Are cell phone samples needed for studies of walking activity?

Ugo Lachapelle
Département d'études urbaines et touristiques
École des sciences de la gestion
Université du Québec à Montréal
Case postale 8888, Succursale Centre-Ville
Montréal (Québec) H3C 3P8
Office: R-4755
Telephone (514) 987-3000
Fax (514) 987-7827
E-mail: lachapelle.ugo@uqam.ca

Marc D. Weiner
Bloustein Center for Survey Research
Edward J. Bloustein School of Planning and Public Policy
33 Livingston Avenue, Second Floor
New Brunswick, New Jersey 08901
Telephone (732) 932-1900, x217
Fax (732) 932-1881
E-mail: mdw@ejb.rutgers.edu

Robert B. Noland
Alan M. Voorhees Transportation Center
Edward J. Bloustein School of Planning and Public Policy
Rutgers, The State University of New Jersey
33 Livingston Avenue, Fourth Floor
New Brunswick, New Jersey 08901
Telephone (732)-932-6812 x536
Fax (732) 932-3714
E-mail: rnoland@rutgers.edu

Paper Prepared for Presentation and Publication at the Transportation Research Board 91st Annual Meeting

Word count: 5,467 + Tables (3) and figures (1) (4*250 words=1000) = 6,467

ABSTRACT

The growth in cell-phone-only households represents a challenge for the collection of survey data. Cell-phone-only households have distinct socio-demographic characteristics, which may result in different travel behavior. To explore those differences, as well to investigate the impact of including a cell phone component in active transportation research, a representative sample of New Jersey households was surveyed with a random-digit dialing survey that included 1,200 completed landline interviews (800 based on a statewide sample, 400 from an oversample of Jersey City) and 311 statewide cell-phone interviews. The survey explored walking behavior and perceived characteristics of the pedestrian environment. Socio-demographic characteristics, the frequency of walking and home location characteristics are compared using Chi-square tests of significance between sample pairs well as multivariate analysis (ordered probit). Cell-phone-only respondents were typically younger and poorer, with a greater proportion of renters, carless households and minorities. We did find that cell-only households walked more frequently, but this was due to their distinct socio-demographic characteristics, not their cell phone use *per se*. The implication for any analysis of rates or trends in walking (and probably other travel behavior) is that these households must be included via cell-phone sample supplementing a landline sample.

Keywords: survey; cell phones; mobile phones; wireless; random digit dial; landlines; physical activity; travel; active transportation; demographics.

1. INTRODUCTION

The widespread introduction and growing reliance on cell phones poses a new challenge to random digit dial (RDD) surveys which are used for many transportation studies *(1-5)*. A growing share of households no longer own landlines and instead rely entirely on cell phones while others maintain a near-vestigial landline, and predominantly use their cell phone(s). First introduced to the consumer market in the early to mid-1990's, by 2005 cell-phone-only households represented 8.4% of the US population and have been steadily increasing *(3)*. By 2008, that number was estimated to be 20.2%, the equivalent of 41 million adults and 14 million children under 18. By 2010, cell-phone-only households represented 26.6% of American households *(6, 7)*. At an estimated 115 million American households, this amounts to 30,590,000 cell-phone-only households or, with a mean household size of 2.6 persons, this represents 79,534,000 Americans who cannot be contacted by traditional landline RDD survey contacting protocols *(8)*.

The research question addressed here is whether this cell-phone gap matters in terms of research on walking behavior and, if so, how? Pedestrian behavior has generated growing interest both in transportation and physical activity research. Health researchers are interested in the health benefits of an active lifestyle *(9)* while transportation practitioners are under pressure to reduce vehicle travel and find ways to increase pedestrian accessibility to destinations, usually through land use, urban design measures and other changes to the built environment *(9, 10)*. Walking is particularly important in urban settings as an access mode or as a link to public transit *(11)*.

Because the cost of conducting cell phone surveys is much greater than landline surveys by an order of magnitude, it is important to evaluate the extent to which landline samples capture a representative cross-section of the population, and, conversely, whether information relevant to the research question is lost by not including a cell phone sample. The objective of this paper is to compare different subsamples of an integrated-dual-frame random digit dial survey. Based on a two-year survey research effort the analysis here includes a comparison of socio-demographic characteristics, walking behavior and home location characteristics of three main subsamples: New Jersey statewide residents (n=800); an oversample of urban Jersey City residents (n=400); and a statewide cell-phone-frame (n=311) consisting of cell-only and cell and landline respondents.

A cell phone sample enables proper representation of the 18-30 year group that is typically under-represented in landline-only random digit dial phone surveys, but is more likely to walk than other age groups. Most importantly, this analysis helps researchers assess potential information loss when omitting cell phone components for transportation research sampling protocols.

Research questions and hypotheses

How do cell-phone-only respondents differ from other telephone survey respondents? Do they walk more frequently? Do they live in areas with distinct built-environment characteristics?

Based on previous research, it is hypothesized that cell-phone-only households are distinct in socio-demographic composition and that they walk more frequently than the population reached by landlines. Finally, because a higher proportion of cell-phone-only households are expected to be renters, they would

be more likely to live near more central areas such as central business districts (CBDs), and have greater access to transit.

Existing literature on trends in cell phone use is first reviewed, followed by a review of research on walking behavior. The sampling strategy and survey instrument are described and compared using univariate analysis. Results are further confirmed in a multivariate analysis. Implications for crafting research protocols and sampling designs for research in transportation are discussed in the conclusion.

Trends in cell phone use

While cell phones have been patented since the 1970's, their widespread use began in the 1990's and they became ubiquitously distributed in the past ten years *(12)*. According to the *Federal Communications Commission* (FCC), there were approximately 24 million cell phone subscriptions in the US in 1994, and 270 million in 2008 *(12)*. In 2010, this number surpassed 300 million, with 302 million subscribers. In 2009 only 14.9% of households had only landline service while 24.5% were cell-phone-only households *(12)*; Table 7.4).

The National Health Interview Survey (NHIS) has become the survey research community's leading reference resource for landline and cell-phone coverage estimates. NHIS interviews are conducted in person and thus reach the designated sample without reference to telephone ownership status, therefore completely disentangling the telephone status from the research protocol. In addition to substantive health data, the interviewer records whether the respondent's household has landlines and cell phones. Beginning in 2007, that survey started including a question on which phone household members used the most, allowing the identification of "cell-predominant" households *(7)*.

Based on this survey, as of 2010, approximately 60% of the United States' population has access to both a cell phone and a landline. One important reason for this is that individuals subscribing to landlines are often required to register a landline as part of an internet or cable deal *(13)*, whether they use it or not. Many report using primarily the cell phone, even when they have a landline. Over 24% of households with both cell phones and landlines were considered wireless *mostly* households. They made up nearly 15% of all U.S. households *(3)*. Cell-predominant-households, despite having a landline, receive most of their calls on their cell phones.

For the purpose of this paper, a cell-only household is defined as a household that does not have any means of telephone communication other than at least one cell-phone. In a recent survey conducted by the Pew Internet and American life Project, 23% of Americans were considered landline-free *(13)*. The proportion of cell-only households has also been growing at a fast pace. According to the Cellular Telecommunications Internet Association CTIA *(6)* cell-only households went from 8.4% of American households in 2005, to 26.6% in 2010. In 2008, cell-phone-only households consisted of 20% of households corresponding to 18% of the total population *(3)*. Together, cell-only and cell mostly households now represent nearly a third of all households in the US *(3, 14)*. Omitting such a sizeable proportion of the population from a sampling plan, especially if it is known to have different characteristics than the general population, may significantly bias survey estimates and may provide inaccurate estimates of the determinants of travel behavior.

Because of the lack of representation of cell only and cell-mostly households, as well as the exclusion process of certain numbers in list-assisted RDD, survey coverage may capture less than 70% of all households in the United States *(14)*. This has considerable implications on the statistical validity and reliability of the data. Coverage bias may exist if persons with and without landlines are different with respect to the variables of interest *(2)*.

Relative costs of landline versus cell phone sampling

One of the challenges in conducting surveys is to balance trade-offs between study costs and sample precision. Including a cell phone component considerably increases expenditures: a cell phone interview costs roughly two-and-a-half to five times the cost of a landline interview primarily because interviews take longer and, often, respondents are paid a cash incentive, ostensibly to compensate for the cost of the air time. Interviews take longer because they typically require more dialing to reach respondents, more screening time, have more quota failures (respondents who do not fit the study's inclusion criteria), and the sample frame of cell phone numbers is generally not screened for known business or out-of-service numbers so more numbers are dialed than a sample frame pre-cleaned of those nonproductive numbers *(14-16)*. All of these factors increase the "cost per completed interview" (CPI). If one is interested in a particular geographic area, costs can be even greater, as cell-phone users may have non-geographically defined numbers, especially if they have moved between regions and maintained the same cell-phone number.[1] A detailed comparison of CPIs *(1)* shows that the cost per sampled telephone number and the cost per completed interviews were respectively $10.85 and $64.25 for landline surveys, $5.79 and $74.18 for cell phone surveys, and $5.10 and $195.78 for cell-only households reached after screening. The difference in cost for completed surveys is striking and attests to the importance of balancing cost versus sampling and data collection precision when designing transportation studies.

A compelling illustration in the transportation literature of concern for proper survey sampling can be found in Sen et al., *(14)*, who compare two sampling strategies: active contact method (cell phone) with passive contact method (mail surveys) in terms of efficiency, data collection effort, response rate and costs per interview of different sampling strategies. Sen and colleagues found that cell phone sampling involved more data collection effort but in turn yielded a higher response rate. Their RDD cell phone survey reported 42% of cell-only households, and 58% of combined cell and landline households. The mail survey, however, reported 30% of cell-only households and 40% of cell and landline households. Hence, RDD cell phone samples were more likely to capture cell-only and cell-landline households than a mail survey. Mail surveys, on the other hand captured a more comprehensive coverage including cell-only, landline only, a mix of both, and no phone households. Data collection efforts for the cell phone survey were more extensive than for address-based surveys, and response rates were higher for cell RDD (19%) than mail survey (8%) *(14)*.

Cell-phone sampling in existing surveys: demographic implications

Since at least 2001, various government-funded surveys, relevant to transportation and walking behavior have incorporated cell-phone sampling to assist in dual frame (i.e., combined landline and cell phone)

[1] Similar issues apply to many voice-over internet protocols (VOIP) services, which may even extend numbers beyond international boundaries.

weighting. Two key questions are relevant: how have these major surveys adapted their sampling plans to accommodate the advent and rise of cell-phone-only and cell-phone-mostly households, and, whether it is useful to use the cell-phone-only and cell-phone-mostly categories as analytical categories in their own right. The evidence suggests that the inclusion of a cell-phone sample is specific to the research question at hand, and, telephone-use status is a proxy that captures the differential adaptation of various demographic segments to changing technology.

The National Household Travel Survey NHTS, sponsored by the Federal Highway Administration, is an extensive nationwide computer-assisted telephone interview (CATI) survey, which uses list-assisted RDD to collect data about the travel behavior of American households. To its credit, the NHTS was early in exploring the impact of cell-phones on survey research; it included questions on cell phone ownership for the first time in its 2001 field administration. In 2009, for the first time, it included a cell phone sample frame as a test of methods *(17)*. The survey team justified this inclusion by suggesting the need to understand if travel patterns of cell-only households were significantly different from households reached via their landlines. This sample allowed survey sponsors "to determine the feasibility of conducting the NHTS interview by cell phone, and also provided some data for research on the differences in demographic characteristics and travel behavior between households that have landlines and those that have only cell phones" *(17)*. This data and the results of any analysis are not yet publically available as it is still being analyzed by FHWA staff.

In the health literature, two large scale survey efforts implemented by the Centers for Disease Control and Prevention (CDC) continue to generate evidence on cell-phone use from a physical activity perspective: the National Health Interview Survey (NHIS) and the Behavioral Risk Factor Surveillance System (BRFSS).

Every three months, the CDC releases estimates for 15 key health indicators using the National Health Interview Survey (NHIS), a face-to-face interview survey that captures information on wide-ranging health and personal data; in 2003, the NHIS began to probe household telephone access and usage. Comparing health outcomes across the telephone use categories, wireless only-households were more likely to binge drink and smoke, but also more likely to report an excellent or very good health status and to engage in regular leisure-time physical activity. They were also less likely to have ever been diagnosed with diabetes *(3, 7)*. Given our knowledge of the demographics of cell only households, these variations are likely functions of the age distribution across the categories of telephone users.

Another major public health survey, the Behavioral Risk Factor Surveillance System (BRFSS) implemented a cell phone component in all states and territories in 2009. The BRFSS is a nationwide health survey with a physical activity component and different modules that can be added on at the request of states. In their comparison of the prevalence of obesity in the 2000 BRFSS and the 2000 National Health and Nutrition Examination Survey (NHANES), Yun et al. *(18)* suggested that the increase in cell-only households raised the need to reconsider the validity of the BRFSS contacting protocol to track trends in obesity. This prompted the CDC to expand their methodology to rely on dual (i.e., combined landline and cell phone) frame samples. The 2008 version of the BRFSS expanded the landline sample frame to a dual frame sample in 18 states as a pilot study *(16)* and moved to a full scale dual frame sample of all 50 states in 2009. Again, differences in health related behavior such as smoking, binge drinking and engaging in physical activity were found between cell phone users and landline

respondents *(16)* with cell phone users being significantly more physically active based on univariate analysis. Telephone usage category is a proxy for other more dominant demographic characteristics. Indeed, once controlling for socio-demographic characteristics, the relationship between phone use category and likelihood of active transportation was no longer statistically significant.

In their analysis of landline and cell phone samples of public opinion surveys, Link et al. *(1)* found that compared to landline only samples, cell-phone-only samples were more likely to be male, African American, Hispanic, under the age of 34, employed, of lower income, and not married. Zuwallack *(4)* found similar results in his dual frame survey sample; cell-phone-only households were younger, and had a higher proportion of minorities. These are some of the same groups that are typically underrepresented in landline surveys due to differential non-response, the lower propensity of low income population to answer surveys *(1)*. Similar findings are presented by Blumberg and Luke *(7)* in their analysis of the NHIS with the addition of renters, residents of the Midwest, and adults living with unrelated adult roommates.

From these efforts we suspect that effects ostensibly attributable to telephone use status are actually the function of socio-demographic factors, particularly age, housing, urbanicity, and employment.

2. DATA AND METHODS

Sampling

We conducted a two-year survey; in November of 2009, we collected 1,200 completed landline interviews, 800 from an area-code-proportional statewide survey of New Jersey households, and 400 from an oversample of Jersey City; in November of 2010, we collected 311 New Jersey statewide cell-phone interviews, drawn from a cell phone frame. The survey explored walking, socio-demographics and perceived characteristics of the pedestrian environment. Weather conditions were similar during both field periods.

The rationale for oversampling Jersey City was that more potential walk-accessible destinations are expected to be found in reasonable proximity in large urban centers such as Jersey City. This sample also provides another point of comparison to assess the statewide cell phone sample frame against an urbanized population. Basic eligibility criteria were defined as: being 18 years of age or older. Eligible participants for the Jersey City oversample had lived in Jersey City for more than one year. To be part of the cell phone sample, respondents had to have been reached on a cell phone. We use an integrated dual frame sampling; for our 2010 sample, we assume that households, for which a completed interview was obtained via cell phone, even if they have a landline, are cell phone predominant households. This is consistent with estimates from the NHIS that suggest approximately 25% of households with both cell phones and landlines predominantly use their cell phones *(3)*.

Response rates, calculated using the American Association of Public Opinion Researchers approach #3,[2] were 20.9 % for the 2009 statewide landline sample, 19.9 % for the 2009 landline Jersey City oversample,

[2] The AAPOR3 response rate was calculated for each sample under the following equations:

and 23.3% for the companion 2010 cell phone sample. The weighting schema was calculated separately for each sample using an ((age x sex) x race) function, and analyses were run with and without weights. Our sample closely matched US estimates for 2010 *(7)*. A Spanish language option was available and about 5% of all interviews were conducted in Spanish. The cell phone sample that we collected was limited in size due to budget constraints. It would have been preferable to obtain a larger sample to enable more sub-group analysis. However, this does not have any implications for the analysis that follows, which has robust and useful results.

Modeling and Analysis

The key dependent variable is the frequency of walking over the past month, coded into four categories from the six original possible answers: "More than once a day", "Once a day", "Several times a week" and "No more than once a week." The independent variables fall into two categories: (a) socio-demographic predictors, and, (b) location/built-environment measures. The household-reporting respondent, i.e., the informant, was asked to report her/his ethnicity, age, education and gender as well as household information including number of children, if any, housing type, rent vs. own, and car ownership. The household's self-reported total annual income was coded into five categories (see below). Working full time and going to school were also considered as dichotomous variables. Of particular interest, in light of the body of research on the enabling effect of built environments on walking *(9)*, were respondents' self-reported measures of 10-minute walk access to their municipalities' central business district (CBD) and to a public transit stop/station.

Using the screening questions, an indicator variable identifying the different subsets of the samples were created: "NJ statewide landline 2009", "Jersey City oversample 2009", "NJ cell phone with landline 2010" and "NJ cell-only households 2010". Each subset's socio-demographic characteristics and walking behavior are compared. Preliminary univariate tests of significance of difference using Chi-square tests for pairs of samples were conducted as follows: The Statewide landline sample vs. the cell-only sample; the Jersey City landline sample vs. the cell-only sample; and the cell and landline sample (dual service households in cell sample frame) vs. the cell-only sample.

Frequency of walking was then modeled in a multivariate framework using ordered probit models. Indicator variables for sample type were assessed while controlling for socio-demographic characteristics. The reference category was the statewide landline. A positive association between sample indicator and

---

$$RR3 = \frac{I}{[(I+P)+(R+NC+O)+e(UH+UO)]}$$

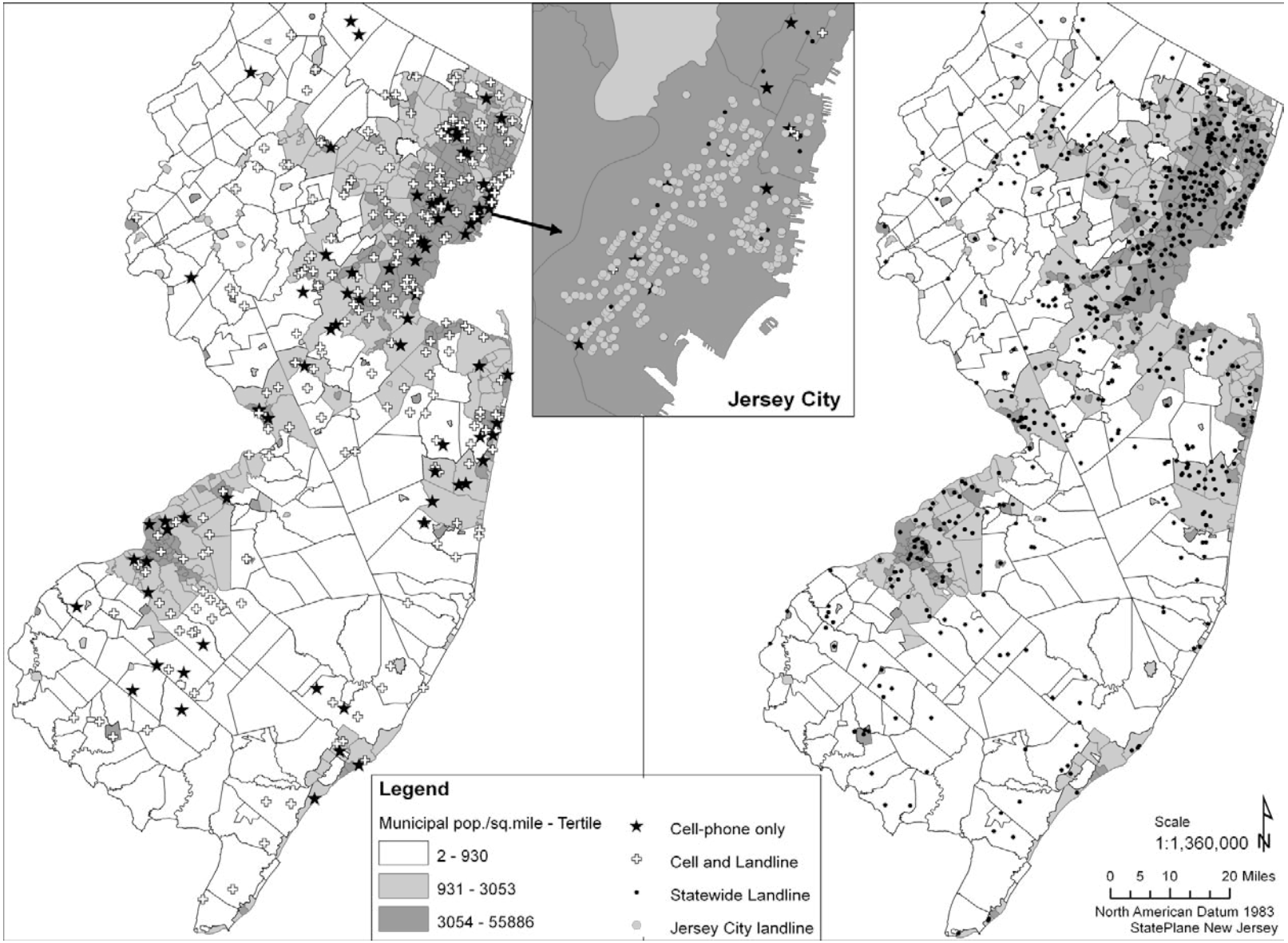
$$e = \frac{(I+P+R+NC+O)}{[(I+P+R+NC+O)+NE]}$$

where I=complete interviews (and screen-outs); P=partial interviews; R=refusals and break-offs; NC=non-contacts; O=other; *e*=the estimated eligibility of unknowns; UH=unknown households; and UO=unknown other and NE=not eligible *(19)*.

dependent variable would suggest that, once accounting for socio-demographic characteristics, being part of the cell phone sample drives up the mean walking frequency. Analyses were conducted using STATA 11 with and without survey weights; inclusion or omission of weights did not substantively affect results. Weighted estimates are provided. \

Respondents were asked to report the nearest intersection to their home, the municipality and the zip code where they resided. Using this information, we mapped completed interviews by subsample to visualize their distribution within the State of New Jersey. In Figure 1, a two-panel map shows that the statewide landline and cell sample respondents are generally well-distributed throughout the state, matching up with population density. Tertiles of population density are used as a backdrop to show where populations concentrate. Thus, visually there is no systematic variation in where these samples reside compared to the general population.

Figure 1: Map of samples



3. UNIVARIATE ANALYSES

The socio-demographic characteristics of each sample are presented in Table 1. Table 2 shows Pearson Chi-square tests of significance for pairs of samples. Statistical significance of differences between sub-samples across socio-demographics, housing and environment, and walking frequency are presented respectively in Table 2. The cell phone and combined cell phone/landline sample frame had respondents who were younger, renters, students, minorities and more carless households compared to the statewide landline sample.

Table 1: Sample description

	Landline sample frame		Cell phone sample frame		
	Statewide NJ	Jersey City	Dual household	Cell-phone-only	Total
	Percent	Percent	Percent	Percent	Percent
Renter	20	58.75	25.11	52.5	32.76
Minorities	32.25	71.25	43.29	57.5	45.6
Women	53.63	53.25	54.55	50	53.47
Have Children	35.63	39.75	47.62	45	39.05
Carless household	7.73	32.89	5.88	18.99	14.69
Household income					
Less than $25,000	12.72	27.61	6.15	34.85	16.86
$25,000 to <$50,000	18.92	22.39	17.95	40.91	20.89
$50,000 to <$100,000	34.82	26.69	36.41	15.15	31.83
$100,000 to <$150,000	17.33	9.51	22.05	4.55	15.3
$150,000 and more	16.22	13.8	17.44	4.55	15.13
Age					
18 to 30	7.12	15.18	28.18	44.74	14.62
31 to 40	14.79	21.95	15	18.42	16.92
41 to 55	31.64	28.46	33.18	23.68	30.61
56 to 70	27.95	23.04	19.09	11.84	24.37
71 and older	18.49	11.38	4.55	1.32	13.48
Education					
High school or less	24.77	32.47	25.66	40.51	27.79
Less than a college degree	26.58	21.13	28.32	35.44	25.89
College degree or more	48.65	46.39	46.02	24.05	46.32
Lives in single family home	73.5	18.5	69.7	38.75	56.52
Employed full time	46.63	50.75	54.98	42.5	48.78
Goes to school	1.25	4	7.79	10	3.44
Has CBD within 10 minutes walk	41.88	48.5	41.99	42.5	43.68
Has transit stop within 10 minutes walk	46.13	82.75	47.62	55	56.52
Frequency of walking					
Less than weekly	13.87	5.06	12.44	5.26	10.86

Several times a week	31.65	21.07	25.35	25	27.51
Once a day	23.11	20.79	17.05	17.11	21.2
More than once a day	31.37	53.09	45.16	52.63	40.43
Total	100	100	100	100	100

Table 2: Chi-square test of significance between pairs of samples

	Cell-phone-only households vs.		
	Landline statewide	**Jersey City**	**Cell and landline (dual service)**
	p-values	p-values	p-values
Renter	0.000	0.302	0.000
Minorities	0.000	0.015	0.028
Women	0.536	0.595	0.482
Have Children	0.097	0.383	0.686
Carless household	0.001	0.015	0.001
Household income	0.000	0.002	0.000
Age	0.000	0.000	0.038
Education	0.000	0.001	0.002
Lives in single family home	0.000	0.000	0.000
Employed full time	0.480	0.178	0.054
Goes to school	0.000	0.025	0.539
Has CBD within 10 minutes walk	0.914	0.327	0.937
Has transit stop within 10 minutes walk	0.129	0.000	0.255
Frequency of walking	0.002	0.832	0.332

Cell-only respondents had lower household incomes, had fewer households with children, were less likely to be women and were less educated than the other samples. The proportion of cell-phone-only respondents without a car was also considerably higher than the state average, but much lower than for the Jersey City sample. With respect to residential location, roughly the same proportion of cell-only households lived within a 10 minute walk of a CBD, as well as closer to transit stops or stations. They were also much less likely to live in single-family homes, as opposed to apartment buildings and other multi-family residences.

Are cell phone users actually more active than others, or is this relationship merely captured by differences in group composition? Table 2 provides Chi-square tests of significance for cell-only respondents paired with other subsamples. Cell-phone-only respondents walked considerably more

frequently than the landline sample, and about as much as Jersey City respondents or the cell and landline households.

In Table 2, column 1 provides Chi-square significance levels for a comparison between the cell-only sample and the New Jersey statewide sample. Both samples were not significantly different in terms of gender, employment status, having children, being employed full time and distance to transit and the CBD. The samples were significantly different on all other characteristics, including the frequency of walking.

Column 2 provides significance levels for a comparison between the Jersey City sample and the cell-only sample. Again, there were no significant differences between samples for gender, employment status, having children and percent renters. There were also no significant differences in the frequency of walking.

In column 3, the cell sample frame with landline is compared with the cell-only households. Gender, going to school and having children were not significantly different across the two categories within the cell phone sample frame. There was also no difference in the frequency of walking.

4. MULTIVARIATE ANALYSIS

The reported frequency of walking over the past month was modeled in a multivariate framework with results displayed in Table 3. Results of multivariate ordered probit regressions are presented for the entire sample for which all variables were available. We modeled the frequency of walking as a function of the variables that were significantly different between groups of interest. Each socio-demographic characteristic is tested individually along with the sample indicators and a final model combines all variables. Survey weights were used.

Table 3: Model estimates for the frequency of walking

WEIGHTED	model 1	model 2	model 3	model 4	model 5	model 6	model 7
Sample							
New Jersey landline 2009 [ref.]							
Jersey City	0.499*	0.423*	0.418	0.499*	0.499*	0.493*	0.471*
Cell sample with landline	0.236**	0.230**	0.219**	0.237**	0.238**	0.236**	0.238**
Cell only sample	0.336***	0.274**	0.291**	0.332***	0.338***	0.333***	0.271**
Renter		0.190*					
White non-Hispanic			0.171*				
Women				-0.06			
Have children					-0.035		
Carless household						0.026	
Household income							
Less than $25,000 [ref.]							
$25,000 to <$50,000							-0.01
$50,000 to <$100,000							-0.225*
$100,000 to <$150,000							-0.196
$150,000 and more							0.038
Age							
18 to 30							
31 to 40							
41 to 55							
56 to 70							
71 and older [ref.]							
Education							
High school or less [ref.]							
Less than a college degree							
College degree or more							
Lives in a single family home							
Employed full time							
Goes to school							
Has CBD within 10 minutes walk							
Has transit stop within 10 minutes walk							
	-	-	-	-	-	-	-
Cut 1 Constant	1.076***	1.037***	1.034***	1.107***	1.092***	1.074***	1.194***
Cut 2 Constant	-0.130**	-0.089	-0.086	-0.160**	-0.145**	-0.128**	-0.242*
Cut 3 Constant	0.400***	0.443***	0.446***	0.370***	0.385***	0.402***	0.291**
Observations	1062	1062	1062	1062	1062	1062	1062
ll (base)	-1536.43	-1536.43	-1536.43	-1536.43	-1536.43	-1536.43	-1536.43

ll (model)	-1525.47	-1522.17	-1522.49	-1525.04	-1525.33	-1525.45	-1519.58
Chi-square	21.9	28.5	27.9	22.8	22.2	22	33.7
Significance	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pseudo R2	0.007	0.009	0.009	0.007	0.007	0.007	0.011

Note for Table 3: model 1-7 on one page and model 8-15 on the following page.

The best way to visualize them is side by side.

Table 3 continued on next page

model 8	model 9	model 10	model 11	model 12	model 13	model 14	model 15
0.446*	0.503*	0.411	0.491*	0.494*	0.402	0.275	0.284
0.136	0.232**	0.235**	0.215**	0.233**	0.241**	0.128	0.143
0.201*	0.313***	0.287**	0.309***	0.345***	0.316***	0.082	0.11
						0.049	0.049
						0.081	0.063
						-0.057	-0.068
						-0.122	
						-0.159	-0.159
						-0.07	-0.044
						-0.237	-0.214
						-0.189	-0.17
						0.047	0.059
0.546***						0.483**	0.448**
0.238						0.232	0.181
0.266*						0.244	0.225
0.235						0.218	0.219
	0.005					0.036	
	-0.096					-0.06	
		-0.156*				-0.105	-0.112
			0.333*			0.088	0.141
						0.194	
				0.203**		0.108	0.109
					0.254***	0.169*	0.185*
-	-	-	-	-	-	-	-
0.833***	1.122***	1.194***	1.071***	0.994***	0.966***	0.977***	0.958***
0.115	-0.176*	-0.245***	-0.125**	-0.043	-0.016	-0.013	0.003
0.650***	0.355***	0.286***	0.407***	0.488***	0.518***	0.533**	0.547**
1062	1062	1062	1062	1062	1062	1062	1062
-1536.43	-1536.43	-1536.43	-1536.43	-1536.43	-1536.43	-1536.43	-1536.43
-1516.51	-1524.33	-1523.1	-1523.3	-1520.58	-1517.75	-1496.96	-1499.61

39.8	24.2	26.7	26.3	31.7	37.4	78.9	73.6
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.013	0.008	0.009	0.009	0.01	0.012	0.026	0.024

Using the statewide landline survey as a reference category, all three other samples, including the cell-only sample, were significantly positively associated with the frequency of walking in model 1. The largest coefficients, as expected were for Jersey City. The significant positive relationship of cell-only households was maintained even when introducing socio-demographic variables one by one in subsequent models (model 2 through 13). Model 14 provides estimates when all significant socio-demographic characteristics are included. Being a renter, all age groups below 71, and having a CBD and a transit stop within 10 minutes from home were all positively associated with more frequent walk trips when controlling for the sample types. Being a woman and living in a single family home were both negatively associated with the frequency of walking, but this was not significant. Model 15 only uses significant or theoretically important variables. In model 15, only the youngest age category, and the walking distance to a transit stop or station were still positively associated with walking frequency once controlling for other variables. The cell-phone-only coefficient was no longer significantly associated with the frequency of walking but remained positive, albeit considerably lower.

5. DISCUSSION

The objective of this work was to determine whether the socio-demographic characteristics and walking patterns of different sampling frames varied. This provides important insights into the potential measurement errors in phone surveys conducted without cell samples.

The analysis suggests that cell phone samples have distinct socio-demographic characteristics and walking patterns. They were from lower income households, were less educated and younger, more were renters, not living in single family homes, and they also tended to live closer to CBDs and to transit stops or stations. Differences in gender were not significant. In univariate analysis, they walked more frequently than statewide landline users, but not as frequently as the oversampled residents of Jersey City.

However, once controlling for socio-demographic characteristics, the walking patterns of respondents selected from a cell phone sample, whether they had a landline or not, were not significantly different from other respondents. Hence, the effect was largely driven by the different socio-demographic characteristics of the samples.

For the purpose of calculating inferential statistics on the correlates of walking activity these results suggest that not having a cell phone sample may be acceptable and should not overly affect estimates, provided there is adequate variation in the sample to capture the demographic distributions that would be collected were a cell phone sample supplement included. However, because of the difficulty of reaching younger and poorer respondents by traditional landlines, it will become increasingly difficult to generate adequate demographic distributions using landline only sampling. As such, if the purpose of a survey is to determine trends and rates across a population, supplementing a landline sample frame with a cell phone sample frame is necessary to accurately represent the population. Thus, this interpretation concurs with Hu et al. *(16)* that surveys carried out by telephone require a dual frame of landline and cell phone

numbers to provide reliable and representative estimates of rates, trends and prevalence. This is particularly important in research on pedestrians for two reasons: first, sampling of pedestrians is typically made harder by the fact that the incidence rates are relatively low, especially when surveys assess specific travel purposes *(10)*; second, because those more likely to engage in walking often have the same characteristics found in cell-only households.

Researchers should be cautious and particularly wary about using landline surveys to draw inferences about sub-populations that are more likely to be wireless only *(7)*. Because of the demographics of cell-phone-only households, research focusing on social disparities, and research on the health or transportation consequences of poverty should consider including cell-only and cell samples to more accurately capture these underrepresented groups.

In their essay on improving research on walking and bicycling, Krizek et al. *(10)* underscore the importance of clear conceptualization, sound research design, measurement innovation and strategic sampling. Whether a cell phone sample will be taken may be a strategic decision that researchers should consider carefully and approaches should be tailored to the different age group and income strata expected to participate in an activity. While caller ID features and voicemail may lower response rates of cell phones, some believe that in the long run, cell phones may make survey respondents more accessible to researchers *(5)*. Understanding the implications of this growing trend is necessary to conduct meaningful and representative survey research in this day and age.

6. CONCLUSIONS

As expected, cell-phone-only respondents were found to typically be younger, have of lower income, with a greater proportion of renters, carless households and non-white minorities. They also tended to walk more frequently than landline-using households. However, once controls for the socio-demographic characteristics of the cell-phone-only sample were included, the differences became non-significant. The distinct socio-demographic characteristics of cell-phone-only households are associated with more walking, but that cell phone users do not otherwise differ fundamentally in their residential location patterns or walking behavior.

While for descriptive and analytical purposes, a cell phone frame makes a properly drawn and executed probability sample more representative, the costs are not insignificant. Hence, researchers should carefully examine their research questions and sample inclusion criteria in light of available resources to make a firm determination of the necessity of including a cell phone component to an RDD landline sample. It will be necessary, particularly for studies targeting minorities and low income populations, and even more so in the future, as cell-phone household become dominant.

Cell phone use is a pervasive and growing trend that influences the way telephone survey sampling is conducted. Whether in transportation planning or in health research, researchers need to seriously consider the impact of omitting a cell phone sample supplement from the previously typical RDD sample survey. These findings suggest that in order to identify trends in the population or calculate prevalence of walking and other physical activity, except under limited circumstances, researchers should deploy dual frame samples to collect data from cell only, cell mostly, and landline telephone users.

ACKNOWLEDGEMENTS

Funding for survey design, deployment and analysis was kindly provided by the New Jersey Department of Transportation (NJDOT). Any statements or conclusions in this paper do not represent the policy or opinion of NJ DOT. Any errors or omissions are the sole responsibility of the authors. We thank M.P.Simon and M.K.Simon-Weiner for encouragement and support.

REFERENCES

References

1. Link, M. W., M. P. Battaglia, M. R. Frankel, L. Osborn, and A. H. Mokdad. Reaching the US Cell Phone Generation. *Public Opinion Quarterly*, Vol. 71, No. 5, 2007, pp. 814.

2. Blumberg, S. J., and J. V. Luke. Reevaluating the Need for Concern regarding Noncoverage Bias in Landline Surveys. *American Journal of Public Health*, Vol. 99, No. 10, 2009a, pp. 1806.

3. ———. Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, July-December 2008. *National Center for Health Statistics*, 2009b.

4. Zuwallack, R. Piloting Data Collection Via Cell Phones: Results, Experiences, and Lessons Learned. *Field Methods*, Vol. 21, No. 4, 2009, pp. 388.

5. Kempf, A. M., and P. L. Remington. New Challenges for Telephone Survey Research in the Twenty-First Century. *Annu.Rev.Public Health*, Vol. 28, 2007, pp. 113-126.

6. CTIA - The Wireless Association. U.S. Wireless Quick Facts. http://www.ctia.org/media/industry_info/index.cfm/AID/10323, Accessed May 31, Accessed on may 31, 2011.

7. Blumberg, S. J., and J. V. Luke. Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, January–June 2010. *National Center for Health Statistics, USA.Available Online at Http://www.Cdc.gov/nchs/nhis.Htm*, 2010.

8. U.S. Census Bureau. Quickfacts. quickfactscensus.gov/gfd/states/00000.html, last accessed June 26, 2011.

9. Transportation Research Board and Institute of Medicine. *Does the Built Environment Influence Physical Activity? Examining the Evidence. TRB Special Report 282.* , Washington, DC, 2005.

10. Krizek, K. J., S. L. Handy, and A. Forsyth. Explaining Changes in Walking and Bicycling Behavior: Challenges for Transportation Research. *Environment & Planning B: Planning & Design*, Vol. 36, No. 4, 2009, pp. 725-740.

11. Hoback, A., S. Anderson, and U. Dutta. True Walking Distance to Transit. *Transportation Planning & Technology*, Vol. 31, No. 6, 2008, pp. 681-692.

12. Federal Communication Commission, Industry Analysis and Technology Division Wireline Competition Bureau. *Trends in Telephone Service.* , 2010.

13. Purcell, K., R. Entner, N. Henderson, Pew Internet & American Life Project, and A.C. Nielsen Company. *The Rise of Apps Culture.* Pew Internet & American Life Project, 2010.

14. Sen, S., J. Zmud, and C. Arce. Efficiency and Effectiveness of Cell Phone Samples in Transportation Surveys. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2105, No. -1, 2009, pp. 44-50.

15. Miles, D. R., H. Herrick, and C. A. Ford. The North Carolina Child Health Assessment and Monitoring Program: Survey Methodology and Data Collection. *Statistical Primer*, No. 18, 2010, pp. 1-14.

16. Hu, S. S., L. Balluz, M. P. Battaglia, and M. R. Frankel. Improving Public Health Surveillance using a Dual-Frame Survey of Landline and Cell Phone Numbers. *American Journal of Epidemiology*, Vol. 173, No. 6, 2011, pp. 703-711.

17. U.S. Department of Transportation and Federal Highway Administration. *2009 National Household Travel Survey User's Guide.* , 2011.

18. Yun, S., B. P. Zhu, W. Black, and R. C. Brownson. A Comparison of National Estimates of Obesity Prevalence from the Behavioral Risk Factor Surveillance System and the National Health and Nutrition Examination Survey. *International Journal of Obesity*, Vol. 30, No. 1, 2006, pp. 164-170.