CONDUCTING EFFICIENT TRANSIT SURVEYS OF HOUSEHOLDS SURROUNDING TRANSIT-ORIENTED DEVELOPMENTS

Marc D. Weiner¹
Orin T. Puniello²
Robert B. Noland³

¹Edward J. Bloustein School of Planning and Public Policy
Rutgers, The State University of New Jersey
33 Livingston Avenue, Room 367
New Brunswick, New Jersey 08901 USA

²Office of Predictive Analytics
Ketchum Global Research & Analytics
1285 Avenue of the Americas
New York, New York 10019 USA

³Alan 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 USA

Corresponding Author: Marc D. Weiner
mdw@ejb.rutgers.edu

Word Count: 6,043
Tables/Figures: 5
ABSTRACT

This paper presents best practices for transit survey protocols targeted at small geographic areas, such as in and around transit-oriented development and for households proximate to transit stations. The widespread dissemination of cell phones and rapidly decreasing presence of in-household landlines has made telephone interviewing prohibitively expensive; moreover, cell-phone number portability has confounded small-area probability sampling. A postal mail push-to-web protocol, described here, is a more cost-efficient approach for small area data collection, particularly where probability samples are sought. Address files are available from postal databases and the strategy is to send respondents details on how to access an on-line questionnaire; continued non-response is then followed up, with the final mailed contact including a paper questionnaire. Embedded in the contacting protocol is a survey research experiment whereby a method for maximizing response rates, the use of web instruction cards, is examined in the context of the mail push-to-web survey protocol.

Results lead to four optimization techniques: First, the mail-push-to-web protocol described in this article minimizes printing and mailing costs as well as computer data entry error from paper survey instruments; second, pre-notification letters also serve to reduce those costs by indicating non-deliverable addresses to be deleted from the sample frame; third, web cards—whether generic or tailored—are ineffective; and fourth, very strong evidence is presented that name association, i.e., having a commercial sample firm append names to the sample frame and using those names to personalize the mailing, has a very strong positive effect on response propensity.
1. INTRODUCTION AND LITERATURE

While many studies have examined the travel behavior of those living in or near Transit Oriented Developments (TODs) (1, 2, 3, 4), this article focuses on optimizing the transit survey methodology for such research. Using the outcomes of a mixed-mode survey conducted on an 8,000 household sample geographically centered around eight train stations in New Jersey, suggestions are offered for a cost-effective and efficient method of mixed-mode surveying of TODs, which, by definition, requires small-area sampling.

The rapid and extensive proliferation of cell phone ownership and use has transformed telephone surveying in general, and has particularly diminished the efficacy of random-digit-dial (RDD) surveys, which are designed to collect data representative of the larger population from which the sample was drawn. Results from the July-December 2014 National Health Interview Survey, the gold standard for assessing the scope of cell phone adoption and use, show that 45.4%—or more than two out of every five U.S. households do not have a landline telephone. Moreover, more than one-half of all adults aged 18-44 and of children under 18 were living in wireless-only households (5). It is clear, then, that landline-only RDD samples alone can no longer provide the coverage needed to produce representative samples.

An additional issue currently vexing protocols of small area surveying, such as within and in close proximity to a TOD, is that the additional expense of cell phone sample frame supplements is often prohibitive, even assuming a cell phone coverage and contacting can be small-area limited, an issue on which methods are still in the trial and error phase (6). This sea change in survey methods has compelled greater exploration of alternative sampling approaches.

The United States Postal Service’s Computerized Delivery Sequence File (CDSF) provides virtually complete coverage (97%) for every delivery point within the United States, with indicators drawing distinctions between seasonal homes, vacation homes, vacant housing, and other occupancy characteristics (7). As a result, address based sampling (ABS), a sampling approach based on contacting data found in the CDSF, is rapidly emerging as a leading sampling method. At the same time, cost-efficiency and the general prevalence of high-speed internet connections have made internet surveys a leading mode of data collection. The marriage of these two relatively new research capacities is a “push to web” approach whereby postal mailing is designed to stimulate participation in an on-line survey.

Sometimes this approach takes a mixed-mode protocol, under which the respondent can opt to complete a mailed paper questionnaire or, using a web-page address and password provided in the postal mailing text, can turn to the internet survey. Other times, the mailing is a straightforward push-to-web, without the paper questionnaire option. Either way, as with all survey research, the key challenge—while endeavoring to maximize response rates—is minimizing non-response bias. Against that background, whether the approach is mixed-mode or push-to-web only, the savings realized from minimizing printing costs, coupled with the minimization of data entry error that accompanies a web survey, leads to a strong researcher preference, even when the paper questionnaire option is offered in the mixed-mode context, for internet survey response over the paper alternative.

As a result, mixed-mode surveys using ABS sampling are becoming an increasingly important survey protocol. ABS sampling frames rely on the U.S. Postal Service’s CDSF, which provides nearly complete coverage of every postal address in the United States, and is continuously updated (7). ABS, however, comes with its own methodological challenges, particularly, how to best employ it in the context of surveys. This is a growing area of interest in
the survey research community. At the 2012 annual conference of the American Association of Public Opinion Research (AAPOR), attendees and presenters spent considerable time contemplating mixed-mode and ABS sampling issues, devoting 15 poster presentations, eight concurrent sessions, six additional papers, two demonstrations, and one short course to those subjects. By 2014, AAPOR had established an Address-Based Sampling Task Force, with subcommittees focused on nine areas of interest: introductions; definitions; frames; designs; quality/cost, weights/response rates, reporting, limitations, and auxiliary data (8). The findings presented in this paper suggest that the use of an ABS sampling approach to frame mixed-mode surveying can be effective when gathering data in and around TODs. What follows is an explication of best practices for that approach, based on a set of conditional logit models exploring the efficacy for key survey research considerations, along with a report of the outcomes of an embedded survey research experiment that explored the value of a web instruction card to help stimulate respondents’ selection of the web questionnaire mode over the return-mail paper questionnaire.

2. SURVEY METHODS

Between July 12th and August 22nd, 2012, the authors surveyed households surrounding eight train-station-proximate TODs in New Jersey. The survey instrument probed household attributes, attitudes toward community, use of public transportation, general travel behavior, and health. For each location, a randomly drawn sample of 1,000 households was stratified, practically speaking, by distance to the closest station entrance. More technically, this was accomplished by using GIS-driven mapping to determine the appropriate Census blocks to sample, and then, in turn, providing those Census block numbers to a commercial sample provider. Under this approach, within each strata, the sample frame represented a randomly drawn sample for which each household had an equal probability of selection. To that extent, the occupants of these households are represented in the sample proportional to their presence in the populations from which the samples were drawn.

To capture new development specifically related to transit, the first stratum was limited to newly built housing units within one-quarter mile of the train station. The overall initial sample frame, then, was stratified concentrically around the station into three geographic categories, with a total of 1,000 housing units per each of eight TODs:

(A) new housing units within a quarter-mile (200 units);
(B) all housing units within a half-mile (400 units); and
(C) all housing units between one-half and two miles (400 units).

Not all of the eight stations investigated had 200 units of new housing within a quarter-mile of the train station; in those cases the undersample count was assigned to the next nearest category, the half-mile strata. Per Table 1, “Sample Eligibility by Location,” and based on postal returns of the pre-notification letter, 6,938 households of the 8,000 household initial sample frame (86.7%) were determined to be eligible, i.e., to have valid contactable addresses, thus constituting the final sample frame. Given the nature of TODs, high turnover in housing for these areas was to be expected, and so the reduction in sample size was not unanticipated.

In terms of survey protocols, while pre-notification letters are primarily used to announce and prime the respondent to the seriousness of purpose of the survey, the use of those letters for
this secondary sampling purpose is common within the survey research community, as it permits adjustment to the sample frame for accuracy in calculating response rates, as well as, where necessary, conducting post-hoc power analyses.

### TABLE 1 Sample Eligibility by Station Location

<table>
<thead>
<tr>
<th>Station</th>
<th>Initial Sample Frame</th>
<th>Eligible Sample Addresses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cranford</td>
<td>1,000</td>
<td>923</td>
</tr>
<tr>
<td>Jersey City, Essex St.</td>
<td>1,000</td>
<td>846</td>
</tr>
<tr>
<td>Metuchen</td>
<td>1,000</td>
<td>925</td>
</tr>
<tr>
<td>Morristown</td>
<td>1,000</td>
<td>887</td>
</tr>
<tr>
<td>New Brunswick</td>
<td>1,000</td>
<td>850</td>
</tr>
<tr>
<td>Newark, Broad St.</td>
<td>1,000</td>
<td>772</td>
</tr>
<tr>
<td>Plainfield</td>
<td>1,000</td>
<td>851</td>
</tr>
<tr>
<td>Rahway</td>
<td>1,000</td>
<td>884</td>
</tr>
</tbody>
</table>

All told, sampled households received five contacts:

1. a pre-notification letter;
2. an invitation letter with a link to the internet survey page, a personalized pass code, and a one-dollar-bill incentive;
3. a three-day reminder postcard;
4. a two-week reminder letter in the same form as the invitation letter; and
5. a final reminder letter with a paper copy of the questionnaire.

Thus, the push-to-web survey protocol repeatedly requested respondents complete the survey via the web, only sending a paper questionnaire to non-responders with the last contact.

The sampling company that provided the CDSF data was able to provide names matched to the household for about 70% of the initial sample frame. Undeliverable returns of the pre-notification letter permitted the purging of the initial sample frame of non-working addresses. Of the final sample frame, 75.6% of addresses were matched to occupants’ names. In these cases, the mailing was personalized with the name of the householder; all other mailings were directed to “Community Resident.” The questionnaire was designed with a multimode unified design to reduce mode effects (9).

The self-administered web survey was programmed in English, and was also passively available in Spanish. Where the respondent’s web browser default language was Spanish, the survey would automatically display the Spanish language version. The paper questionnaire was printed in English and while the letter indicated that a Spanish-language version was available on request, no respondent requested it.

### A Contemporaneous Survey Research Experiment

Nonresponse is one of the most studied areas of survey error, and as response rates continue to decline it becomes even more crucial (10). Nonresponse takes two forms, unit and item; in light of the overall nation-wide reduction in response rates in all modes of survey, unit-nonresponse
was the key concern, i.e., the exploration focused on best practices for the maximization of
survey cooperation and instrument completion. Considerable time and energy has been spent
researching response rate maximization methods (11, 12, 13, and 14). A 2009 meta-analysis
examined the effects of 110 different methods to increase postal and electronic response rates,
but did not report any studies on the effect of web instruction cards (14). To that end, a survey
research experiment was embedded in the study to examine the possibility of reducing unit non-
response by inserting web instruction cards with “push-to-web” mailings.
Survey researchers have theorized that including a web instruction card with the mailed
contacts would increase response rates. Dillman, et al. (9) assert that the “web card,” as they
termed the insert, has two functions: simplifying the survey process, and, communicating the
importance and professionalism of the survey effort. Messer and Dillman (15) tested the use of
web cards by experimental design, hypothesizing that the web card will increase the proportion
of responses via the web option. However, their finding was contrary, i.e., the generic web card
did not increase response rates, whether overall or for the web component of the mixed-mode
survey. An earlier study by Lesser (16) also examined the use of web cards, and found that the
web card suppressed, rather than enhanced, response.

The efficacy of personalization and tailoring of survey protocols and contacts to improve
nonresponse is well documented (9, 15, 17, 18, 19, 20, and 21). These studies show that
“tailoring” and “personalizing” survey protocols and contacts typically lead to higher response
rates. In the case of web cards, all previous efforts used “generic” inserts, i.e., they were neither
tailored (meaning including some specific information thought to be useful to the respondent)
nor personalized (specifically referring to the respondent by name). There is, however, an
unpublished, non-peer-reviewed master’s thesis that implicitly tested whether tailoring with the
passcode would result in higher response rates; it did not (22, at pp. 28, 38, and Figure 3).

Building on this research, the survey protocol included the tailoring of an experimental
group’s web cards by providing the respondent’s unique login code. This embedded experiment
explicitly tested the hypothesis that the tailoring of the web card would increase the response
rate, as opposed to the negative and no effect of web cards on response rate found by Lesser (16),
Messer (22), and Messer and Dillman (15). The basis of the hypothesis follows the tailoring and
personalization literature, i.e., including helpful information on the web card—in this case, the
household-respondent’s unique passcode—would simplify the survey completion process by
eliminating the step by which respondents would need to refer back to the invitation letter for the
individualized information necessary to access the web survey. In more direct terms, the
motivation was to explore whether the tailoring of the web card would reduce respondent burden
and, in turn, increase the unit response rate.

For the web card experiment, within each of the eight station subsamples, respondents
were randomly distributed into three nearly equal sized groups:

(A) the control group, receiving no web card (N=2,306);
(B) the replication treatment group, receiving the generic web (N=2,318); and
(C) the experimental treatment group, receiving the personalized web card (N=2,314).

Web cards, printed in English on the front and Spanish on the back, were included in the 2nd and
4th contact to respondents in the “generic” and “tailored” groups. The web cards were modeled
after those used by Messer and Dillman in 2010 for comparability of replication and analyses
(15). See Figure 1, “Web Cards,” for a digital reproduction of the generic and personalized web cards, which were 3.375” wide by 7.75” tall.

FIGURE 1  Generic And Personalized Web Inserts
(Actual Size: 3.375” Wide by 7.75” Tall)

Distinctions between versions:

Text here refers to passcode location: generic “sent to you by postal mail,” / personalized: “printed below.”

Passcode appears here only on personalized version.

3. STATISTICAL METHODS AND RESULTS

Putting the experiment aside for the moment, in an effort to identify a set of best practices for mail push-to-web surveys at TOD locations, conditional logit models, controlling for each of the eight subsample’s proximate station location, were estimated. Two dependent variables were
tested: First, using the entire sample of 6,938 eligible households, a model was estimated with a binary dependent variable representing whether the respondent completed the survey (by either web or mail); second, using only those respondent households that completed the survey, a model was estimated with a binary dependent variable that indicated the survey mode, web or mail, by which the respondent answered the questionnaire.

The primary focus of this research is to highlight best practices for surveying of TODs and transit station areas. As such, the models’ independent variables were limited to only those considerations that, with some certainty, would be known to a survey researcher prior to receiving survey responses. This replicated the limitations on information known to the survey research when designing the survey protocol. As a result, such typical demographic indicators as the gender, age, ethnicity, and race of the within-household respondent, as well as number of persons or vehicles in the household, the household income category, and other standard controls were excluded from the set of independent variables.

The independent variables included were:

(A) whether the mailing was personalized due to the availability of the name of the unit occupant (“name association”);

(B) whether the unit was single family, as opposed to part of multifamily housing (with single family housing taken as a proxy for home ownership, which captures the greater sense of permanence and investment in the respondent’s housing unit when assessing their propensity to respond);

(C) whether the unit was new housing within a one-quarter mile radius of the train station, (which likely indicated whether the household was built specifically in connection with the TOD); and

(D) whether the household was within a half-mile (taken as a proxy for walkability to the proximate train station).

Upon the purchase and refinement of the CDSF-derived ABS sample frame, these predictors, unlike the demographic characteristics of the occupants of the household, would be known to the survey researcher prior to fielding the survey. These, then, are tested to determine how to best and most cost-efficiently design a survey research protocol for a TOD study to maximize the response rate and, in turn, minimize non-response bias.

The first conditional logit model (N=6,938) tests the effect these four independent variables have on the overall response rate, without reference to the survey mode of questionnaire completion. The dependent variable then, as noted, is binary under which 0 indicates “no response” (n=5,309) and 1 indicates “response by either web or mail” (n=1,629). These results, shown in Table 2, first confirm the vital importance of personalization of the mailings. That indicator, “Mail_Name,” with a z-score of 10.59, shows an odds ratio point estimate of 2.65 (p=0.000), which, per the 95% confidence interval, may be only as low as 2.21 and may reach as high as 3.17.

Thus, the odds that the overall response rate will be improved due to name-of-household-occupant personalization ranges from a low of 2.21 to a high of 3.17, compared to mailings to “Community Resident” or some similar generic address form. In terms of percentages, where the survey researcher has a name associated with a household, there is a 121% to 217% increase in the odds that the household will respond, compared to a household with no name association.
These values correspond to an increased range of probability of response due to name association from 0.688 to 0.760 greater, compared to a “Community Resident” mailing.

As discussed above, the independent variables indicating whether the household is a single family unit, or new housing within one-quarter mile of the train station, are taken as proxies, respectively, for home-ownership and new-TOD-related construction. While not as powerful as the effect of personalization due to name association, the confidence intervals for these odds ratios also show a response rate improvement due to each, holding all else constant. To the contrary, the indicator for a household within a half-mile from the train station did not return a statistically significant coefficient ($p=0.271; z=1.10$).

**TABLE 2  Conditional (Fixed-Effects Controlling for Station) Logistic Regression**

<table>
<thead>
<tr>
<th>DV, 1=Completed Survey; 0=Did Not Complete Survey (N=6,938)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR chi2(4)= 184.36; Prob &gt; chi2 = 0.0000</td>
</tr>
<tr>
<td>Log likelihood = -3582.7661</td>
</tr>
<tr>
<td>Pseudo R2 = 0.0251</td>
</tr>
<tr>
<td>Odds Ratio</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>Mail_Name</td>
</tr>
<tr>
<td>SingleFamily</td>
</tr>
<tr>
<td>NewHousing</td>
</tr>
<tr>
<td>HalfMile</td>
</tr>
</tbody>
</table>

The predictor variable indicating a single-family unit returns an odds ratio of 1.24 ($p=0.006; z=2.77$), which generates a confidence interval showing that under that condition, the respondent household is between 6.4% to 44.0% more likely to respond than a unit in a multi-family building; in terms of probabilities that range is from 0.516 to 0.590 greater likelihood of response. For new construction within a quarter-mile, the outcome is similar: A point estimate of 1.40 ($p=0.001; z=3.22$) generates a 95% confidence interval indicating that new housing in the closest stratum (here, one-quarter mile), is 13.9% to 71.2% more likely to respond than units of existing building stock outside the closest stratum; in terms of probabilities, there is a 0.533 to 0.631 greater likelihood of response.

The second conditional logit model ($N=1,629$) assumes the respondent has completed the questionnaire and so further tests the effect these four independent variables have on survey mode choice. The dependent variable, as noted, is binary under which 0 indicates “mail response” ($n=418$) and 1 indicates “web response” ($n=1,211$). These results, shown in Table 3, indicate that the strong effect of mailing personalization due to name association with the household unit, and the impact on response propensity of home-ownership, taken as a proxy from the predictor variable for a single-family unit, are already accounted for in the initial decision to respond. In turn, then, the service these variables provide speaks to enhancing general response propensity, rather than affecting the choice of survey response mode.

The strongest indicator in this model is whether the household unit is new housing within a quarter-mile of the train station, followed by whether the unit is within a half-mile of the station. The new housing/closest stratum’s odds ratio point estimate, 2.60 ($p=0.000; z=4.06$), yields a confidence interval showing a percentage increase in the likelihood of a web response over a mail/paper response ranging from 63.8% to 311.4% for new housing over existing...
building stock; this corresponds to increased probabilities of web over paper responses for that
category ranging from 0.621 to 0.804. The half-mile stratum’s odds ratio point estimate, 1.32
\(p=0.031; z=2.15\), yields a confidence interval indicating a percentage increase in the likelihood
of a web response over a mail/paper response ranging from 2.5% to 70.1%, which corresponds to
enhanced probabilities of a web response ranging from .507 to .630. It is likely that these
proximity indicators show greater web response compared to mail/paper response because in the
new housing instance there is greater probability that the housing was built and sold with internet
connection capacity. In both instances, they place the housing unit in question squarely within a
high-density area in close proximity to a train station, where internet connectivity is likely to be
readily available, and the respondent likely more motivated than one beyond walking distance to
the train station.

**TABLE 3** Conditional (Fixed-Effects Controlling for Station) Logistic Regression
DV, 1=Completed Survey Via Web; 0=Completed Survey Via Mail (N=1,629)

|                | Odds Ratio | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|----------------|------------|-----------|-------|-----|----------------------|
| Mail_Name      | 0.892393   | .1839817  | -0.55 | 0.581 | 0.595754    1.336736 |
| SingleFamily   | 1.177927   | .1799439  | 1.07  | 0.284 | 0.873145    1.589098 |
| NewHousing     | 2.596101   | .6098346  | 4.06  | 0.000 | 1.638215    4.114076 |
| HalfMile       | 1.320423   | .170447   | 2.15  | 0.031 | 1.025264    1.700555 |

Shifting gears to the web card experiment, the first point of interest is the differential
response rates for the control group, the replication group, and the experimental group. Response
rates were calculated using the relatively conservative AAPOR RR2 definition (23 at p. 52). Overall, the survey returned a gross 23.5% response rate, providing 1,629 survey interviews over
a sample frame of 6,938. The response rates for the control group (no web card, sample \(N=2,306\)
yielding 558 completes) and the experimental group (tailored web card, sample \(N=2,314\) yielding
563 completes) were statistically indistinguishable, at 24.2% and 24.3% respectively. The
replication group (the generic web card, sample \(N=2,318\) yielding 508 completes) response rate,
however, at 21.9%, was about 2.3 percentage points (about 10%) lower than the other two, a
finding in line with Lesser’s suppression effect (16).

To give these findings meaning within the transit survey context, a final set of conditional
logit models was estimated, again controlling for the location of each subsample’s proximate
train station; each set was distinguished by the use of one of three dependent variables (complete,
any survey mode; complete via web; and complete via mail/paper), and added to the \(a priori\)
independent variables tested above, were one of three web-card categories. This created a three-by-three set of nine conditional logit models, from which the relative probabilities of including
one of the three web card categories were calculated. Relative probabilities were calculated by
summing the logit coefficients for each independent variable, taking the exponentiation of that
sum to determine the overall odds ratio when all effects are combined, and then converting that
to value to a probability (in the form of the odds ratio divided by one plus the odds ratio). These
results are shown in Table 4.
As a preliminary note, each of these nine models strongly confirms the importance of obtaining names to match the household address so as to be able to personalize the mailing. More to the point of this exercise is the comparison of the relative probabilities of survey completion (or survey mode choice) in light of the extra effort and expense of including a web card. For the first set of three models, testing whether the survey was completed or not by either mode, the probability of completion under conditions of the tailored web card was—at .003 difference—essentially statistically equal to not inserting any card (0.157 to 0.160). The generic card improves the probability, but only by about 0.03 (0.189 to 0.160). Thus, there appears to be no meaningful effect on response propensity as a result of the inclusion of a web card, whether tailored or not.

**TABLE 4 Set Of Conditional (Fixed-Effects Controlling for Station) Logistic Regression**

| Complete, any Survey Mode | logit | P>|z| | Complete, any Survey Mode | logit | P>|z| | Complete, any Survey Mode | logit | P>|z| |
|---------------------------|-------|-------|---------------------------|-------|-------|---------------------------|-------|-------|
| No Card                   | 0.063 | 0.306 | GenCard                  | -0.148| 0.018 | PsCard                    | 0.083 | 0.175 |
| Mail Name                 | 0.974 | 0.000 | Mail Name                | 0.976 | 0.000 | Mail Name                | 0.974 | 0.000 |
| Single Family             | 0.213 | 0.006 | Single Family            | 0.214 | 0.005 | Single Family            | 0.214 | 0.005 |
| New Housing               | 0.335 | 0.104 | New Housing              | 0.338 | 0.001 | New Housing              | 0.335 | 0.001 |
| Half Mile                 | 0.071 | 0.271 | Half Mile                | 0.073 | 0.254 | Half Mile                | 0.072 | 0.261 |
| Sum of Logits             | 1.656 |       | Sum of Logits            | 1.454 |       | Sum of Logits            | 1.679 |       |
| Odds Ratio                | 5.239 |       | Odds Ratio               | 4.281 |       | Odds Ratio               | 5.358 |       |
| Probability               | 0.160 |       | Probability             | 0.189 |       | Probability              | 0.157 |       |

| Complete, Web | logit | P>|z| | Complete, Web | logit | P>|z| | Complete, Web | logit | P>|z| |
|---------------|-------|-------|---------------|-------|-------|---------------|-------|-------|
| No Card       | 0.058 | 0.394 | GenCard       | -0.133| 0.056 | PsCard        | 0.073 | 0.287 |
| Mail Name     | 0.874 | 0.000 | Mail Name     | 0.875 | 0.000 | Mail Name     | 0.873 | 0.000 |
| Single Family | 0.238 | 0.006 | Single Family | 0.239 | 0.006 | Single Family | 0.239 | 0.006 |
| New Housing   | 0.553 | 0.000 | New Housing   | 0.556 | 0.000 | New Housing   | 0.553 | 0.000 |
| Half Mile     | 0.169 | 0.020 | Half Mile     | 0.171 | 0.018 | Half Mile     | 0.170 | 0.019 |
| Sum of Logits | 1.893 |       | Sum of Logits | 1.709 |       | Sum of Logits | 1.909 |       |
| Odds Ratio    | 6.636 |       | Odds Ratio    | 5.523 |       | Odds Ratio    | 6.743 |       |
| Probability   | 0.131 |       | Probability  | 0.153 |       | Probability  | 0.129 |       |

| Complete, Mail/Paper | logit | P>|z| | Complete, Mail/Paper | logit | P>|z| | Complete, Mail/Paper | logit | P>|z| |
|----------------------|-------|-------|----------------------|-------|-------|----------------------|-------|-------|
| No Card              | 0.042 | 0.698 | GenCard              | -0.125| 0.254 | PsCard              | 0.081 | 0.449 |
| Mail Name            | 0.994 | 0.000 | Mail Name            | 0.996 | 0.000 | Mail Name            | 0.994 | 0.000 |
| Single Family        | 0.062 | 0.631 | Single Family        | 0.063 | 0.630 | Single Family        | 0.063 | 0.626 |
| New Housing          | -0.412| 0.052 | New Housing          | -0.410| 0.053 | New Housing          | -0.412| 0.052 |
| Half Mile            | -0.172| 0.116 | Half Mile            | -0.170| 0.121 | Half Mile            | -0.171| 0.119 |
| Sum of Logits        | 0.513 |       | Sum of Logits        | 0.352 |       | Sum of Logits        | 0.554 |       |
| Odds Ratio           | 1.670 |       | Odds Ratio           | 1.422 |       | Odds Ratio           | 1.740 |       |
| Probability          | 0.375 |       | Probability          | 0.412 |       | Probability          | 0.365 |       |

Generally speaking, the same relative results obtain for the set of models testing web response compared to no response, and testing mail/paper response to no response. For the
former, the tailored web card category and no web card category both returned statistically
indistinguishable probabilities of response (0.129 to 0.131), with the generic web card insert
increasing the probability of response by a non-zero but trivial amount (0.22) to 0.153. For the
latter, the tailored card to no card differential was 0.010 (0.365 to 0.375), with the generic card
insert increasing that probability by 0.037. It does not appear then that under any configuration
the web card insert helps improve response rates.

4. DISCUSSION AND CONCLUSIONS

The intention of this article is to bring forth transit survey protocol elements to help optimize
data collection in and around TODs, which are by definition targeted to geographically small
areas. Modeling results indicate four main points of optimization, all of which affect the cost-
benefit ratio of a TOD survey.

First, the mail-push-to-web protocol described in this article is strongly recommended.
To the degree that a survey respondent can be pushed to respond to a web questionnaire, printing
and mailing costs are minimized, as is computer data entry error from paper survey instruments.
By allowing a mailed paper response only with the final communication, the contacting protocol
described maximizes the push to web. Since those respondents with the propensity to respond by
web have, for the most part, already done so, this sequence also saves resources by minimizing
the number of paper questionnaires that need to printed, mailed, and data-entered.

Second, for this type of mailed survey, researchers should always use pre-notification
letters. In addition to their intended survey research function of announcing and underscoring the
importance of the survey, “undeliverable” returns of pre-notification letter serve as indicators by
which to purge the sample of non-functional addresses so that they receive no further contacts,
which in turns saves money on printing and mailing costs and increases the accuracy of response
rate and post-hoc power analyses calculations.

Third, Table 4 and the companion discussion present strong evidence that web cards—
whether generic or tailored—are, plainly speaking, a waste of money. That said, while all of the
web card research to date, including these findings, has demonstrated null or negative effects, it
is still possible that a differently designed card may be found effective. Indeed, as more people
move away from desktop computers to tablets and smartphones as the vehicle for completing
web surveys, it is not impossible to contemplate different designs and content for web cards that
may be helpful.

Fourth, Tables 2 and 3, together with the attendant discussion, present very strong
evidence that name association, i.e., having a commercial sample firm append names to the
CDSF-driven ABS sample frame, and using those names to personalize the mailing, has a very
strong positive effect on response propensity. While this is well established in the general survey
research literature, this article applies this principle directly to transit survey efforts to collect
data in and around TODs. To be sure, any financial resources that would have gone to web
inserts of any stripe would be better spent increasing the name association proportion of the ABS
sample frame. It may be, as in this case, the sampling firm can only return a 70% name
association rate; however, the strong evidence in favor of name personalization shown here
suggests that resources expended to locate and append names for the remaining 30% would have
been well spent.
Finally, while this article did not address the issue of incentives, there is a large literature indicating that a pre-paid incentive—here, a one-dollar bill with the first mailing—aids to increase the overall response rate.
FUNDING
The Voorhees Transportation Center and the Public Policy Graduate Studies Program at the Edward J. Bloustein School of Planning and Public Policy, Rutgers, The State University of New Jersey, conducted this research on behalf of the New Jersey Department of Transportation and NJ TRANSIT. Funding for the survey was provided as part of a larger project for the New Jersey Department of Transportation and also supported by the U.S. Department of Transportation’s University Transportation Centers Program under Grant Number DTRT12-G-UTC21.

ACKNOWLEDGEMENTS
The authors appreciate Stephanie DePetrillo and Andrew Kay of the Voorhees Transportation Center for support and advice throughout the project. Marc Weiner thanks M. Patrick Simon, and K. T. & G. Weiner-Simon for their ongoing support and encouragement.

DISCLAIMER
The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation’s University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

REFERENCES


