UC DAVIS
SUSTAINABLE TRANSPORTATION CENTER
of the Institute of Transportation Studies

Final Research Report S5-02

On-Line vs. Phone Surveys: Comparison of Results for a Bicycling Survey

November 2011

Yan Xing
Susan Handy

Institute of Transportation Studies
University of California, Davis
One Shields Ave., Davis, CA 95616

DISCLAIMER

ON-LINE VERSUS PHONE SURVEYS: COMPARISON OF RESULTS FOR A BICYCLING SURVEY

ABSTRACT

Researchers in the transportation field rely heavily on the traditional random-digit dialing phone survey and increasingly on on-line surveys. Many studies have looked at the strengths and weaknesses of the two survey methods with respect to the representativeness of the resulting sample as well as descriptive differences in responses to the survey questions. However, few of them have examined the inferential differences between the survey methods, for example, by comparing the coefficients of models of travel behavior estimated for each sample separately, to assess the degree to which the models yield consistent conclusions. In this paper we compare both descriptive and inferential results from on-line and phone surveys with identical questions conducted in Davis, CA. A split-sample approach was employed to examine the performance of models developed from the on-line survey data. Results show that although bicycling behavior does not differ across the two survey samples, many socio-demographic characteristics do. The models developed from each sample have several statistically indistinguishable coefficients but also notable differences in key explanatory factors. In addition, the models of bicycling behavior estimated with on-line data do not do a good job of predicting bicycling behavior as measured in the phone survey. Thus, the two survey methods in this case lead to different inferential results with different policy implications.

INTRODUCTION

For decades, travel behavior researchers and transportation planning agencies have relied on phone surveys as the primary means of collecting data on household travel patterns, especially for large-scale, general population surveys [1]. Although the speed and efficiency of phone surveys make it an appealing option, this approach is increasingly problematic. According to Dillman [1], as of 2000, about one third of the US population had unlisted phone numbers and 38% had answering machines (not counting those with voice mail); call-blocking was also common. More recently, the shift from land-lines to cell phones has made it more difficult than ever to achieve a representative sample.

With technological advances, on-line surveys offer an intriguing alternative, particularly given their relatively low cost [2,3] and the feasibility of multimedia content [2,4]. But for these surveys, too, sampling is difficult. For a general household survey, no complete sampling frame of email addresses yet exists. Instead, researchers have used letters sent via regular mail to recruit households to participate in the on-line survey. Because not all households have access to the Internet, they may be given the option of requesting a paper survey instead, but this puts an extra burden on the respondent and discourages participation. Non-response bias is thus a serious concern for on-line surveys, as it is for phone surveys.

But how do the biases compare between the two types of surveys? In this paper we compare results from an on-line and a phone survey on bicycling conducted in Davis, CA. The primary purpose of the surveys was to measure bicycling and potential explanatory factors. Although we did not directly examine the non-response bias in each survey, we compare the descriptive results for each survey as an indicator of differences in biases. In addition, we also examine the inferential differences between them by comparing the coefficients of models of bicycling behavior estimated for each sample separately, to assess the degree to which the models yield consistent empirical conclusions. Finally, we assess the ability of models developed from the on-line survey data to predict results from the phone survey.

LITERATURE REVIEW

Previous studies have compared the advantages and disadvantages of both phone and on-line surveys with respect to the representativeness of the resulting samples [e.g. 5, 6, 7]. These studies indicate that on-line surveys have poorer coverage of the general population than random-digit dialing (RDD) phone surveys. Additionally, good sampling frames of Internet users are usually lacking. On-line surveys in which the number of people surveyed is known, as is the case when participants are recruited by email invitation rather than by a blanket email request to a listserve, so that a response rate can be calculated, have lower response rates than is typical for phone surveys [2]. On the other hand, on-line survey method employs a self-administration approach that decreases the possibility of social desirability concerns in comparison to phone surveys. On-line surveys also enable respondents to control the time and pace of completing the survey questions themselves, potentially reducing the level of satisficing[1] and distraction compared to phone surveys [5, 6]. Visual presentation of the on-line survey also helps to reduce measurement error.

[1] Satisficing is related to task difficulty, respondents' ability, and motivation to do survey questions. The formulation of the probability of satisficing is expressed as: P(Satisficing)=a_1(Task difficulty) / (a_2(Ability)*a_3(Motivation)) [8].

Studies have also compared responses to on-line and RDD phone surveys, checking the similarity of descriptive characteristics of the samples with respect to socio-demographics as well as responses to factual and attitudinal questions. For example, Al-Subaihi [5] found that males were more likely to respond to an on-line survey than were females. Fricker et al [6] showed that the demographic characteristics of on-line and phone respondents who have access to the Internet do not differ significantly, a finding also confirmed by Vehovar et al [10]. These studies also found that responses to attitudinal questions that were not sensitive did not differ significantly across the two survey types. In another study, responses to factual questions and questions with two response categories, compared to multiple choice categories, were not significantly different after the two samples were weighted based on socio-demographic characteristics to match the population [9]. However, Greene, et al. found that phone respondents were more likely to provide socially desirable responses to personal lifestyle questions than were on-line respondents [11]. Several studies [11, 12, 13, 14] have found that the phone survey response are more towards the positive end of the scale for a variety of scale types than on-line responses. The evidence thus suggests that socio-demographics and responses to more straight-forward questions are less likely to differ than responses towards more subjective questions, particularly those with a greater range of response categories.

Few studies, however, focus on inferential differences between samples from on-line and phone surveys. Transportation planners rely on survey data, collected by phone or on-line surveys or other methods, to build models for predicting future transportation demand. The possibility that different survey methods would generate different predictions is thus a significant concern. However, prior studies have not measured differences in the predictive tendencies of models developed from data from the two different survey methods. Further, the performance of models developed from on-line data is unclear: evidence of the ability of these models to accurately predict the behavior of the population is lacking.

The purpose of this study is to fill this gap. We examine differences in models estimated separately with samples from an on-line survey and a phone survey, both by testing for the significance of differences in the coefficients of the models and by evaluating how well models estimated with on-line data predict bicycling behaviors for each individual in the phone survey sample, commonly believed to be more representative of the population than on-line survey samples. We aim to address the following research question through the analyses: do the two survey methods lead to different conclusions about which factors contribute and to what degree to explaining travel behavior?

METHODOLOGY

Survey Sampling and Administration

The on-line survey was conducted in Davis in early fall 2006. A recruitment letter was sent to a random sample of Davis households, purchased through a commercial vendor. Recipients of the letter had the option of completing the survey on-line or requesting a hard copy to be sent and returned via the mail. Two reminder postcards were sent. The final response rate in Davis was 18.8%, after accounting for bad addresses, yielding a final sample of 354. The survey showed that 78.0% of respondents owned or had access to a functioning bicycling and 53.0% bicycled at least once in the last 7 days. While vastly higher than the US average, these results seem plausible for Davis, the self-proclaimed “Bicycle Capital of the US.” However, only one reliable measure of bicycling was otherwise available for Davis, the share of workers usually bicycling to work, from the 2000 US Census, and comparing this value to that from the survey suggested significant bias: 23.7% of workers usually bicycled to work according to the survey, versus 14.4% according to the Census. It was unlikely that bicycling had more than doubled between 2000 and 2006.

To assess the accuracy of the results from the on-line survey, we hired a survey research firm to conduct a phone survey in Davis in 2008. The survey included a subset of the questions in the original on-line survey. The firm used random-digit dialing to achieve a sample of 400, representing14.7% of the phone numbers where a household member was reached.

Methods for Comparing Surveys

To examine the possibility that the characteristics of the sample and the relationships inherent in each could differ, we analyzed the data in two ways. First, the two samples were compared with respect to socio-demographic characteristics measured in the survey. A chi-square test was used to test whether there is a relationship between two categorical variables, and analysis of variance (ANOVA) to test differences in continuous dependent variables (e.g. Age) between the two surveys. We use a p-value of 0.05 as a cut-off for identifying statistically significant differences.

Second, to test whether the type of survey influences inferential results, we estimated models for each sample for “bicycled or not in the last 7 days” as well as “days rode a bicycle during the last week” as dependent variables. The models test several different potential explanatory variables: gender, age, income, work status, auto ownership, travel constraint, and a variety of attitudinal factors. According to the specific properties of the two dependent variables, a binary logistic model and a negative binomial model were chosen respectively. The statistical significance of the differences between coefficients of the two models was tested by employing the market segmentation method [15]. Using the type of survey as the segmentation basis, a chi-square test was applied to determine whether the coefficients collectively are significantly different between the segments:

$$-2[LL(\widehat{\beta}_{pooled}) - \Sigma_g LL(\widehat{\beta}_{segmented})] \sim \chi^2_{(G-1)K}$$

Where:
(G-1) K = degrees of freedom (df)
G = the number of segmented models
K = the number of variables in the pooled model

Even if coefficients collectively are not statistically different, some individual coefficients may still differ significantly. Thus we also checked for specific differences between coefficients for the two models by employing a t-test. The steps involved in carrying out these tests are described more fully in the Results section.

Method to Measure the Performance of Predictive Model

One of the main purposes of travel demand models is to forecast future travel. Therefore, a model’s predictive capability has long been a concern of researchers, who have developed a variety of validation methods. In the transportation field, models are commonly validated by applying the model to a data set other than that used in calibrating the model so as to check model estimates against observed data. The split-sample approach, in which one sample is split into two, with half of the data used for calibration and half for validation, is sometimes used.

In this study, we borrowed from this method by using the on-line survey data for calibration and phone survey data for validation. The best models developed from the on-line survey data were applied to the phone survey data, which we treated as observed data from a representative sample of the population, to measure their predictive capabilities.

There are several common measures to examine the accuracy or error of the predictions. For continuous dependent variables, the variance between estimated and observed values is often used to assess how well a model from one data set fits the observed behavior in another sample. One measure of variance is the mean squared error of validation (MSE), equal to the average squared error and calculated as the sum-of-squares of errors divided by the sample size. Acceptable values of this measure range from 0, indicating a perfect prediction, to 0.25, a cut-off value implying the worst acceptable prediction [16]. Root mean square error (RMSE), the square root of MSE, is often reported instead of the MSE, and they yield the same conclusions about the predictive performance of a model.

For discrete choice models, the success table [17] is used to measure how well a model preforms in predicting the choice of all alternatives. The success table indicates the extent to which the model misclassifies each alternative by comparing the predicted versus observed choice of an alternative for each individual in the sample. In this table, the predictive success of the model is summarized by a success index, the prediction success proportion normalized by the sample observed shares. A success index greater than one refers to good predictive success (the model predicts better than simply applying market share probabilities).

RESULTS

Descriptive Comparison

Comparison of Socio-demographics

Socio-demographic characteristics of the two samples were compared to data from the 2005-2009 American Community Survey (ACS) 5-Year estimates or the 2000 Census, the best available data on population characteristics (Table 1). The comparison shows that gender and car ownership distributions in the phone sample more closely resemble those of the census than do those from the on-line sample. Respondents older than 65 were over-represented in the samples from both surveys compared to the Davis population, at least in part because the survey sampling frames are likely to disproportionately exclude university students who live within the city. Both the phone and on-line surveys found a higher share of workers usually bicycling to work than in the ACS, and the on-line survey had a higher share than the phone survey.

TABLE 1 Comparison of Socio-demographics Characteristics of Survey Samples to Census Data

	Census (%)	Phone survey (%)	On-line survey (%)
Female	51.4 (1)	51.5	46.6
Percent 65 years and over among people who are 18 years and over	9.7 (1)	25.4	15.1
Percent of households owning a car	93.4 (2)	93.5	96.9
Percent biking to work	14.4 (2)	17.5 (3)	23.7 (3)

[1]2005-2009 American Community Survey 5-Year Estimates. They are based on data collected over a 5-year time period. The estimates represent the average characteristics of population and housing between January 2005 and December 2009 and DO NOT represent a single point in time.
[2]2000 census data.
[3] Percent of respondents who used a bicycle as the primary mode of travel to and from work at least 3 days in a typical week with good weather, which was calculated to match the percent of population who responded “Bicycle” to the census question “How did you usually get to work last week”.

The socio-demographic characteristics of the two samples are compared in Table 2. The statistics show that gender, household annual income level, years living in Davis, and working at least one day a week do not significantly differ between the two samples, whereas the mean age and age group, UCD student or not, employment status, vehicle ownership, and the share of respondents who have physical limitation on bicycle riding differ significantly between the two surveys. The on-line survey sample is characterized as having a significantly younger average age, higher percentages of UCD students, being employed, owning cars, and lower share of physical limitation on bicycle riding..

TABLE 2 Comparison of Socio-demographic Characteristics of Survey Samples

	Phone survey		On-line survey		p-value
Female	51.5%		46.6%		.183
Mean Household income($)[1] (std dev)	64,200	(.093)	66,600	(.095)	.363
Mean years living in Davis (std dev)	3.85	(.077)	3.71	(.087)	.235
Mean Age (std dev)	52.26	(.880)	48.46	(.848)	.002
Age Group:					.003
18-34	17.6%		20.8%		
35-54	36.5%		45.6%		
55+	45.9%		33.5%		
A student at UC Davis	9.3%		14.1%		.038
Currently employed (**Emp Status**)[3]	65.8%		76.2%		.002
Work outside of the home at least one day a week	91.3%		91.7%		.845
Vehicle ownership	93.5%		96.9%		.032
Physical limitation on bicycle riding (**Biking Limit**)[2,3]	14.7%		9.6%		.040

[1]1=”<$20K”; 2=”$20-40K”; 3=”$40-60K”; 4=”$60-80K”; 5=”$80-100K”; 6=”>$100K”. The mean was calculated using the mid-point of each range.
[2] “Do you have any condition that seriously limits or prevents you from riding a bike?” 1=Yes; 0=No.
[3] Bold in bracket is the corresponding variable name used in the following models.

Comparison of Bicycling Behaviors

The phone survey yielded slightly lower levels of bicycle ownership and use, but the differences between most measures of bicycling frequency between the on-line and phone survey were not statistically significant (Table 3). In both surveys, almost half of respondents reported that their last bicycle ride was within the last week, but beyond a week, on-line respondents reported more recent trips. However, the nature of bicycling differs between the two surveys. Respondents in the on-line survey report more days bicycling as their primary mode to and from work, and they are more likely to report that their bicycling is either all for recreation or all for transportation.

TABLE 3 Phone Survey vs. On-line Survey Results for Davis Bicycle Survey

	Phone Survey	On-line Survey	P-values
Share bicycle ownership	76.3%	78.0%	0.576
Share biking in last 7 days	47.0%	53.0%	0.101
Share biking within last year	72.5%	74.1%	0.630
Share biking to work	29.5%	32.3%	0.502
Days biking within last week (std dev)	1.79 (.119)	2.04 (.131)	0.157
Primary purpose for taking the last bike ride[1]			0.166
Transportation to or from work or school	32.2%	31.7%:	
Recreational --for pleasure or exercise	46.7%	41.1%	
Last time rode a bike[2]			0.005
Within the last week	49.0%	49.7%	
Between one week and one month ago	7.9%	13.6%	
Portion of bike ride for transportation and recreation[3]			0.009[4]
All for transportation	19.7%	16.7%	
All for recreation	21.1%	12.6%	
Days biking as primary mode to or from work (std dev)	1.54 (.147)	3.54 (.186)	0.000
Number of respondents	400	354	

[1]1=Transportation to or from work or school; 2=Transportation to a friend's house, store or other destination; 3=Recreational -- for pleasure or exercise.
[2]1=last week; 2=one week to one month ago; 3=one month to one year ago; 4=one year to 10 years ago; 5=more than 10 years ago; 6=never.
[3]1=all for transportation; 2=most for transportation; 3=half for each; 4=most for recreation; 5=all for recreation.

Comparison of Attitudes

In addition to reporting their bicycling behaviors, the respondents in both surveys also answered several attitudinal questions. The results show significant differences in perceptions of the bicycling culture in Davis (Table 4). More people in the phone survey agree that "Bicycling is a normal mode of transportation for adults in this community."

Almost all of the attitudes toward physical exercise and travel modes differ significantly across the two samples. A higher percentage of the respondents in the phone survey agrees that physical exercise is important or that they enjoy physical exercise than in the on-line survey. A higher share of respondents in the phone survey also reports being in good health. A higher percentage of on-line respondents agrees that they need a car to do many of the things they like to do, and they agree less that they try to limit driving, that they like walking, and and that they like taking transit. These differences suggest that the phone sample might be more inclined to bicycle. However, equal shares of the two samples agree that they like riding a bike.

TABLE 4 Comparison of Attitudes toward Physical Exercise and Travel Modes in the Two Surveys

	Phone survey Agree or Strongly Agree (%)	On-line survey Agree or Strongly Agree (%)	p-value
Bicycling is a normal mode of transportation for adults in this community (**Biking Normal**)[1]	81.6	52.1	.000
It is important for me to get regular physical exercise	99.0	94.3	.000
I enjoy physical exercise	87.7	76.4	.000
I am in good health	88.7	80.0	.001
I need a car to do many of the things I like to do (**Need Car**)	78.2	83.3	.046
I try to limit my driving as much as possible (**Limit Driving**)	80.0	56.8	.000
I like riding a bike (**Like Biking**)	74.8	76.9	.278
I like walking (**Like Walking**)	85.9	81.4	.060
I like taking transit	40.1	23.6	.000

*The scale of all the variables is from "Strongly disagree" to "Strongly agree". We re-categorized it into dichotomous variable (agree or strongly agree vs. not). Then the Chi-square test was applied to the new binary variables.
[1] Bold in bracket is the corresponding variable name used in the following models.

Inferential Comparison

Although the descriptive comparison shows that the bicycling levels in the phone and on-line samples are similar, the relationships inherent in each could differ. Thus we estimated models for the two samples to explore whether relationships between bicycling and potential explanatory variables differ.

Comparison of Models for "Bike or Not"

The first model uses regular bicycling as the dependent variable, derived from a survey question that asked, "During the last week, on how many days did you ride a bicycle?" We categorized the respondents who reported bicycling at least one day during the last week as regular bicyclists and the rest as not. A binary logit model was used to explore factors that explain regular bicycling behavior. The market segmentation method was used to examine whether the model estimates from the two survey samples are significantly different. We look first at whether the coefficients collectively differ across the surveys, then at whether individual coefficients differ.

Using the type of survey as the basis for segmentation, three best (most parsimonious) models were first estimated: two segmented models, the best on-line model (with on-line data only) and the best phone model (with phone data only); and the best pooled model (with pooled on-line and phone survey data). The best models were estimated by entering socio-demographic, travel constraints, attitudinal factors, and social environment factors as sets in steps into the binary logistic regression. At each step, only the statistically significant ($p<0.5$) variables were retained and insignificant variables were dropped by using a backward stepwise process. In backward selection, SPSS enters all the predictor variables in the model. The weakest predictor is then removed and the regression re-calculated. If this significantly weakens the model then the predictor variable is re-entered, otherwise it is deleted. This procedure is repeated until only significant variables remain in the model. Removal testing is based on the probability of the likelihood-ratio statistic based on the maximum partial likelihood estimates.

In order to properly conduct the chi-square test for the difference in the coefficients collectively, we re-estimated the pooled and segmented models with the superset of explanatory variables contained in all three models. Assuming the original segmented models differed with respect to the significance of at least some variables, then naturally some variables in the superset will be insignificant when they are all entered into one model. But these insignificant variables must be retained in order to ensure that the same set of variables is included in each model so that the chi-square test is legitimate. Next, a chi-square test was used to determine whether the coefficients collectively are significantly different between the segments. The chi-square value of 15.532 (df = 9) with a p-value of 0.077 (Table 5) indicates that collectively the coefficients are not significantly different between the segments at the 95% significance level (though they are at the 90% significance level).

TABLE 5 Chi-square Test for Difference of Collective Coefficients across Surveys

	Bike or Not Model	Biking Days Model
$LL(\hat{\beta}_{pooled})$	-314.997	-1123.250
$LL(\hat{\beta}_{online})$	-132.507	-518.892
$LL(\hat{\beta}_{phone})$	-174.724	-595.673
Chi-square	15.532	17.370
df	9	10
p-value	0.077	0.067

Although at the 5% level of significance we do not reject the null hypothesis that coefficients collectively are not different, it is possible that specific coefficients differ significantly. To test this possibility, the each of the best segmented models was re-estimated using data from the other segment, e.g. the best on-line model was re-estimated with the phone survey data, and vice versa. The two sets of coefficient were compared using t-tests to find segment-specific factors and generic factors. Significant coefficients in the best on-line model that were not different in the phone version of this model point to generic variables, while those that were different in the phone version are specific to the on-line model. Phone-specific factors were found following the same approach.

The results show that Age, Bike Limit, Limit Driving, and Biking Normal are generic variables that have equivalent effects in both survey samples. However, Female and Like Biking are survey-specific variables (Table 6). Female is significant only for the on-line sample. Liking Biking is significant in both the best on-line and best phone models, but it is associated more strongly with bicycling in the on-line survey than in the phone survey.

Comparison of Biking Days Model Estimates

The second set of models we developed to compare the inferential differences between the two types of surveys is for the dependent variable, Bicycling Days, derived from the same survey question as for the Bike or Not variable but keeping the original responses to this question. Values range from 0 to 7 days during the last week.

TABLE 6 Comparison of Models for Bicycled or Not

Variable	Best on-line model coefficient	sig	s.e.	Best on-line model with phone data coefficient	sig	s.e.	Comparison t statistic[1]	sig	p-value
Age	-.0260	**	0.011	-0.039	***	0.009	0.961		0.337
Biking Limit	-2.629	**	1.211	-0.848	*	0.512	1.354		0.176
Limit Driving	0.317	**	0.159	-0.031		0.148	1.597		0.111
Like Biking	1.799	***	0.254	1.152	***	0.174	2.101	**	0.036
Constant	-6.858	***	1.234	-2.533	***	0.868	2.866	***	0.004
McFadden R^2			0.218			0.149			

[1]df= N1+ N2-2*K=661

Variable	Best phone model coefficient	sig	s.e.	Best phone model with on-line data coefficient	sig	s.e.	Comparison t statistic[1]	sig	p-value
Female	-0.608	**	0.255	0.115		0.302	1.827	*	0.068
Age	-0.039	***	0.008	-0.031	***	0.010	0.603		0.547
Like Biking	1.205	***	0.168	1.954	***	0.252	2.474	**	0.014
Biking Normal	0.242	*	0.141	0.071		0.139	0.860		0.390
Constant	-3.646	***	1.013	-6.531	***	1.263	1.782	*	0.075
McFadden R^2			0.154			0.201			

[1]df= N1+ N2-2*K=667

Notes: *10% significance level, ** 5% significance level, *** 1% significance level
N_1 is the sample size of the first segment model, N_2 is the sample size of the second segment model, K is the number of explanatory variables.Variables in bold are potential generic variables in the two models. Variables in italic are potential survey specific variables in the two models.

Because the dependent variable (Bicycling Days) is a count variable that does not follow the normal distribution, Poisson regression is a better choice than linear regression. Another model form, negative binomial regression, performs better than Poisson regression in cases of overdispersion, i.e. when the is variance much larger than the mean. The likelihood ratio test based on Poisson and negative binomial distributions is commonly used to test for over-dispersed data.[1] This test showed that both the phone and on-line survey data are over-dispersed[2] and so we used negative binomial regression to estimate this set of models.

We followed the same process as for the Bike or Not model to test the significance of the difference in coefficients collectively and individually. The collective test produced a chi-square of 17.370 (df=10) with a p-value of 0.067 (Table 5) and again suggesting, that the coefficients of the segmented models collectively do not differ at the 5% significance level (though they do at the 10% level). The comparison of the individual coefficients shows that only a few variables differ significantly across the two surveys (Table 7). Generic variables were Female, Age, Emp Status, Like Walking, and Biking Normal. Like Biking was significant in both the best on-line and the best phone models, though the results suggest that the magnitudes of the coefficients may be different. Biking Limit and Limit Driving appear to be specific to the on-line survey. .

[1] http://www.uky.edu/ComputingCenter/SSTARS/www/documentation/P_NB_3.htm
[2] The phone survey data yielded a value of 57.856 (p = 0.000); for the on-line survey data, the value is 25.382 (p = 0.000).

TABLE 7 Comparison of Models for Bicycling Days

Variable	Best on-line model coefficient	sig.	s.e.	Best on-line model with phone data coefficient	sig.	s.e.	Comparison t statistic[1]	sig.	p-value
Emp Status	0.368	***	0.128	0.093		0.176	1.267		0.206
Biking LImit	-1.883	***	0.519	-0.730	**	0.295	1.932	*	0.054
Limit Driving	0.328	***	0.056	-0.088		0.087	4.008	***	0.000
Like Biking	0.823	***	0.075	0.871	***	0.101	0.381		0.704
Like Walking	-0.179	***	0.054	-0.200	**	0.083	0.212		0.832
Biking Normal	0.122	***	0.045	0.134	*	0.077	0.135		0.893
Constant	-3.972	***	0.418	-2.486	***	0.572	2.096	**	0.036
McFadden R^2			0.125			0.081			

[1]df= N1+ N2-2*K=313+364-2*7=663

Variable	Best phone model coefficient	Sig.	s.e.	Best phone model with on-line data coefficient	Sig.	s.e.	Comparison t statistic[1]	sig.	p-value
Female	-0.357	***	0.093	-0.151		0.139	1.237		0.217
Age	-0.022	***	0.003	-0.016	***	0.005	1.029		0.304
Like Biking	0.795	***	0.066	0.997	***	0.095	1.740	*	0.082
Biking Normal	0.114	**	0.050	0.127	**	0.065	0.160		0.873
Constant	-2.116	***	0.388	-3.203	***	0.549	1.618		0.106
McFadden R^2	0.046			0.111					

[1]df= N1+ N2-2*K=364+313-2*5=667

Notes: *10% significance level, ** 5% significance level, *** 1% significance level
N_1 is the sample size of the first segment model, N_2 is the sample size of the second segment model, K is the number of explanatory variables.
Variables in bold are potential generic variables in the two models.
Variables in italic are potential survey specific variables in the two models.

Performance of Predictive Model Developed from On-line Survey

In addition to comparing descriptive and inferential differences between the two survey methods, the predictive capabilities of the best models estimated with the on-line survey data were also tested. In this test, we applied the best on-line models for both Bike or Not and Biking Days to the phone survey data to predict bicycling behaviors for each of the respondents in the phone survey data base. In other words, we plugged socio-demographic characteristics and attitudes of each respondent in phone survey into the on-line models' equations to generate the predicted bicycling behaviors of each individual. Then we compared each individual's predicted bicycling behaviors to his/her corresponding observed behaviors. As measures of predictive capability, we calculated the mean square error (MSE), root mean square error (RMSE), and percent of observations correctly predicted for the model for Bike or Not.

Two measures indicating the on-line model's predictive performance are shown in Table 8. The MSE and RMSE show poor predictive performance for both models, according to the acceptable range of MSE described in Steyerberg et al [16]. The success table for the Bike or Not model shows yields a success index of less than one, indicating that the model does not perform even as well as the market share model and is thus considered not a good predictive model (Table 9). Note that the on-line models had relatively low predictive power even for the on-line sample, as indicated by the r-square and McFadden r-squared values for the models (see Tables 6 and 7). The low sample size and limited number of explanatory variables clearly limit the predictive capabilities of the models. Thus, the low predictive performance for the phone sample is not a conclusive test of inferential differences between the survey methods.

TABLE 8 Predictive Performance of Bike or Not Model and Biking Days Model

Model	Measures of Predictive Performance	
	MSE	RMSE
Bike or Not Model	0.519	0.720
Biking Days Model	6.131	2.476

TABLE 9 Success Table of Bike or Not Model

Alternative		Predicted			
		Did not bike	Biked	Row totals	Observed share
Observed	Did not bike	N_{11}=79	N_{12}=105	$N_{1.}$=184	$N_{1.}/N_{..}$=0.513
	Biked	N_{21}=151	N_{22}=24	$N_{2.}$=175	$N_{2.}/N_{..}$=0.487
	Column totals	$N_{.1}$=230	$N_{.2}$=129	$N_{..}$=359	1
	Predicted share	$N_{.1}/N_{..}$=0.641	$N_{.2}/N_{..}$=0.359	1	
	Success proportion [1]	$N_{11}/N_{.1}$=0.343	$N_{22}/N_{.2}$=0.186		
	Success index [2]	$(N_{11}/N_{.1})/(N_{1.}/N_{..})$=0.670	$(N_{22}/N_{.2})/(N_{2.}/N_{..})$=0.382		

[1]Proportion of those predicted to choose alternative i who actually chose i.
[2]Success proportion normalized by observed share scales success to market share predictions.

DISSCUSSION AND CONCLUSIONS

In this study, results from bicycling surveys with the same questions but using two different survey methods – a traditional RDD phone survey and an on-line survey – were compared as to differences in the characteristics of the samples they produced, the responses to survey questions, and the explanatory models derived from these responses. Differences were also tested by assessing the ability of the on-line-based models to predict the levels of bicycling reported in the phone survey.

Socio-demographic characteristics of the respondents in each sample differed significantly. Higher percentages of on-line respondents are younger, are UCD students, have a job, own a car, and have no physical limitation on bicycle riding than of phone survey respondents. This difference is consistent with studies that show that web users are usually younger and better educated than people having no access to the Internet (Fricker et al. [6]). These socio-demographic differences may explain some of the observed differences in responses to attitudinal questions, particularly those having to do with exercise and travel modes. It is also possible that the observed differences in attitudes stem from differences in the way that individuals respond to the same question if it is asked by a person on a phone or by a computer. For example, previous studies have found that when people are surveyed orally (e.g. by telephone), they may not remember all the items for a survey question for a long time and are more likely to choose the last categories in the list, a tendency that is called a recency effect [18]**.** Surprisingly, despite these differences, measures of bicycling were not significantly different across the two survey samples. This finding may reflect the widespread nature of bicycling in Davis, an exceptionally bicycle-friendly city.

A comparison of the inferential implications of the two survey methods also shows mixed results. In models for both regular bicycling (bicycled or not in the last seven days) and number of days bicycling (in the last seven days), the coefficients collectively do not differ significantly across the two samples at a 95% significance level – but they do at the 90% significance level. The comparison of individual coefficients shows that most of the individual coefficients are generic (the same for both samples). The policy implications we might draw from the models developed with each sample are largely similar: efforts to increase bicycling should focus on attitudes toward bicycling as well as differences in bicycling needs by age. However, the few variables that were specific to the different surveys are still notable. The fact that gender is significant in only the best phone models is especially notable, given the strong influence of gender found in most previous studies of bicycling behavior With only the on-line survey model in hand, local planners might overlook the need for bicycling programs targeted at women that the phone survey model would suggest. Note that a larger sample size might yield more survey-specific variables and thus more significantly different policy implications.

Overall, this study suggests that on-line and phone survey methods have the potential to produce significantly different results, both descriptively and inferentially. Our results are somewhat ambiguous, with the bulk of the analysis leaning toward largely insignificant differences but some evidence pointing towards potentially important ones. Whether the observed differences reflect systematic biases arises from the two survey methods is also not clear, given the lack of obvious ties between the methods and the results. From a policy standpoint, the observed differences in this case were not substantial, but in other studies they might be. This analysis points to the need for further research on the potential effect of survey method on the analysis and understanding of travel behavior.

ACKNOWLEDGEMENTS
This study was funded by a grant from the UC Davis Sustainable Transportation Center, supported by the U.S. Department of Transportation and the California Department of Transportation (Caltrans) through the University Transportation Centers program.

REFERENCES

[1] Dillman, D.A. 2000. *Mail and Internet Survey. The Tailored Design Method*, 2nd ed. New York: John Wiley and Sons.
[2] Couper, M. P. Web Surveys: A Review of Issues and Approaches. *Public Opinion Quarterly* Vol. 64, 2000, pp. 464-494.
[3] Taylor, H. Does Internet Research Work? *International Journal of Market Research*, Vol. 42, 2000, pp. 51-63.
[4] Best, S. J., B. Krueger,C. Hubbard, and A. Smith. An Assessment of the Generalizability of Internet Surveys. *Social Science Computer Review* 19, 2001, pp. 131-145.
[5] Al-Subaihi, A. A. Comparison of Web and Telephone Survey Response Rates in Saudi Arabia. *The Electronic Journal of Business Research Methods* Vol. 6, 2008, pp. 123 - 132, available on-line at www.ejbrm.com
[6] Fricker, S., M. Galesic, R. Tourangeau, and T. Yan. An experimental comparison of web and telephone surveys. *Public Opinion Quarterly* Vol. 69, 2005, pp. 370-392.
[7] Chang, L.C., and J. Krsnick. National surveys via RDD telephone interviewing vs. the internet: comparing sample representativeness and response quality. *Public Opinion Quaterly*, Vol. 73, 2009, pp. 641-678. Manuscript under review.
[8] Krosnick, J.A., S.S. Narayan, and W.R. Smith. Satisficing in surveys: Initial evidence. In M.T. Braveman & J.K. Siater(Eds.), Advances in survey research (pp. 29-44), 1996. San Francisco: Jossey-Bass.
[9] Schonlau, M., K. Zapert, L.P. Simon, K. Sanstad, S. Marcus, J. Adams, M. Spranca, H. Kan, R. Turner, and S. Berry. A comparison between responses from a propensity-weighted web survey and an identical RDD survey. *Social Science Computer Review* Vol. 21, 2003, pp.1-11.
[10] Vehovar, V., Z. Batagelj, and K. Lozar, K. Web Surveys: Can the weighting solve the problem? Paper presented at the American Association for Public Opinion Research Conference, St. Petersburg, FL. 1999.
[11] Greene, J., H. Speizer, and W. Wiitala. Telephone and Web: Mixed-Mode Challenge. *Health Serveries Research* Vol. 43, 2008, pp. 230-248.
[12] Dillman, D.A., and L.M. Christian. Survey Mode as a Source of Instability in Responses across Surveys. *Field Methods* vol. 17, 2005, pp. 30-51.
[13] Christian, L.M. How Mixed-Mode Surveys are Transforming Social Research: The Influence of Survey Mode on Measurement in Web and Telephone Surveys. Ph.D. Dissertation. 2007. Washington State University, Pullman.
[14] Dillman, D.A., G. Phelps, R. Tortora, K. Swift, J. Kohrell, J. Berck, and B.L. Messer. Response Rate and Measurement Differences in Mixed-Mode Surveys Using Mail, Telephone, Interactive Voice Response (IVR) and the Internet. *Social Science Research* Vol. 38, 2009, pp. 1-18.
[15] Engel, J. F., H.F. Fiorillo, and M.A. Cayley. Market Segmentation: Concepts and Applications. 1972. New York: Holt, Rinehart and Winston, Inc.
[16] Steyerberg E.W., F.E. Harrell, G.J.J.M. Borsboom, M.J.C. Eijkemans, Y. Vergouwe, and J.D. Habbema. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology* Vol. 54, 2001, pp. 774-781.
[17] McFadden, D. The theory and practice of disaggregate demand forecasting for various modes of urban transportation. Prepared for the Office of the Secretary, US DOT, for the

Seminar on Emerging Transportation Planning Methods, Daytona Beach, Florida, 1976, December, pp. 5-9.
[18] Krosnick J., and D.F. Alwin. An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, Vol. 51,1987, pp. 201–219.