THE TRANSITION FROM DIARY-BASED TO SMARTPHONE-BASED TRAVEL SURVEY DATA: Implications for Travel Demand Modeling

Mark Bradley, RSG
Elizabeth Greene, RSG
Robert Wertman, MCCOG
Suzanne Childress, PSRC

RSG
the science of insight

A very recent innovation in data collection for travel demand modeling is the development of travel survey data collection completely via smartphones, as a replacement for more traditional diary-based travel survey methods. This shift in data collection methodology over time is likely to affect the data and methods used to model travel demand. The research brief describes two pilot surveys carried out in May 2015 using RSG's rMove smartphone application for travel demand surveys, one in Indiana (a 7 day survey with 275 respondents) and the other in the Seattle region (a 3-day survey with 520 respondents). A comparison of the smartphone-based survey data with analogous diary-based survey data from smartphone-owning adults in the same region shows that the smartphone-based approach captures about 25% more trips, with the greatest increase in walk and bike trips. There is also evidence that sampling non-response and self-selection biases are quite different for smartphone-based surveys compared to more traditional methods. In the period of transition toward full smartphone ownership, it is likely that many surveys will need to use a mix of smartphone-based and diary-based methods. It is important that the two methods can provide fully analogous data and can be used jointly in model estimation, so that any method-specific response biases can be estimated as part of the modeling process.

OBJECTIVES AND MOTIVATIONS

A key innovation in data collection for travel demand modeling is the development of smartphone-based travel data collection as a replacement for more traditional diary-based travel surveys. It is probable that smartphone-based household travel surveys will gradually replace telephone/internet-based diary surveys over the next five to ten years. This prospect raises the question: How is this shift in data collection methodology likely to affect the data and methods used to model travel demand?

This research brief is based on evidence from two pilot surveys carried out in May 2015 using RSG's rMove smartphone application for travel demand surveys, one in Indiana and the other in the Seattle region. After providing a brief description of those surveys and the methods used, we present a comparison of the data from those two surveys with analogous data from recent diary-based household travel survey data collected in those same two regions. Last is a brief discussion of the implications for travel demand modeling.

THE SURVEY METHODOLOGY

The Indiana Pilot Survey

The Indiana pilot smartphone survey project was funded by the Madison County Council of Governments (MCCOG) in Anderson, Indiana and the Federal Highway Administration (FHWA). Only a brief description of the survey is provided here, but a more complete description can be found in Greene, et al. (1).

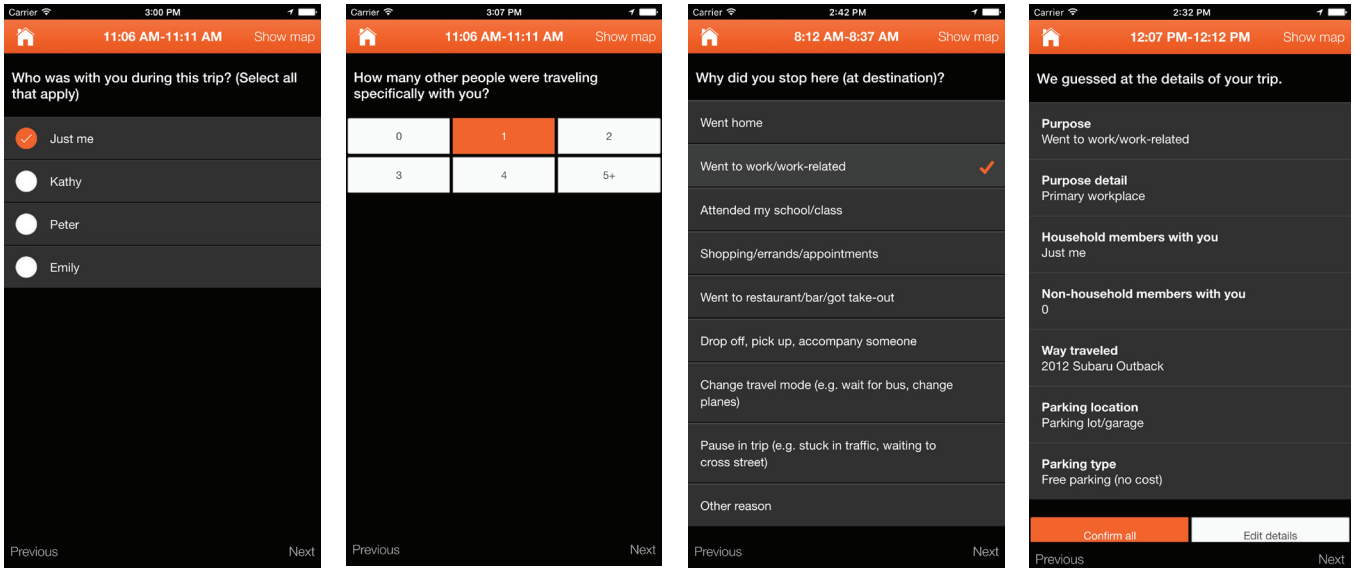
MCCOG conducted a diary-based household travel survey in Spring 2014, with retrieval of data from one-day travel diaries via internet or telephone. More than 1,400 households stated willingness to participate in further surveys. These households formed the smartphone-based survey recruitment pool in Spring of 2015. If at least one household member had an eligible smartphone, then the person (and therefore their household) was invited to participate. Household members with an eligible smartphone were asked to download the rMove app and participate for seven days. In total, 275 participants in 186 households downloaded rMove.

Comparing the demographic distribution of the 2014 survey respondents to the 2015 smartphone-based survey respondents, the largest difference found is related to age. People ages 65 and older have lower representation and people age 25-44 have higher representation in the sample of 2015 participants who downloaded rMove compared to the 2014 sample ($p < .0001$ in both cases). This result is of interest, because traditional approaches to household travel surveys tend to elicit over-representation among older ages (and smaller household sizes) and lower-than-desired representation among younger age groups

Each participant was asked to carry their smartphone everywhere they went for seven days, with rMove running continuously in the background. Whenever the app detects that a person has made a trip, a trip survey appears asking them to provide all of the same details for that trip that would typically be asked in a travel diary survey. This includes trip purpose, travel group composition, and mode of travel, as well as some further details that are mode-dependent (transit route, which car used, parking type and cost, etc.). Trip end locations and departure/arrival times do not need to be asked, as they are detected automatically and shown to respondents on a map (see Figure 1). All unanswered trip surveys are "queued up" on a daily summary screen.

The final screen in Figure 1 shows a feature to reduce respondent burden over multi-day surveys. If the app detects that a person has made a very similar trip

FIGURE 1: rMove Example Questions



previously (between the same two points at a similar travel speed), then the responses are pre-populated based on the answers given for the previous trip(s). In the example, the person had previously reported making a walking trip alone to go grocery shopping, and that information is "guessed" when the person repeats the trip, but the respondent has the option of changing any of the details. (In the data collected, respondents changed one or more of the pre-provided details about 30% of the time.)

At midnight each night, a short daily survey also appeared to participants, asking if any trips had been missed, or, if no trips had been made at all, the reason for not traveling. The goal was to focus on testing the paradigm in which the entire travel survey experience occurs on the smartphone. This is in support of the theory that having respondents answer questions in real-time as trips occur and for a longer period of time will lead to additional data and increased data quality.

TRIP SURVEY COMPLETION

Of the 275 participants who downloaded the app, 240 fully completed (by answering every single question) every survey in rMove for all seven assigned travel days. This statistic of nearly 90% of people answering every single trip survey for seven consecutive days is encouraging for the viability of longer data collection periods, completely via the smartphone. By comparison and recognizing a somewhat different selection process, the 2014 study sample had an overall 81% conversion rate, whereby just over 4 out of 5 households that were recruited went on to fully complete the household travel diary.

Respondents generally answered the trip-specific questions either within a few hours of the trip or after the participant's travel appeared to be done for the day. Although the survey completion peaked in the evening, 40% of surveys were completed within one hour of the trip ending and 71% of surveys were completed within five hours of the trip ending. The median time between making the trips and answering the related questions was 1.7 hours. This rapid response should help to alleviate the recall bias that can impact diary-based surveys when there is a time lag between the travel and the data retrieval.

In some cases, respondents reported that what rMove thought was an activity stop (a trip end) was actually part of travel, such as a long stop at a traffic light. On average, there was about one such spurious trip end per respondent per day. (In general, rMove errs on the side of including some false trips ends rather than missing valid trip ends.) In data cleaning, those stops were eliminated and the two adjacent trips were merged into one. Additionally, for 18% of respondent-days, respondents indicated that the app had missed one or more trips that they actually made—typically because they did not have their smartphone with them or it was turned off. In the pilot survey, there was no way for respondents to provide full details for their missing trips on the phone, although that feature is being added for future surveys. Therefore, the trip rates from the smartphone survey reported in the following section are likely somewhat lower than what will be obtained after this reporting feature is added.

The Seattle Pilot Survey

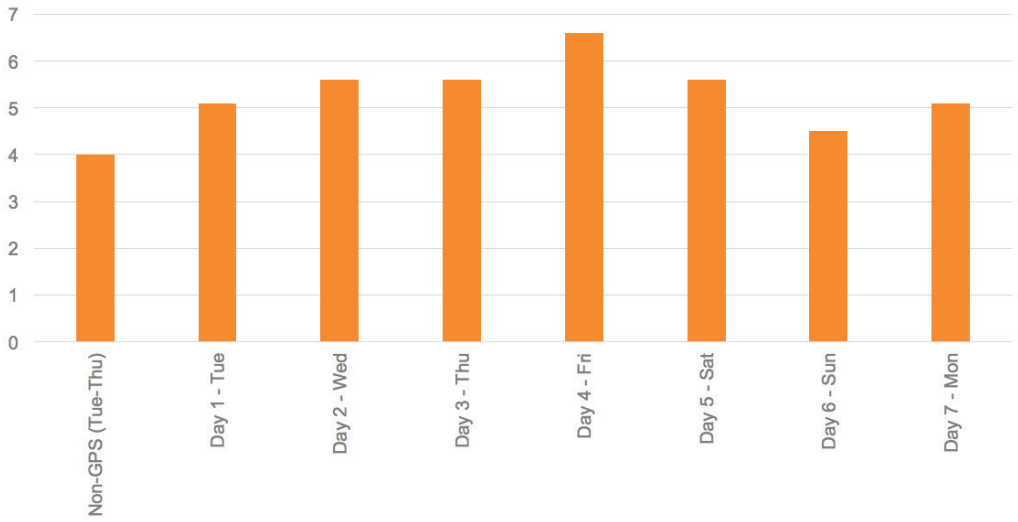
The Seattle region rMove pilot survey was funded by the Puget Sound Regional Council (PSRC) as part of a multi-year program that included diary-based household travel surveys in both the spring of 2014 (almost 6,000 households) and the spring of 2015 (820 new cross-sectional households and 1600 "repeat" panel households). The Seattle smartphone pilot survey was much like the Indiana pilot in that the respondents were recruited from the 2014 travel survey sample, and the same rMove app was used for a pilot survey administered in May of 2015. The main differences between the two pilots are: (a) the Seattle pilot was for only 3 days (Tue, Wed, Thu) rather than 7 days, and (b) the Seattle pilot had a somewhat larger sample size, with 566 persons participating and 460 answering every question for all 3 days.

Overall, the response rates and characteristics reported above for the Indiana pilot were very similar for the Seattle pilot.

SYNOPSIS OF MAJOR RESULTS

To compare the 2014 diary-based data to the 2015 smartphone-based survey data for Indiana, the analysis of the diary-based sample was restricted to those adults who own smartphones, so that we would be comparing similar segments of the population. Figure 2 shows that the one-day diary-based survey, (conducted only for Tue-Thu travel days during Spring 2014) yielded an average of 4 trips per person-day, while the rMove data has an average of more than 5 trips for every weekday, with a high of about 6.5 trips/day on Friday. Note that in traditional (non-smartphone) two-day diary surveys, we tend to see

FIGURE 2:
Indiana Pilot Survey- Comparison of Average Trips/Person-Day from the Diary-Based 2014 Travel Survey and Each Day of the 7-Day 2015 Smartphone-Based Survey



lower reported trip rates on the second day compared to the first, presumably due to respondent burden and fatigue. For the rMove pilot, every person started their 7-day period on Tue, so it is not possible to separate possible survey duration effects from day-of-week effects. However, there is certainly no evidence that rMove trip rates decline during the week, with Tue, Wed and Thu all showing similar trip rates.

A similar comparison on the Seattle pilot data (not shown graphically) gives average trip rate from the one-day diary-based survey (carried out on Tue, Wed or Thu) at about 4.4 trips/person-day, and the rMove pilot survey giving average trip rates of 5.2 on Tue, 5.7 on Wed and 5.4 on Thu. Those results are similar to the Indiana results, indicating that the smartphone-based method yields about 25% more trips per person-day compared to the online diary-based trip rates (which are already fairly high compared to surveys from past years using less sophisticated travel diary data retrieval methods such as paper or telephone).

One reason for the higher trip rate is that there are fewer respondents who fail to report any trips at all in any given survey day (5% of person-days captured by smartphone have 0 trips, compared to 10% of person-days in the diary-based data). There are also more respondents who report many trips (10% of person-days captured via smartphone have 10 or more trips, compared to 5% of person-days in the diary-based data.). This indicates that the smartphone-based method helps to overcome non-reporting biases at both ends of the trip frequency spectrum.

Another possible reason for higher trip rates with smartphone-based methods is a change in how self-selection bias may affect the data. With diary-based surveys, it has been suspected that even after accounting for demographic differences, those people who are most busy and travel the most may be somewhat less likely to complete diary-based surveys due to respondent burden. There is at least some evidence that this type of self-selection bias is less pronounced for smartphone-based data collection, because of less perceived burden and/or the technological aspect being more appealing to certain types of people. For example, when analyzing the PSRC diary-based survey data across different age groups (including only smartphone-owning adults), we find that the 18-24 and 25-34 age groups report markedly fewer trips per day, on average than other age groups—even the 65-74 age group. When analyzing the smartphone-based data, however, this age difference disappears, with no clear trend across the age groups from 18 to 74. This finding suggests that (a) the younger age groups are "laziest" when filling in diary-based surveys, but are more conscientious when using their phones, and/or (b) the subset of younger people who are willing to complete smartphone-based surveys tend to travel more, on average, than the subset of younger people willing to complete diary-based surveys. It is likely that both of these statements are true to some extent, and it will be an interesting research question to analyze in future smartphone-based surveys.

We also compared the distribution of trips from the PSRC smartphone-based and diary-based data across a number of different dimensions (mode split, purpose split, travel party size distribution, trip distance distribution), and found surprisingly

few differences, suggesting that smartphone-based methods tend to capture higher rates of all types of trips. There is a smaller percentage of trips returning home in the smartphone-based data, meaning more trips per home-based trip chain (tour). There are also more short-distance auto trips under 1 mile in the smartphone-based data—trips that may often be omitted in diary-based reporting.

IMPLICATIONS FOR TRAVEL DEMAND MODELING

The main implications for travel demand modeling include a number of issues that were touched upon in the previous sections, plus some that were not:

- Compared to more-traditional methods, smartphone-based surveys can provide a more complete "inventory" of household trip-making, with particular benefits for shorter trips such as walk and bike trips, including "loop trips" for exercise or recreation.
- It appears viable to complete smartphone-based surveys for up to 7 days per respondent with no apparent drop-off in survey participation or completion rates. Furthermore, all travel days have full trip details and can be used in modeling. This not only provides more useful data per respondent, but can enable new types of models, such as the allocation and substitution of activities across days of the week.
- Initial evidence suggests that smartphone-based surveys are less prone to some of the types of non-response bias and self-selection bias that have been prevalent in past diary-based travel surveys—particularly the biases towards older households and less "active" respondents.
- Smartphone-based location and time-of-day data are inherently more accurate than the data reported by respondents. This is even more the case for smartphones, which people tend to carry with them everywhere, than it is for the GPS devices used in previous travel surveys, which people are more likely to forget or leave in their vehicles.
- Smartphone-based surveys also provide trace data, from which respondents' travel routes and speed profiles can be derived. (Here, there is a trade-off between accuracy of the data and the amount of drain on the phone battery, which should become less of an issue in the future.)

Currently, about 70% of U.S. adults own smartphones, with the percentage increasing rapidly. However, many adults do not own smartphones, and our experience has shown that some of them are not willing or able to complete surveys by smartphone even if one is provided to them for free. This, along with the fact that we still need to survey young children who do not have (or are not allowed to have) smartphones, means that for the foreseeable future we will be using mixed methods, with some respondents providing data via smartphone and others using more traditional diary-based methods. Using mixed methods need not be an impediment to modeling, however, as long as the survey is designed so that the different methods provide the same data items, meaning that the data can be merged and used jointly in analysis. When that is the case, one can estimate "bias parameters" on the non-smartphone data cases in order to identify and adjust for any method-specific differences. In a sense, this is the reverse of the way GPS data has been used in the past for trip-rate correction, and is much more powerful in this case because both types of data can be used jointly in modeling. Thus, bias parameters can be estimated not only for trip or tour generation rates, but also for other variables such as mode choice constants and time of day choice constants.

REFERENCES

1. Greene, E., L. Flake, K. Hathaway, M. Geilich. "A Seven-Day Smartphone-Based GPS Household Travel Survey in Indiana." Paper submitted for the 2016 TRB Annual Meeting, Washington, D.C.

This paper is a reprinted version of an accepted paper for TRB's ITM Conference 2016.

FOR MORE INFORMATION

Mark Bradley

RSG
55 Railroad Row
White River Junction, VT 05001
Tel 415-328-4766
E-mail mark.bradley@rsginc.com

Elizabeth Greene

RSG
41 North Rio Grande Street, Suite 106
Salt Lake City, UT 84101
Tel 801-456-4907
E-mail elizabeth.greene@rsginc.com

Robert Wertman

Madison County Council of Governments
16 E 9th Street, Suite 100
Anderson, IN 46016
Tel 765-641-9482
E-mail robert@heartlandmpo.org

Suzanne Childress

Puget Sound Regional Council
1011 Western Ave, Suite500,
Seattle, WA 98104
Tel 765-641-9482
E-mail schildress@psrc.org