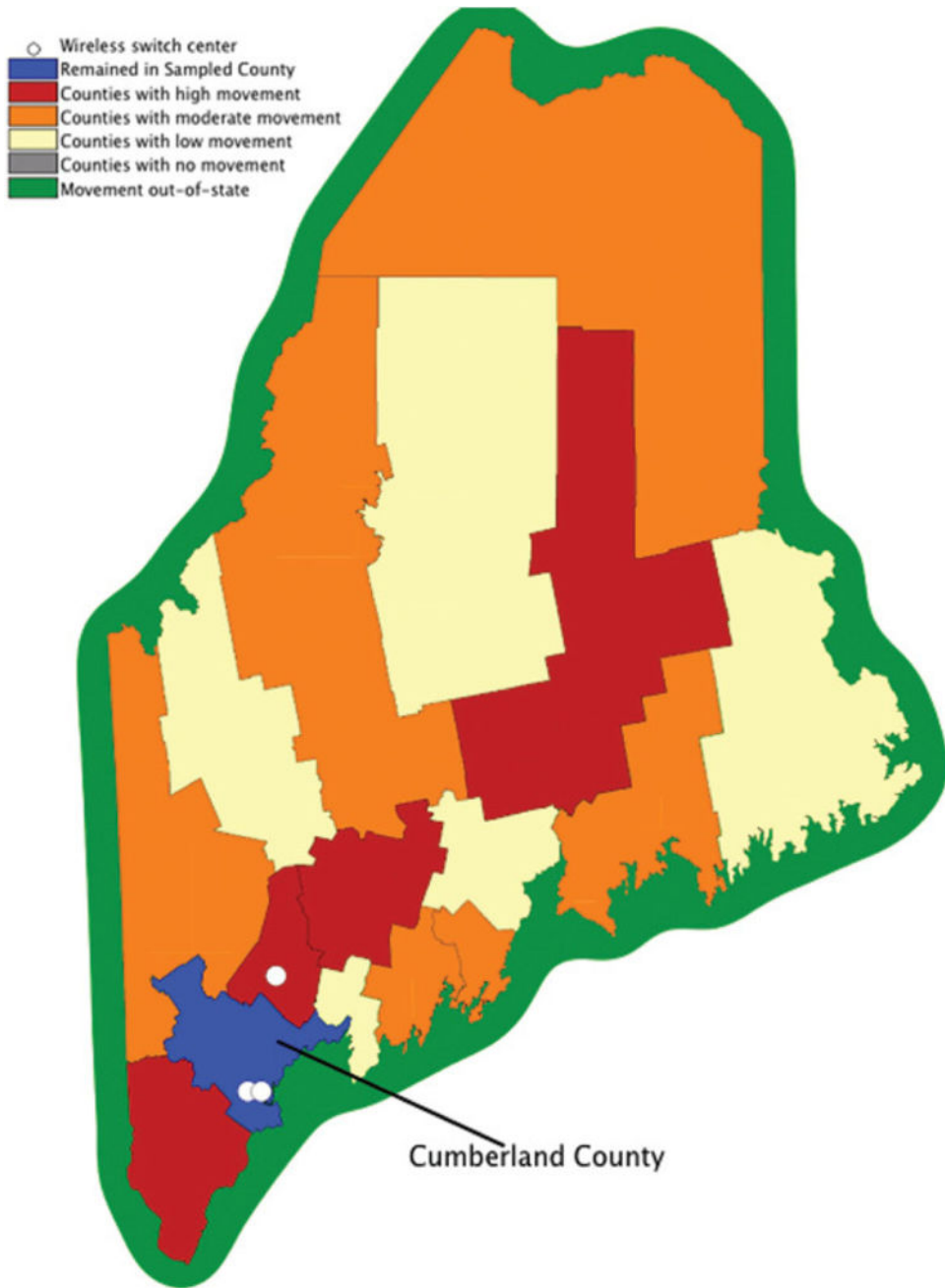
Figure 1.
The location of sampled cases in Tennessee.

Figure 2.
The location of sampled cases in Maine.
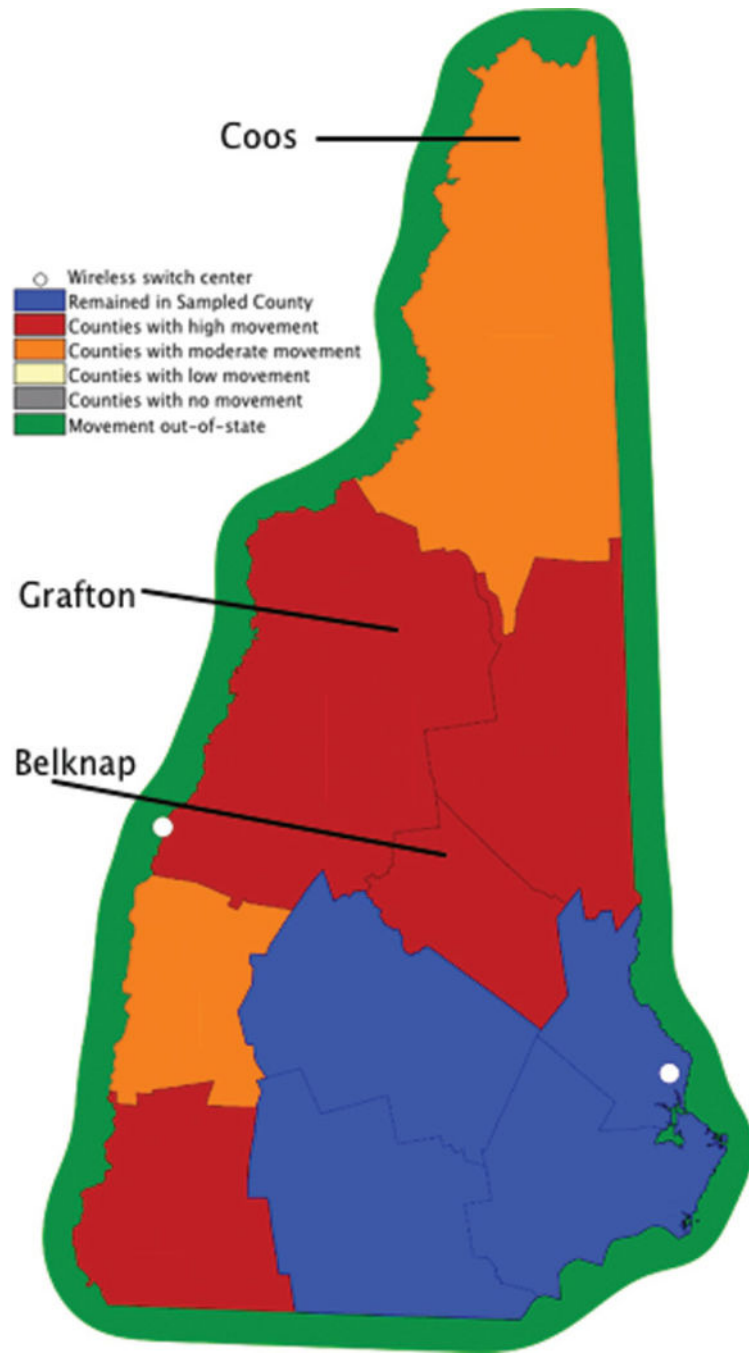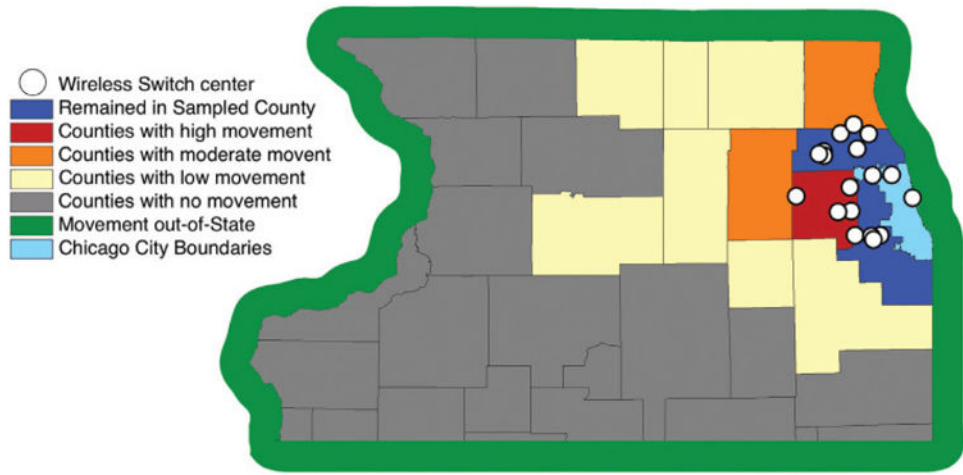
Figure 3.
The location of sampled cases and switches around New Hampshire.

Figure 4.
The location of sampled cases and switches around Chicago.

Table 1

Summary geographic accuracy rates, National Flu Survey, selected local areas, 2010–2011 influenza season.[a]

Accuracy rate[b]	Wireless-only/mainly	Landline
Census region	93.4%	99.8%
Census division	91.1%	99.7%
Bordering state	91.1%	99.7%
State	85.9%	99.5%
In-state, bordering county	65.8%	99.3%
County/county group	42.4%	96.0%
Original sampled estimation area	40.5%	95.5%
Sub-county (where appropriate)	36.5%	95.2%

[a] All geographic differences were significantly different between cell and landline samples (x2>2,738.3, DF=1, $p<0.001$).

[b] DC was included in all calculations.

Table 2

Geographic accuracy of the wireless-only/mainly cases by local area, National Flu Survey, 2010–2011 influenza season.

Area name	In area matched	Out of area			Total counts
		In-state	In bordering state	In other state	
Minnesota	75.70%	12.50%	3.30%	8.50%	543
New York	74.50%	8.60%	9.30%	7.60%	419
New Mexico	69.40%	20.00%	4.70%	5.80%	569
AZ-Maricopa County	65.30%	23.30%	4.10%	7.30%	763
CA-Los Angeles County	62.50%	29.10%	0.80%	7.50%	491
TX-Bexar County	59.60%	33.70%	1.20%	5.50%	688
WA-King County	53.70%	36.10%	1.70%	8.50%	762
Arkansas	52.10%	40.70%	5.00%	2.20%	1,070
Colorado	50.60%	38.90%	2.90%	7.60%	864
CA-Fresno County	48.30%	47.40%	0.60%	3.80%	661
Connecticut	46.10%	33.90%	8.50%	11.50%	566
MI-Washtenaw County	43.90%	34.50%	2.30%	19.20%	990
IL-City of Chicago	36.60%	51.60%	2.90%	8.90%	907
TX-City of Houston	36.50%	56.70%	1.10%	5.60%	887
PA-Philadelphia County	36.30%	44.20%	12.20%	7.40%	720
TN-Davidson County	33.30%	57.10%	4.80%	4.80%	1437
District of Columbia	31.60%	N/A	55.60%	12.80%	915
Georgia	26.40%	64.00%	3.00%	6.60%	1,258
ME-Cumberland County	25.30%	58.40%	1.10%	15.30%	1,328
New Hampshire	9.50%	67.60%	10.20%	12.70%	2,218

Surv Pract. Author manuscript; available in PMC 2018 April 17.