Discovering Rare Urban Populations for Community Health Surveys: Area Probability versus GIS-based Telephone Sampling Approaches

Ned English, National Opinion Research Center
Whitney Murphy, National Opinion Research Center

This paper examines disparate approaches used by NORC for the CDC-sponsored *REACH 2010* Risk Factor Surveillance Survey to interview rare populations in two urban settings. In the first instance an area probability methodology was utilized in 2002 to interview Cambodian households in Lowell, Massachusetts. Cambodians comprised a very small portion of the city's population, but even less so of the total households, which had a significant impact on data collection costs. The second round of the survey used a stratified telephone sampling procedure which is compared to the field method. In the second instance a GIS (Geographic Information Systems)-based telephone sampling approach was utilized to identify African-American and Hispanic households in Hillsborough County, New Hampshire. These two populations together comprised 3 percent of the total county households. Using a sophisticated sampling method, we were able to realize a racial eligibility rate of 11 percent among sample households which allowed telephone interviewing to be conducted instead of a more costly field interviewing approach. This paper examines the cost and effectiveness of each approach and discusses their applicability to discovering rare populations in similar urban settings.

I. Introduction

If not confronted, the issue of rare populations can both inconvenience survey production by extending field periods and affect survey results in unpredictable ways. Traditional area probability or telephone-sampling approaches can become prohibitively inefficient in cases of rare populations because extensive screening is required to identify eligible respondents. GIS, or geographic information systems, provides an advantage in that it can permit the linkage of data sets based on geographic variables, such as census or postal areas, and so facilitate targeting special populations with *a-priori* demographic information. GIS can thus enable sampling methods otherwise unfeasible and potentially increase screener efficiency.

This paper focuses on two communities in the *REACH 2010* study, a project funded by the Centers for Disease Control and Prevention to understand how race and ethnicity affect community health. Both communities presented special challenges to sampling and screening in that the target populations—Cambodians in Lowell, Massachusetts, and African Americans and Hispanics in Hillsborough County, New Hampshire—were extremely rare and not always well documented.

The purpose of this paper is to examine how rare populations can affect results in public health surveys in terms of response rates, dispositions, and actual response data. It does so by comparing two rounds of results in Lowell, Massachusetts, in which round one was conducted via face-to-face interviews and round two via the telephone using a GIS-enhanced stratified sampling method. This paper also discusses how a GIS-based sampling method was utilized to increase screener efficiency in the New Hampshire study area under even rarer population circumstances. The overall theme is that the issue of rare populations can be overcome from a number of different angles.

II. Background and Problem

In recent decades, a considerable amount of social research has been dedicated to understanding the particularities of rare or special populations, often discrete ethnic groups (Blair 1999, Sudman et al. 1988). It has often been the experience that sampling rare populations is both technically and administratively challenging, as screening costs can exceed those from interviewing (Blair 1999, Sudman et al. 1988). This problematic situation is exacerbated in instances where no reliable frame or list exists, which is usually the case with rare populations not enumerated by the census (Wells et al. 2003, Tourangeau and Smith 1985). Having to screen large quantities of sample for eligibility in a special population can render a survey completely unfeasible, and so certain probability-based techniques have been developed for greater efficiency in these circumstances (Tourangeau and Smith 1985).

Various probability-based methods of selection are often employed to sample rare populations due to their benefits of providing estimates and sampling variances without having to interview more than a fraction of the population at-large (Sudman et al. 1988). At a high level, this advantage is due to each member of the target population having a known non-zero chance of being selected, permitting generalization to the population with weights (Blair 1999, Sudman et al. 1988). One common probability-based method is *cluster sampling*, where multiple interviews are completed in

proximate segments. Cluster sampling reduces costs by minimizing travel, but increases sampling variance in comparison with a simple random sample of the population of interest (Sudman et al. 1988).

Other methods often used for rare population sampling include *network sampling*, in which respondents are questioned about relatives or friends that are also members of the target population; *two-stage telephone sampling* in which geographically-clustered 100-banks containing some target population are discovered and then over-sampled in comparison to the overall sample (Blair 1999, Sudman et al. 1988, Tourangeau and Smith 1985); *piggybacking*, in which data from one survey are used as screener information for another; and *adaptive cluster sampling*, where, when a rare population member is discovered, additional units are added to the sample from its immediate neighborhood (Thompson 1990). All of these listed techniques for sampling rare or specialized populations trade sample costs for design effect or reduced precision in estimates.

In a geographically targeted RDD survey, a sample of telephone numbers is selected from an area's exchanges, which produces a probability sample of residential telephone numbers when businesses and disconnected numbers are removed (Freeman et al. 1982). It has been found that telephone surveys are substantially less expensive than face-to-face surveys, which is exaggerated in cases of rare populations requiring large sample sizes (Freeman et al. 1982). Even in less-expensive telephone surveys, there is the issue that errors in estimated hit-rates will have considerably more weight in rare populations than in more common populations (Blair 1999). One example used in Blair's (1999) probability study of a rare, undocumented population—urban gay males—was that if a target group is 50 percent of the population, a 3 percent error in that estimate increases the sample size by 6 percent. If the target population is only 5 percent of the population, however, a 3 percent error would double the sample size and any assumed costs and effort. It is thus especially desirable to improve the estimation accuracy in the case of rare populations. To do so, however, requires precise *a-priori* census or other demographic data and the means to process and understand it. Geographic information systems can be seen one avenue to meet these needs (Bond and Devine 1991).

GIS is defined as a combination of spatial database management and spatial analytical tools, along with computerized cartography, used to facilitate the accumulation and manipulation of geocoded objects (Goss 1995, Bond and Devine 1991). GIS has been commonly employed in social research to link attribute information, often from the census, to spatial or geographic information (Srinivasan 2002, Bond and Devine 1991). The ability to process large geographic and demographic data sets has developed considerably since the early 1990's with the overall industry-driven advances in computing storage and power, leaving us today with a large capacity for spatial data processing (Bond and Devine 1991).

One major advantage of GIS to survey sampling is that if the locations of the population of interest can be ascertained, it is possible to eliminate areas with no population a-priori by linking addresses to geographic areas (Sudman et al. 1988). If the available information is not complete or accurate, however, the methodology becomes more complex (Blair 1999). While much literature exists as to what GIS is and how it can be used in surveys, there is little information as to actual methods to assist in discovering rare populations, as will be explained herein.

III. Methodology

REACH 2010 is a multi-year, community-based program that targets racial or ethnic minority populations in study areas across the United States, focusing on particular health-priority areas. *REACH* communities can be either urban or rural environments, and focus on one or more of the following ethnic groups: African Americans, Native Americans, Hispanic Americans, Asian Americans, and Pacific Islanders. NORC has the role of collecting data in the various communities by conducting screening surveys and interviews through both face-to-face and telephone methods.

Two *REACH 2010* study communities in particular are the focus of this paper. The study in Lowell, Massachusetts targets Cambodians in this former industrial city of approximately 100,000 residents. The other study is concerned with African and Hispanic Americans living in Hillsborough County, New Hampshire, the most urban county in the state. While the targeted populations remained the same, the sample designs, and in one case mode of data collection, were changed between the first and second rounds in both communities. This section details the differences in methodology between rounds for the two communities.

Methodology: Lowell, MA

Cambodians in the city of Lowell, Massachusetts, were the targeted population for the study, and questionnaire materials were available in both English and Khmer. Based on census data, the expected telephone coverage for the target population was approximately 95 percent, which was considered

sufficient for telephone data collection. The community coalition, however, advised NORC to avoid telephone interviewing due to perceived difficulties in gaining cooperation with this cultural group. Thus, in the first round of data collection, interviews were conducted in person using a modified area probability method.

The Cambodian population constituted a relatively small percentage of the total Lowell population; approximately 6 percent of Lowell households were estimated to be occupied by Cambodians.[1] Consequently, traditional area probability sampling would have necessitated listing most of the city to provide enough Cambodian households for sampling, even with stratification. To keep costs at a manageable level, NORC devised a sampling plan whereby households were selected through a stratified "count and go" procedure. To simplify field work, the study area was divided into strata of relatively similar geographic units before sample selection. Within each stratum, a random sample of blocks was selected with equal probability. Every housing unit in the selected blocks was then automatically selected into the sample. Thus, in the end we had a clustered design, with housing units clustered by blocks.[2]

The theoretical advantage of the count-and-go design was that an advance listing of all addresses in the selected areas was not necessary, saving considerable costs. One disadvantage was that selected housing units were more highly clustered, increasing the design effect more than with traditional area probability sampling.

NORC also employed stratification to improve the overall sampling efficiency by selecting a larger proportion of the sample from areas of higher concentration of the target population. For this community, we stratified census tracts according to the percentage of Asians among the adult population (2000 census Cambodian percentages were not available at the time of the survey). The high-density stratum consisted of 11 census tracts in Lowell, and the low-density stratum included the remaining 15 tracts. Then, blocks were randomly selected within strata to yield sufficient interviews for the study. Screening interviews were attempted at all occupied housing units within the selected blocks.

Although this "count-and-go" procedure eliminated the need to list all addresses in the city, it was still very inefficient in practice. Many blocks in Lowell contained no Cambodians, and the time required to screen such blocks was enormous, especially for interviewers whose primary language was Khmer. Consequently, NORC made several additional design and operational changes after working the first 15 replicates, including: hiring experienced English-speaking interviewers to help with initial screenings, eliminating from the sample all blocks with no Asians according to the 2000 census (not available in the initial design stages), and only working selected blocks whose ratio of Asian population to housing units, per 2000 census, exceeded a certain threshold[3] or where field observation indicated a likely Cambodian presence. One key change was the introduction of a simple question at the beginning of the screener instrument that eliminated the need to roster all adults in non-Cambodian households; this was adopted after determining from the early data that mixed households were rare.

In-person data collection in round one was thus complicated by several challenges that the scarcity of the target population presented in a field environment. Further conversations with the community intervention group led to the determination that telephone interviewing could be feasible from a cultural perspective in future rounds of data collection. Consequently, a study was conducted before round two to analyze the possibility of gaining cooperation and collecting data via telephone. The feasibility study proved that this option was viable; NORC was able to collect data in Lowell via telephone interviews in round two.

Because Cambodians constituted a relatively small percentage of the total Lowell population, traditional RDD or even white-pages listed telephone sampling was deemed an inefficient design. NORC recommended that a dual frame sample design be used, drawing on both listed and unlisted[4] telephone frames for sample selection. Additional stratification techniques were employed to increase the efficiency of the sample: specifically, the listed frame was partitioned into a Cambodian surname targeted list and a non-Cambodian surname list, both of which

[1] This estimate is based on 2000 census data for Asian population (16%), Cambodian population (9%), and Asian householders (11%) in Lowell.

[2] After the Lowell sample blocks were selected, we learned that the University of Massachusetts was conducting a similar one-time survey for the Lowell Community Health Center. Wishing to avoid interviewing in the same blocks, we agreed to remove from our sample any blocks selected by the university.

[3] The threshold was lowered throughout data collection in order to obtain the required 900 completed interviews to a final level of 0.4.

[4] This frame was constructed by generating an RDD frame for the area and removing all listed telephone numbers.

were utilized. Because telephone sample vendors do not provide Cambodian surname lists, NORC created the targeted surname listed frame in-house.[5] Thus, for round two, numbers were disproportionately sampled from the three strata. A portion of the sampled telephone numbers were selected from the unlisted frame, a portion from the listed frame linked to known Cambodian surnames (the Cambodian surname list), and a portion from the remaining listed frame (the non-Cambodian surname list).[6]

Methodology: Hillsborough County, NH

The second REACH community characterized by a rare target population was Hillsborough County, New Hampshire, which contains the cities of Manchester and Nashua in the east and rural areas towards the west. African Americans and Hispanics/Latinos were the racial/ethnic groups targeted for the study, and questionnaires were available in both English and Spanish. Based on 2000 census data, the expected telephone coverage for the target population in Hillsborough County was approximately 96 percent for both African Americans and Hispanics, which was considered sufficient for a telephone survey.

Because of the scarcity of the target population in Hillsborough County (3 percent of households and 4.5 percent population, according to the 2000 census), the first round of data collection used a modified list design. Using this alternative to RDD telephone sampling, we took advantage of the geographically determinant nature of the target population to focus telephone interviewing at the block level. A set of census blocks containing high densities of the targeted population individually and representing 90 percent of the targeted population was selected. NORC calculated that calling numbers only in these targeted blocks would raise ethnic/racial eligibility from 3 percent to approximately 11 percent. It is not possible, however, to acquire telephone numbers at the census block level. Thus, NORC first purchased the entire frame of listed telephone numbers within the 81 census tracts that define Hillsborough County, New Hampshire. Because listed numbers have address information attached to them, NORC was able to geocode the numbers exactly to the address level and create a reduced list consisting only of telephone numbers within the targeted set of blocks. Sample numbers were then selected from the reduced frame.

NORC planned to use the same design for this community in the second round. However, because the entire frame was purchased in the previous round and only a small proportion of unused numbers remained, NORC decided to reduce the degree of sample overlap between rounds by supplementing the frame. As in round one, NORC ordered the listed frame and created a reduced frame consisting only of numbers whose addresses fell within the targeted set of blocks. In addition to the listed numbers, NORC also ordered a sample from the frame of unlisted numbers in New Hampshire. The unlisted numbers were then sent to a vendor that was able to provide address data for unlisted numbers using non-traditional, non-directory sources such as warranty cards, web surveys, and even pizza delivery records. On average, we receive usable address data for approximately 60 percent of unlisted telephone numbers from this vendor. Using this address information, NORC was able to geocode the addresses to the block level and create a reduced frame of unlisted numbers, similar to that of the listed numbers, from which a supplemental sample was drawn. The addition of the unlisted portion not only increased the pool of numbers from which the full sample is drawn, but provided for coverage of a population that was not covered in the first round of data collection, i.e., members of the target population with unlisted telephone numbers.

IV. Results and Discussion

This section outlines the differences between round one in Lowell, which was conducted face-to-face, and round two, which was conducted via the telephone. Comparisons are made through unit and item response rates as well as actual screener and survey results. No data are yet available for the New Hampshire study site as this paper is being written, and so it is presented as a novel methodology at this stage.

Response Rates

Unit and item response rates were calculated following data collection in Lowell using standard methods. The unit response rate is the product of the screener response rate and the interview completion rate; the screener response rate is the proportion of

[5] NORC constructed an electronic database of Cambodian surnames identified in the Lowell city census and verified by Cambodian interviewers. Approximately 10 percent of the full listed frame fell into the Cambodian-surname frame.

[6] While the listed numbers designated as non-Cambodian were expected to have a much lower racial/ethnic eligibility rate, the feasibility study did not provide enough evidence for us to assume that none of those numbers would belong to Cambodian households, so we did not eliminate these cases from the frame altogether. However, due to the lower expected eligibility rate, we expected (and obtained) a much smaller proportion of completed interviews from this stratum.

eligible households that completed a screening interview. The item completion rate for a questionnaire variable is defined to be:

$$\frac{\textit{the number of respondents answering the question}}{\textit{the number of respondents asked the question}}.$$

One caveat for the item response rate is that if the number of persons asked a question is small, the rate will be unstable. In general, the numerator does not include "don't know" and "refused" responses unless these are legitimate, listed responses for an item. There are exceptions to this rule, however, and each item's responses should be carefully interpreted.

The unit and item response rates, as well as the distribution of disposition codes, were compared for Lowell rounds one and two. As described in Section II, Lowell was originally a field site in which data were acquired through face-to-face interviews. Beginning in the second round, interviews were collected via telephone.

It is useful to compare results from the two rounds from a number of angles to determine how the differences in methodology affect particular parameters for these rare populations. Unit non-response is a notable measure of comparison because of the effects it can have on sample size and bias. Table 1 compares unit response rates between the field in round one and the telephone in round two. The 'screener overall' is the proportion of eligible households that completed a screening interview, while the 'unresolved' rate identifies the percentage of numbers to which calls are never answered and household status cannot be determined. Member completion rates were adjusted by subgroup and then combined.

Table 1. Comparison of Unit Response Rates Between Rounds in Lowell

	Lowell Round 1- Field	***Lowell Round 2- Phone***
Screener Overall	93.5%	39.2%
Screener Overall Without Unresolved Numbers	93.5%	44.6%
Member adjusted O class	93.5%	55.7%
Member adjusted F class	92.6%	63.8%
Member combined adjustment	92.8%	57.9%

As Table 1 shows, the response rates were considerably higher in the field than via the telephone, as expected. These results are indicative of the obvious fundamental differences between the two data collection methods. We do face the question of how the low screener response rates on the telephone affect coverage and bias in round two in Lowell. This is a salient issue, but one that is beyond the scope of this paper.

The second major comparison of this analysis was of item response rates between the two rounds. Response and non-response at the item level is an important measure because item non-response can significantly reduce the sample size for particular critical items and introduce bias. Item non-response, however, is often ignored, and only the overall number of respondents is reported.

In comparing item response rates, most items had very high rates in both rounds, with six measures having 100 percent. A question regarding adult vaccines for pneumonia had the lowest rates in both rounds, being 89.8 percent in round one and 89.6 percent in round two. The only performance measure item with a response rate that was significantly different between the two rounds was the percentage of women who had a pap-smear in the previous three years, which was significant at the 95 percent alpha-level. These results may imply that items considered to be particularly sensitive are easier to refuse on the phone then in a face-to-face situation, where there are more opportunities for an interviewer to develop trust with the respondent.

Screener Data

We were also interested in comparing the actual results in the screener and interview survey data. Several demographic variables were examined at the screener level in both rounds of data collection. Table 2 (see Appendix A) displays the differences in means, between rounds, for several demographic indicators for eligible households in Lowell. The t-values that represent significant differences at the 95 percent level are highlighted.

There were very few demographic differences between the samples in the two rounds, with the only significant difference being the number of selected adults. The selection procedure for adults within households was changed between rounds, however, and the difference is in the direction expected. While the mean number of adults aged 65 and older is not significantly different between rounds, the means look quite different, with twice the rate found on the phone as in the field.

Interview Data

Thirteen variables of interest, five presented as means and eight presented as proportions, were chosen for comparison at the interview level and are presented in Tables 3 and 4 (see Appendix A). The main criterion for variable selection was the

availability of sufficient data. In other words, if a question was intended to be answered by all or most of the respondent population, it was included in the analyses. Selected interview variables fell into two groups: demographic data (respondent age, income, weight, and education) and health data (physical and mental health status, disease status, smoking habits, eating habits, and exercise habits). Means are presented for the appropriate variables in Table 3, while proportions are presented for the remaining variables in Table 4. T-values representing significant differences are highlighted in both tables.

The means in Table 3 show some significant differences between rounds. In round two, respondents report more days per month that their physical and mental health were not good, though the number of days poor health kept them from activities is not different. Also, the respondents in round two are slightly older (40 versus 38), but mean weight does not differ.

Table 4 shows more differences between rounds. Firstly, the proportions in Table 4 portray slightly different demographic groups in the two rounds. A larger proportion of round two respondents are female (59 percent versus 53 percent) and have a lower family income level (64 percent versus 54 percent earn less than $25,000 a year) than in round one. Education data are less straightforward; a larger proportion of round two respondents have only an elementary education (26 percent versus 17 percent), but a larger proportion are college graduates (7 percent versus 2 percent).

In terms of health, the results in Table 5 suggest that round two respondents are healthier overall than those in round one. A larger proportion of round two respondents report excellent or very good general health status (17 percent versus 10 percent), eat three or more pieces of fruit per day (62 percent versus 48 percent), and get at least 10 minutes of moderate exercise at a time (77 percent versus 33 percent) than round one respondents. There were no differences in reports of vigorous exercise between the two rounds.

The differences illustrated in the previous tables may be explained in a few different ways. First, self-presentation tends to be stronger in person, so it may be more important for respondents to make a positive impression in person than on the phone. Second, although the eligible households appear demographically similar between rounds, it is likely that the two modes yielded different types of respondents. Field interviewers are more likely to contact a household during the day, when particular types of people are at home (e.g., stay-at-home moms, unemployed people, elderly people, etc.). The telephone center, however, was much more likely to contact people in the evening, which would cover more employed people away from home during the day. One would expect the telephone data, then, to reflect more working college graduates and working women, as exhibited by the data.

A third discrepancy between the two rounds is seasonality, which is known to affect factors such as exercise, fruit consumption, and cigarette smoking. The field survey was conducted between October 2001 and June 2002, in the fall and winter, while the phone was conducted between March and July 2003, in the spring and summer. In fact, the majority of the face-to-face interviews were conducted between October and January, the cold season in Massachusetts. Not unexpectedly, the colder face-to-face results reflect significantly less moderate exercise and fruit consumption, and significantly more cigarette smoking.

Finally, the costs associated with the two rounds, and modes, of data collection differ quite substantially. Although the fixed costs for field work in this community were relatively low due to the small geographic area, the overall data collection costs were higher in round one. In the first round, interviewers worked approximately 3 times as many hours in the field for every completed interview compared to the telephone in round two. Overall, the field was close to 4 times as costly as the telephone and required a full day longer on average.

The majority of this analysis has focused on the Lowell survey experience because New Hampshire was still in data collection as this paper was being written. The GIS-based New Hampshire methodology is presented as a means to discover rare populations that are enumerated by the census, which wasn't the case in the Cambodian Lowell scenario. It will be most interesting to compare the round one results in New Hampshire, conducted using the list sample, with round two results, which included unlisted telephone sample.

V. Conclusions

Overall, these results show that rare populations can be difficult to sample, survey, and completely understand no matter what the methodology. Our results contain differences, and one may not know which are closest to the truth. Even with the differences, however, we have shown that rare populations can be successfully surveyed and produce useful data, as most of the results were generally comparable between the two rounds. Our demographic results shown in Table 2, for example, were the same between rounds in Lowell with very different methodologies. The survey results summarized in Tables 3 and 4 do contain differences, both predictable as per seasonality, and

unpredictable, perhaps due to mode-effects. We do argue that the telephone may be preferable from a practicality standpoint in such cases, due to its relative efficiency advantage over face-to-face. The authors are planning to focus on the New Hampshire data when it is complete as an interesting comparison and source of future research in the near term.

VI. References

Blair, Johnny. 1999. A probability sample of gay urban males: The use of two-phase adaptive sampling, *Journal of Sex Research*, 36(1), pp. 39-44.

Bond, Derek and Paula Devine. 1991. The role of geographic information systems in survey analysis, *The Statistician*, 40, pp. 209-215.

Downey, Liam. 2003. Spatial measurement, geography, and urban regional equality, *Social Forces*, 81(3), pp. 937-952.

Freeman, Howard E., K. Jill Kiecolt, William L. Nicholls II, and J. Merrill Shanks. 1982. Telephone sampling bias in surveying disability, *Public Opinion Quarterly*, 46, pp. 392-407.

Goss, Jon. 1995. We know who you are and we know where you live: The instrumental reality of geodemographic systems, *Economic Geography*, 71(2), pp. 171-198.

Srinivasan, Sumeeta. 2002. Quantifying spatial characteristics of cities, *Urban Studies*, 39(11), pp. 2005- 2028.

Sudman, Seymour, Monroe G. Sirken, and Charles D. Cowan. 1988. Sampling rare and elusive populations, *Science*, 240, pp. 991- 995.

Thompson, Steven K. 1990. Adaptive cluster sampling, *Journal of the American Statistical Association*, 85(412), pp. 1050- 1059.

Tourangeau, Roger and A. Wade Smith. 1985. Finding subgroups for surveys, *Public Opinion Quarterly*, 49, pp. 351-365.

Wells, Yvonne, Walter Petralia, David De Vaus, and Hal Kendig. 2003. Recruitment for a panel of Australian Retirees, *Research on Aging*, 25(1), pp. 36-64.

Appendix A: Additional Tables

Table 2. Comparison of Demographic Screener Variables Between Rounds in Lowell

Screener Variable	*Mean-Round 1 Field (n = 593)*	*Mean-Round 2 Phone (n = 731)*	*t-value*
# Adults/ Eligible HH	2.85	2.77	0.50
# Selected Adults/Eligible HH	1.75	2.09	4.06
# F Selects/ Eligible HH	0.60	0.55	0.82
# O Selects/Eligible HH	2.19	1.99	1.40
# Adults 65+/HH	0.08	0.16	1.77
# Males/Eligible HH	1.31	1.26	0.55
# Females/Eligible HH	1.53	1.46	0.67
# Asians/Eligible HH	2.82	2.64	1.11
# Cambodians/Eligible HH	2.81	2.55	1.74

Table 3. Interview Results (Means) Between Rounds in Lowell

Variable	*n-Round 1 (field)*	*Mean-Round 1 (field)*	*n-Round 2 (phone)*	*Mean-Round 2 (phone)*	*t-value*
Days Physical Health Not Good (per month)	1,036	3.74	882	4.56	2.15
Days Mental Health Not Good (per month)	1,032	4.82	871	5.87	2.38
Days Poor Hlth Kept from Activities (per month)	484	6.96	506	6.38	0.60
Weight (lbs)	1,031	135.42	876	134.82	0.53
Age (years)	1,034	38.37	903	40.02	2.42

Table 4. Interview Results (Proportions) Between Rounds in Lowell

Variable	*n- Round 1*	*%- Round 1*	*n- Round 2*	*%- Round 2*	*t-value*
Gender:	1,040		907		
Male		46.64		41.25	2.40
Female		53.36		58.75	2.40
Education:	1,036		903		
Kindergarten or less		16.98		17.08	0.06
Elementary (G1-8)		16.76		25.79	4.85
Some High School (G9-11)		17.59		12.13	3.40
GED/HS Grad (G12)		34.26		26.84	3.56
Some College (1-3 years)		12.47		11.63	0.57
College Grad (4 years)		1.94		6.52	4.94
Income:	984		729		
<$25,000 per year		53.75		63.97	4.29
$25,000-$50,000 per year		33.18		27.58	2.51
>$50,000 per year		13.07		8.45	3.10
General Health Status:	1,040		905		
Excellent		4.49		7.56	2.82
Very Good		5.86		9.12	2.71
Good		49.01		48.21	0.36
Fair		28.80		20.57	4.23
Poor		11.84		14.43	1.69
Fruit Serving Index:	1,039		904		
None-<1 per day		5.34		5.59	0.24
1-<3 per day		46.94		32.46	6.60
3-<5 per day		32.62		37.30	2.16
5 or more per day		15.10		24.65	5.27
Smoking Status:	1,031		907		
Current smoker (every day)		21.13		15.83	3.02
Current smoker (some days)		6.30		7.01	0.62
Former smoker		6.33		12.28	4.48
Never smoked		66.24		64.89	0.62
Get Moderate Exercise 10 min	1,039	63.27	905	76.80	6.60
Get Vigorous Exercise 10 min	1,035	40.63	901	37.97	1.20