Assessing the representativeness of a smartphone-based household travel survey in Dar es Salaam, Tanzania

As Published: https://doi.org/10.1007/s11116-017-9851-6

Publisher: Springer US

Persistent URL: https://hdl.handle.net/1721.1/131519

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Assessing the Representativeness of a Smartphone-based Household Travel Survey in Dar es Salaam, Tanzania

Abstract The household travel survey (HTS) finds itself in the midst of rapid technological change. Traditional methods are increasingly being sidelined by digital devices and computational power – for tracking movements, automatically detecting modes and activities, facilitating data collection, etc. Smartphones have recently emerged as the latest technological enhancement. FMS is a smartphone-based prompted-recall HTS platform, consisting of an app for sensor data collection, a backend for data processing and inference, and a user interface for verification of inferences (e.g., modes, activities, times, etc.). FMS, has been deployed in several cities of the global north, including Singapore. This paper assesses the first use of FMS in a city of the global south, Dar es Salaam. FMS in Dar was implemented over a one-month period, among 581 adults chosen from 300 randomly selected households. Individuals were provided phones with data plans and the FMS app preloaded. Verification of the collected data occurred every three days, via a phone interview. The experiment reveals various social and technical challenges. Models of individual likelihood to participate suggest little bias. Several socioeconomic and demographic characteristics apparently do influence, however, the number of days fully verified per individual. Similar apparent biases emerge when predicting the likelihood of a given day being verified. Some risk of non-random, non-response is, thus, evident.

P. Christopher Zegras (corresponding author)
Associate Professor, Transportation & Urban Planning
Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge, MA 02139
Tel: 617 452 2433 | Fax: 617 258 8081 | czegras@mit.edu

Menghan Li
Department of Urban Studies and Planning
Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge, MA 02139
Email: mhli@mit.edu

Talip Kilic
Senior Economist
Development Data Group – Survey Unit
The World Bank
Email: tkilic@worldbank.org
Phone: +39-0677710236

Nancy Lozano-Gracia
Senior Economist
Social, Urban, Rural and Resilience
The World Bank
Tel: 2024732779
Email: nlozano@worldbank.org

Ajinkya Ghorpade
Department of Civil Environmental Engineering
Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge, MA 02139
Email: ajinkya@mit.edu

Marco Tiberti
Survey Specialist
Development Data Group
The World Bank
Via Labicana 110, 00184 Rome, Italy
Tel: +39-06-77710245
Email: mtiberti@worldbank.org

Ana I. Aguilera
Urban Economist
Social, Urban, Rural and Resilience
The World Bank
Email: aaguileradellano@worldbank.org

Fang Zhao
Principal Research Scientist
Singapore-MIT Alliance for Research and Technology, Future Urban Mobility Lab
1 CREATE Way, #09-02 CREATE Tower, Singapore 138602
Tel: 65-6601 1547, Fax: 65-6778 5654
Email: fang.zhao@smart.mit.edu

1 INTRODUCTION

The household travel survey (HTS), a basic building block for urban transportation modeling and analysis, dates, at least, to the home interview origin-destination surveys of 1940s in the USA (Weiner, 2013). Traditional means of collecting such data are often viewed as cumbersome and costly, typically leading them to be carried out irregularly, at best. In the industrialized world, data collection at a scale and resolution necessary to estimate and update traditional transportation planning models often happens infrequently. For every place like Singapore, which has carried out relatively wide-scale (e.g., ~0.5% of households) travel surveys every four years since 2004, we can find dozens of places like Boston, where 20 years passed between its most recent surveys. The non-systematic evidence available suggests that across much of the global south such data are more the exception than the rule (Ampt and Ortúzar, 2004; McDonald et al., 2003; Ortúzar, 2006; Schäfer, 2000).

The original HTS method was in-person interviewing to complete a "trip report" (Weiner, 2013), normally asking people to recall travel for the previous day (Stopher and Greaves, 2007). Techniques have evolved with technologies and experiences. One key variation in practice is whether the instrument is self-completed (by the interviewee) or completed by an interviewer. The medium of collection represents another point of variation. Today, self-completed travel surveys can be by paper or computer (e.g., computer-assisted web interviewing [CAWI]), while interviewer-assisted surveys can be by paper (paper-and-pencil interview [PAPI} or via telephone) or computer (e.g., computer-assisted telephone interview [CATI] or CAWI). The diary format for travel reporting emerged by the late 1970s (Stopher, 1992) as a way to make recall more intuitive for the subject. A concomitant change was the introduction of "prospective" recall, setting a future day for the diary completion, often with inclusion of a "memory jogger" for people to carry with them to assist in recalling trips on the assigned day (Stopher and Greaves, 2007). A related evolution has been in the survey's "object of recall" (Behrens and Masaoe, 2009); early diaries were "trip based", focusing on the trips undertaken; activity-based approaches, focusing on the sequence of activities undertaken, emerged in the 1980s; and more recently place-based approaches, focusing on the sequence of places visited, have appeared.

Beginning in the mid-1990s, advances in geolocation and telecommunications technologies found their way into HTS practice. The use of Global Positioning System (GPS) devices, first piloted in the USA in 1996 (Stopher and Greaves, 2007), has since become widespread (e.g., Bricka et al, 2012; Shen and Stopher, 2014), also ushering in a methodological advance: the "prompted recall" (PR) survey (Bachu et al., 2001). In the PR approach, an individual is presented with a map of her travel as derived from the GPS traces and asked to verify the data presented and provide additional information (e.g., mode, purpose), as relevant. GPS-based PR surveys have since benefited from enhanced post-processing of the data to better infer routes, stops, activities (purposes), and modes and a move to the web to enhance the user interface (e.g., for verifying information). GPS-based advantages have now been widely recognized. These include the increasing portability (and declining prices) of the devices, the high resolution of the data (and improved inference capabilities), and the relative ease of collecting data for many days (Shen and Stopher, 2014). Pure GPS logging does suffer from limitations, however, such as requiring the purchase and distribution of purpose-specific devices and requiring participants to

remember to carry them (Zhao, et al, 2015).

Mobile phones also hold promise, with two basic approaches of relevance: passive or purposeful. Passive approaches typically use a byproduct of commercial telecoms activity: call detail records (CDRs) which can approximate a phone's location based upon the cell phone tower through which phone activity (e.g., call or SMS) is transmitted. On the positive side, CDRs provide massive amounts of data, over time, on millions of phones, at relatively low cost. Such data have been used, for example, to estimate origin-destination matrices for metropolitan areas (Alexander et al., 2015) and activity-based mobility patterns (Jiang et al., 2017). The relative "low" CDR cost comes, however, with disadvantages. Geographic precision is generally conditioned by cell tower coverage, which might or might not coincide with other spatial units of relevance (e.g., census tracts). The representativeness of the data is also unknown, as telecom operators rarely release individual data on the phone (technically, the SIM card) user. The latter-related disconnect between analyst and subject also hampers inferences – of destination, activities, routes, modes, etc. – and verification vis-à-vis "ground truth."

The possibilities for "purposeful" use of mobile telephones for HTS emerged with smartphones and their apps. Smartphones have several theoretical advantages. They are increasingly ubiquitous and users carry them habitually. They have intuitive user interfaces, a range of sensors, and powerful computing capabilities. They clearly have potential for collecting high resolution travel data beyond that enabled by GPS alone. Despite this promise, inter-related technological, algorithmic, and sociocultural challenges exist. For example, the need to preserve battery life requires phased GPS use, limiting locational precision. Subsequent sparse and noisy data hinder the backend inferences (e.g., of modes) carried out upon the data. Finally, smartphone penetration rates vary by income and other factors (e.g., age, technological 'savvy'), as might tendencies towards, and capabilities for, using them. The latter factors – potentially affecting battery life, device durability, sensor activation – could impact the quality of the data collected.

This paper assesses the potential of a smartphone-based HTS in the Global South, specifically in an African city. Mobile phone usage in Africa has proliferated in recent years, growing at about 10% per year and reaching a penetration rate on the order of 48% by 2016 (GSMA Intelligent, 2016). Mobile telephony's potential value for better understanding travel behavior and planning transportation in the region has not gone ignored. Demissie et al (2016), for example, used CDRs to estimate national-level commute origins and destinations in Senegal. Pinelli et al (2016) also used CDRs from Abidjan, Ivory Coast, to identify possible design improvements to public transport routes in the city. With increasing smartphone penetration on the continent (smartphone adoption increased from 4% in 2010 to 23% in 2015; GSMA Intelligent, 2016), these devices have also been applied to understand urban Africa's urban mobility conditions. Ndibatya, et al (2017), for example, used a smartphone-based app to map the semiformal minibus services in Kampala (Uganda). Saddier et al (2016) performed a similar mapping exercise of minibuses in Accra (Ghana), including counts of passenger boarding and alighting.

We know of no prior efforts, however, to use smartphones for collecting HTS data in Africa. This paper examines a pilot in a sub-Saharan African city: Dar es Salaam, Tanzania. Given the promise and challenges of such technologies, we focus on evidence of possible biases in the representativeness of the data collected. How effective might such a data collection approach be in capturing the travel of different residents? Specifically, what factors might influence

individuals' likelihood to participate in the survey and the rate of participation (in terms of verified days)? The following section presents basic travel survey challenges and the smartphone-based data collection platform used. The third section presents the research context, the specific data collection approach applied, descriptive statistics, specific challenges, and the research questions. Section four presents the modeling methods and section five presents the modeling results. A final section concludes.

2 FUTURE MOBILITY SENSING (FMS)

2.1 Travel survey challenges

A fundamental concern with HTS is non-random, non-response, which can introduce problematic bias at three levels. The first is with the recruitment of participants (i.e., "cooperation rate"; McGuckin, et al., 2004). Simply, it may be difficult to recruit certain travel survey subjects, including for reasons of interest to the survey itself (e.g., busy people, who travel a lot may be harder to reach and/or less inclined to want to spend time on the survey). A second level is the actual participation by those recruited. For example, an individual may agree to participate in a survey, but purposefully report no travel on a given day to reduce response burden, what Madre et al (2007) call "soft refusal." A third level is selective non-response within reporting by participants. Examples include individuals purposely omitting trips or tours due to complex days (high levels of activity) or privacy concerns (e.g., not wanting to disclose children's locations and schedules) and/or misremembering activities (e.g., forgetting short trips or short duration activities) (e.g., Rieser-Schüssler and Axhausen, 2014).

Theory suggests that cooperation rates and response rates will decline with the anticipated survey burden. Burden is a function of: the length of the survey period; the length and difficulty of the survey instrument(s); amount and type of interaction (if any) with interviewers; the "reporting unit" (e.g., stage, trip); and, various characteristics of the household and individuals (Rieser-Schüssler and Axhausen, 2014; Stopher et al., 2007). Madre et al (2007) also note the important role played by the survey firm, as interviewers can have an incentive to, e.g., "invite the respondents to use soft refusal to avoid the task at hand" (p. 123). Relatedly, de Leeuw and de Heer (2002) highlight the need to strictly supervise interviewers.

The traditional retrospective survey method has long been assumed to result in under-reported trips (Stopher and Greaves, 2007). A problem to understanding the degree of resulting biases, however, is the lack of "ground truth" for comparison. The emergence of GPS-based surveys offered a new opportunity for comparing traditional survey accuracy. Stopher et al (2007), for example, used a GPS-based survey to validate a face-to-face interview survey in Sydney (Australia). They find a 7.4% under-reporting rate for trips, with evidence of under-reporting for people making many trips, for night trips, for shorter distance (and duration) trips and social trips. Their analysis also illuminates the heterogeneity in subjects' preferences for different survey implementation strategies, and how this can impact participation and response rates. For example, they find that, in the Sydney case, nearly the majority of households preferred phone-based validation (i.e., CATI), with internet-based (CAWI) having the second highest preference. The latter, however, had high drop out and non-compliance rates, while phone-based validation had the highest compliance. Similarly, Kagerbauer et al (2013), in comparing PAPI, CATI and CAWI

methods (in Germany) found CATI to have a higher participation rate (measured by useful questionnaire/cooperative households) than CAWI in a 7-day survey in Stuttgart. Kagerbauer et al (2013) conclude that multimode implementation methods will be necessary moving forward, due to differences in cohort preferences, a conclusion echoed by Wolf et al (2013).

2.2 FMS: A smartphone-based HTS platform

In a recent review of GPS-based HTS approaches, Shen and Stopher (2014) gave little attention to smartphones. While recognizing their increasing ubiquity and relatively quick geo-positioning capabilities, Shen and Stopher (2014) discounted smartphones due to their somewhat short battery life and, curiously, their "high cost of transferring data from phones to data centres" (p. 321). In contrast to these misgivings, the viability of smartphone-based travel surveys has now been demonstrated. Future Mobility Sensing (FMS) is one available smartphone-based approach, originally developed and piloted in Singapore in 2012-2013, piggybacking on the government's household interview travel survey (HITS) (Cottrill, et al., 2013). The FMS platform consists of three components: the smartphone app, backend server, and web-based user interface for verification (Figure 1).

FMS aims to identify a user's trip origins (starts), destinations (stops), non-travel activities, modes, and routes. The smartphone app, running in the background of the phone, collects sensor data (from GPS, WiFi, GSM, and accelerometer) without requiring user intervention. The application is designed to be lightweight (in terms of memory use), non-intrusive, and energy efficient. FMS uses various approaches to minimize battery consumption (Nawaranthne et al. 2014), a major *technical* concern for location-based applications (posing an energy use versus location precision tradeoff). A *sociocultural* challenge relating to locational precision comes from individuals' possible discomfort with being "followed." The latter concern has not been examined to date, but could increase bias in participation, affecting likelihood to participate and/or collect data (purposely not carry phone, turn it off, or turn off relevant sensors).

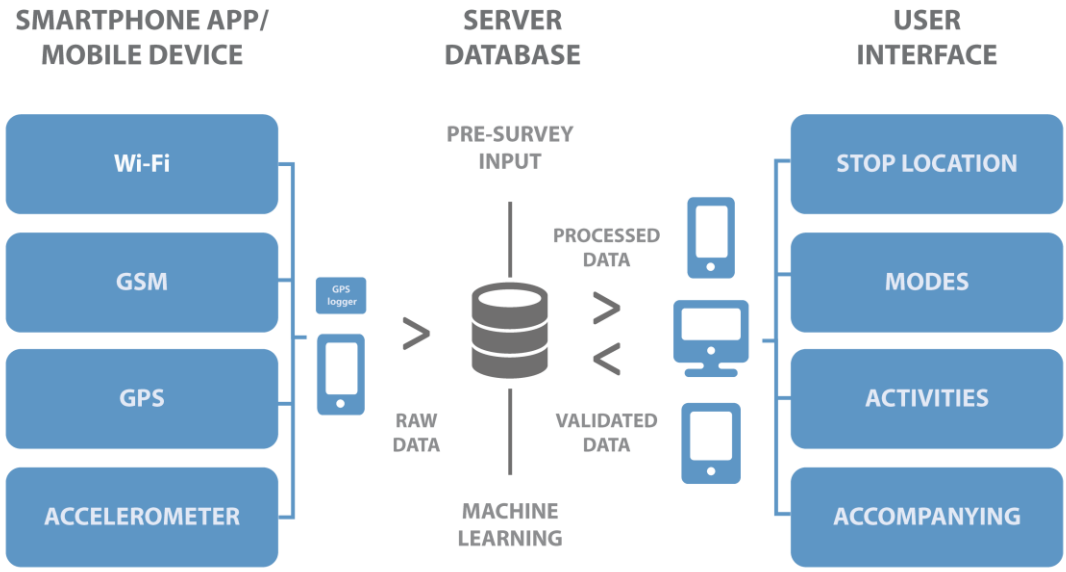

Figure 1 FMS Architecture.

FMS uses a prompted-recall protocol. The sensor data collected on the phone are transferred to the back-end server through either the cellular network or WiFi. The server includes the database and the data processing and learning algorithms for inferring stops, stop duration, stop location, travel modes, etc. The server-based inference is designed to minimize user interaction burden during verification (a main objective being to minimize false stops presented since users tend to be reluctant to change system-generated stops; Zhao, et al., 2015). Detected stops play a key role in identifying activities and travel modes; the latter because detected stops also affect mode inference. FMS defines a stop based on how much time a user spends in a place; for activity inference, detecting a change of modes is relevant, for example, while delays due to traffic are not. A main challenge to stop detection comes from the data gaps produced by the battery-saving phased sampling approach.

FMS' first round of stop detection uses location data, with GSM, Wi-Fi and accelerometer information used to merge stops that would otherwise be interpreted as being distinct. Travel mode inferences use GPS and accelerometer and public transit network information (if available). Inferences of non-travel activities (e.g. home, work, shopping) use the participant's previous verifications and points of interest (POI) data and other contextual information, if available. When a GPS reading is unavailable, location data come from cell tower information and WiFi (if on). These factors impact the availability and quality of the location data and the inferences made and presented for verification (and, thus, verification burden). The backend learns via verification to improve subsequent inferences.

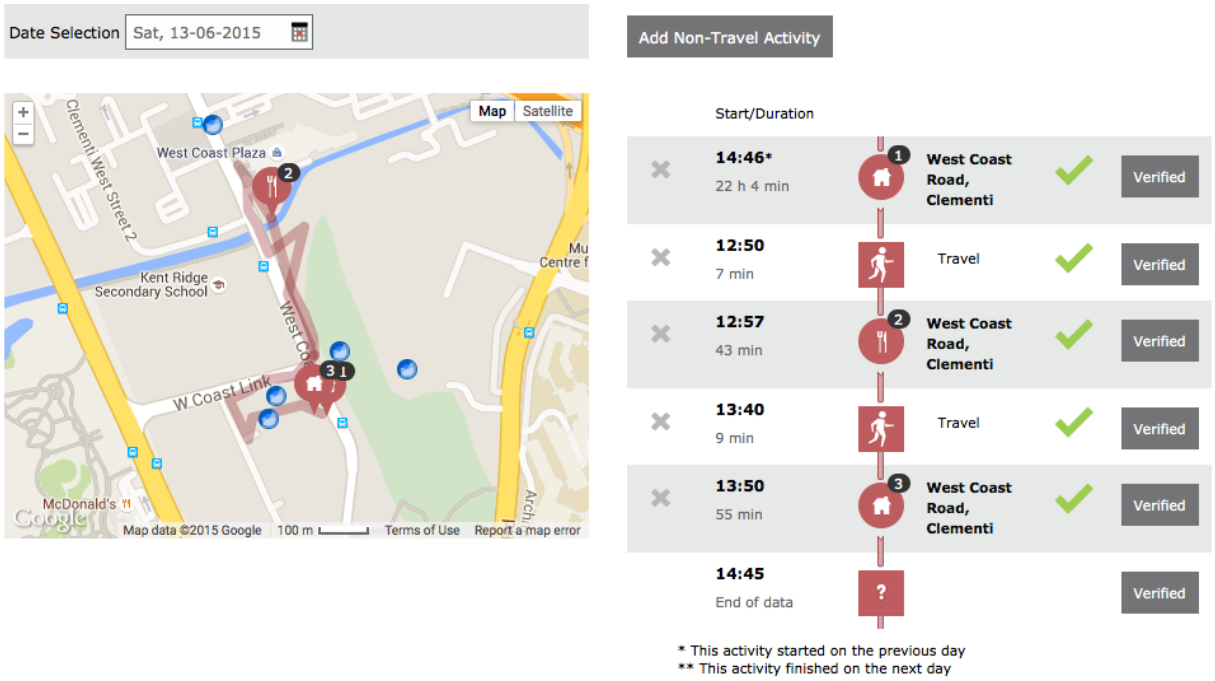

Figure 2 FMS activity diary timeline as presented on the FMS website for verification.

Typical FMS implementations can be categorized as CAWI (with help line services to assist respondents as needed), using a timeline-based recall approach. The website presents a user her

FMS inferences as an activity diary timeline (Figure 2). Users can make changes including: adding or deleting stops, editing stop locations and times, answering additional pertinent questions (e.g., amount paid for parking), changing travel mode, and adding or changing activity purposes. Inferences must be verified and only after verification of an entire day's stops and trips will the system mark the day as verified.

FMS applications to date reveal a number of important implications for minimizing user burden and maximizing data quality. Stop detection accuracy plays a critical role; erroneous stops increase verification burden. False positives are preferred over false negatives (Zhao, et al., 2015), since users find it easier to delete stops rather than add them. Sensor limitations combined with travel behaviors can make verification cumbersome and error-prone. Long times indoors can generate signal jumps (false short trips), since inactive GPS will prompt the app to find cell tower-based location readings. Capturing public transport alighting can be especially challenging, as can very short trip purposes. Mode detection can be tough in especially congested situations where signals among different modes are difficult to distinguish.

Existing FMS implementations have not allowed for experimental designs enabling direct comparison to other methods, as Stopher et al (2007), for example, do with a GPS-based method. The Singapore 2012 pilot did provide the possibility to compare reported user behavior collected via FMS versus HITS (Zhao, et al., 2015). HITS was implemented as a face-to-face, interview-assisted CAWI. As part of the HITS interview, selected participants were invited to subsequently take part in the FMS pilot. Nearly 800 participants verified at least five days via FMS. A comparison of the FMS-recorded behavior to the HITS-reported behavior (same individuals, different days) suggested FMS had more accurately reported travel times and time-of-day of travel, and less under-reporting of trips. The results also revealed the relative heterogeneity of most people's travel and their tendency to self-report "typical" days (Zhao, et al., 2015). As of early 2017, FMS has been deployed for official HTS in cities in the USA, Israel, and Singapore (HITS 2016-17).

3 FMS IN DAR ES SALAAM (DAR)

3.1 Context

Dar is Tanzania's largest and fastest growing city, with 4.4 million persons in 2012 (NBS, 2014). The nation's economic hub and principal port, the city has a GDP per capita (US$3,000 in 2012 at PPP) 1.7 times higher than the rest of the nation (ESRF, 2015). Still, approximately 60 percent of the city's residents work in the informal sector (Kiunsi, 2013b) and more than 70 percent live in informal settlements scattered across the city (Behrens, et al., 2016). The city is relatively monocentric and has been characterized as "low rise" and "sprawling" (Kiunsi, 2013a; p. 96) with an average population density of 24 persons per hectare. Density increases greatly near the central business district (CBD), with some areas exceeding 300 persons per hectare (Kiunsi 2013b). Consistent with the city's relative level of economic development, households in the city are smaller than the national average (3.9 versus 4.7; NBS, 2014).

Travelling in Dar is notoriously challenging. The city has five main radial roads and one circumferential road, all converging on the CBD. Only about a quarter of the city's roads are

paved (Kiunsi, 2013a). Industrial activity and the port, combined with poor road conditions, further exacerbate congestion. Melbye et al. (2015) estimate average speeds in the city at 12 km/hr. Similar to other cities in the region, public transport suffers from low quality and unsafe services (Nkurunziza et al. 2012). Dar's public transportation is dominated by privately operated, loosely regulated minibuses (daladalas) and three-wheeled auto-rickshaws (bajajs). The city government operates a few traditional buses and introduced a bus rapid transit line (BRT), DART, in April 2015.

The city has been rapidly motorizing, with an estimated 600,000 to 700,000 automobiles by 2012 (Kiunsi, 2013a; Melbye et al, 2015). This figure would give it a relatively high motorization rate of 100-125 cars per 1,000 persons. Other data suggest lower motorization rates; Masaoe et al (2011) report, for example, that 90% of households have zero automobiles. The estimated number of motorcycles ranges from 250,000 (Kiunsi, 2013a) to almost 500,000 (Mkalawa and Haixiao, 2014). According to Mkalawa and Haixiao (2014), public transit accounts for 68% of passenger trips, walking 17%, and private vehicles 12%.

Nationally, Tanzania's mobile phone ownership increased from just 10% in 2002 to 73% (of adults) in 2015 (Poushter, 2016). Smartphone penetration levels are still relatively low, but increasing rapidly (10% in 2014, 15% in 2015) (Poushter 2016). In Dar, the smartphone penetration rate is likely higher than the national average, given the city's relatively higher levels of income and lower rates of poverty. Nonetheless, various indicators of "smartphone divide" prevail, with ownership rates nearly twice as high for men, for younger adults (18-34 versus 35+)) and for higher income individuals (Poushter 2016).

3.2 Travel survey precedents and challenges

Behrens et al. (2006) review travel survey practices in seven sub-Saharan African countries (not including Tanzania), finding home-interview-based PAPI, with a trip-based diary to be common HTS practice. Based on their review and experiences, they identify a range of problems with HTS experiences in the region. These include: challenges in defining sampling unit (e.g., definition of household); high variation in literacy and numeracy; sociocultural heterogeneity which challenges linguistic, conceptual, and measurement equivalence and familiarity; and various survey administration challenges, including interviewer training, respondent distrust, interviewer bias, and difficult field logistics. Behrens and Masaoe (2009) catalogue a number of commonly resulting errors in HTS in the region: item non-response, recording errors, missing trips, missing trip stages, and incomplete diary period recall.

In Dar, passenger travel survey data have been collected in various ways over the years. The Japanese government supported a large household travel survey in 2007 as part of a transport master plan. Also in 2007, a stated preference survey was conducted among 740 commuters traveling to the CBD to estimate demand for the proposed BRT (Nkurunziza, et al., 2012). In 2010, university researchers implemented an HTS among Dar households (Masaoe et al., 2011), designing the survey after an experiment of different survey diary strategies (trip-based, place-based, or activity-based) (Behrens and Masaoe 2009). Those experimental findings suggested that place-based diaries (chronological reporting of the place of travel inquired, followed by an inquiry of means of travel) have the best recall and least error while trip-based diaries are most preferred by interviewers (due to survey duration). Behrens and Masaoe (2009) recommend

place-based diaries, with sequential question format and associated memory joggers, as most appropriate for an HTS in African cities.

3.3 FMS implementation

FMS was deployed in Dar over four weeks in November-December 2015, through a World Bank project supported through an internal "Data Innovation Challenge" grant. Three hundred households were selected at random from a larger sampling frame (a sample of 2,400 households previously interviewed by the Measuring Living Standards within Dar es Salaam Survey (MLDS) in January-February 2015). For each selected household, up to two smartphones were deployed among adult household members, aiming to strike a gender balance; 581 candidates were chosen. Each was administered a short multi-topic questionnaire (pre-survey) with individual-specific MLDS modules and given an Android smartphone with free data bundles and equipped with the FMS application. Table 1 compares socio-demographic characteristics of FMS candidates, FMS participants, citywide MLDS respondents (providing the best available estimate for citywide statistics), and the Tanzania national population. A few variables require explanation. More than half Dar's residents work in a family business; we use monthly sales to proxy business size. The household wealth index is the composite assessment of durable goods and small items ownership (see Appendix). The depression score is the test score of 20 questions regarding emotional depression status from the questionnaire. The following sub-section discusses the implications for the relative representativeness of the sample.

Compared to other FMS implementations (i.e., self-completed CAWI), the Dar case used an interviewer-assisted CAWI approach for verification. A number of reasons led to this approach. Generally, while participant self-correction of processed GPS data has been presumed as "ground truth," studies have shed doubts on the validity of this assumption without interviewer guidance (Rieser-Schüssler and Axhausen. 2014). More specific to the Dar context, one concern was that not all users would be familiar with interacting with a web-based interface, which would have implied major training efforts and/or risks of introducing too many errors via self-verification. Even more basically, some users may not have Internet access. Finally, using interviewers was seen as a way to reduce risk of participant selective non-response. Interviewers made verification calls to participants every three days. Interviewers verified inferred travel modes, activity purposes and locations, and costs and challenges associated with trips, as displayed (to the interviewer) via the web interface. Importantly, interviewers would delete wrongfully inferred "phantom" trips/stops, but were instructed not to add additional trips/stops not inferred by FMS.

Table 1 Socioeconomic and Demographics: FMS, MLDS, Nation

		FMS candidates		FMS participants		MLDS		Nation	P-value:
		Mean/ Proport.	*Std. Dev.*	*Mean/ Proport.*	*Std. Dev.*	*Mean/ Proport.*	*Std. Dev.*	*Mean/ Proport.*	(FMS vs. MLDS)
Categorical variables	Urban	100%	-	100%	-	100%	-	29%	-
	Rural	0%	-	0%	-	0%	-	71%	-
	Gender: female	51%	-	51%	-	52%	-	51%	0.70
	Gender: male	49%	-	49%	-	48%	-	49%	0.70
	Type of primary job: Government	4%	-	3%	-	4%	-	-	0.01
	Type of primary job: Private company	15%	-	16%	-	17%	-	-	0.18
	Type of primary job: Private individual(s)	14%	-	14%	-	16%	-	-	0.27
	Type of primary job: Public company	3%	-	2%	-	2%	-	-	0.91
	Type of primary job: self-employed	64%	-	65%	-	45%	-	-	0.03
	Type of primary job: others	0%	-	0%	-	1%	-	-	0.00
	Type of primary job: no job	24%	-	20%	-	16%	-	-	0.43
	Marriage: Married (monogamous)	64%	-	66%	-	50%	-	51%	0.00
	Marriage: Married (polygamous)	2%	-	2%	-		-		
	Marriage: Living w/ partner, unmarried	5%	-	5%	-	-	-	7%	-
	Marriage: Divorced	2%	-	1%	-	1%	-	3%	0.17
	Marriage: Separated	3%	-	2%	-	3%	-	1%	0.42
	Marriage: Widowed	3%	-	3%	-	4%	-	3%	0.26
	Marriage: never married	22%	-	21%	-	42%	-	36%	0.00
	Driving license - car	11%	-	11%	-	-	-	-	-
	Driving license - motorbike	2%	-	2%	-	-	-	-	-
	Driving license - none	86%	-	87%	-	-	-	-	-
Continuous variables	Age: <18 (proportion)	0%		0%		38%		50%	-
	Age: >=18 (mean and sd)	35.83	10.53	36.06	10.59	35.19	13.30	-	0.07
	Education years	11.02	3.45	11.11	3.47	10.01	4.29		0.00
	Distance from home to CBD	13.14	8.36	13.16	8.18	13.77	8.46	-	0.09
	Wealth Index per household member	-0.03	0.37	-0.02	0.36	0.01	0.31	-	0.03
	Depression score	36.41	6.98	36.23	6.95	-	-	-	-
	Total working hours per day	7.82	5.47	8.10	5.40	8.11	4.54	-	0.23
	Monthly sales of the family business	226900	734075	204200	503156	335900	3974462	-	0.10

Notes: The job-related statistics in MLDS are only for household head; national data come from the 2012 Tanzania mainland census. FMS participants are those who generated >0 verified record). The p-value is the significance of the difference of means or proportions of the FMS candidates and the MLDS respondents.

3.4 Research Questions: Bias in FMS Participation

The Dar FMS implementation revealed a range of social and technical challenges which lead to basic questions regarding the viability of a smartphone-based HTS in a context like Dar. As described in the previous section, an essential concern lies in possible biases resulting from non-random, non-response in the cooperation rate (effective recruitment), the participation rate (recruited participant), and the reporting rate (participants fully respond).

This paper does not focus on cooperation rate, due partly to lack of data. To give an indication, however, the sample descriptive statistics (Table 1) suggest modest possible bias in cooperation, based upon observed socioeconomics and demographics between FMS candidates (recruited), participants, and the city-wide population (approximated by the MLDS sample). Keeping in mind that, in contrast to MLDS respondents, FMS candidates were constrained to being over 18 years old, we can see that FMS candidates and participants: had higher reported levels of self-employment, had higher rates of being married and lower rates of never being married, were slightly older (p-value = 0.07) and slightly more educated, lived slightly closer to the CBD (p-value = 0.09), and had slightly lower levels of estimated wealth. Sample distributions for other socio-demographics are similar between FMS and MLDS. That said, we had no access to other data (e.g., on travel or time use characteristics) that would be necessary to gauge problems in travel behavior representativeness that may arise from the cooperation rate biases.

Table 2 Example user problems identified in Dar FMS implementation

Type of challenge	Reasons	Count
Deliberate non-cooperation (18)	Refusing to meet with the support team for technical issues	6
	Refusing to answer the phone interview	4
	Refusing to participate (e.g. tired of carrying the phone, disappearing)	4
	Selling the phone after deployment	3
	Constantly removing the SIM card	1
Limited ability to participate (32)	Mixed up username	14
	Poorly recalling activities	12
	Rarely, if ever carrying the phone	3
	Insufficient battery charge (e.g. due to low access to electricity)	3
Forced low degree of participation (22)	Phone being stolen	8
	Technical problems (e.g. battery problem, not showing data, internet connection problem, wrong phone time)	7
	Phone damage (e.g. dropped in the water)	4
	Poor Tigo network	3

Among recruited participants, a number of user problems can be identified from interviewer notes (Table 2), including deliberate non-cooperation (e.g., selling phone), limited ability to participate (e.g., forgetting to carry the phone), and forced low degree of participation (e.g., stolen or damaged phone). An important technical challenge related to the sparse cell tower coverage, resulting in large data gaps. Other technical issues related to the smartphone or app, including wrong system time, connectivity problems, battery problems, etc. These challenges

resulted in a relatively low number of fully verified days and may result in biased representativeness of the data collected.

We have two different ways to measure verification completeness: FMS diary-based, using the FMS diary and considering a "complete" day to be one in which every trip and stop has been verified; or interviewer-based, using the interviewer's evaluation of the day verification status. The latter is arguably a less rigorous definition of verified days since interviewers may tolerate incompleteness and still label a day as fully verified in the accompanying interview notes. Out of 581 candidates recruited (completed the pre-survey and received the free smartphones), 482 people (83%) provided at least one verified trip/stop, 354 people (61%) had at least one fully verified day using the interviewer-based criterion, and 329 (57%) had at least one fully verified day according to the FMS diary-based criterion. Out of 13,944 person-days (581 candidates times 24 active survey days), the FMS diary-based criterion indicates 1087 fully verified days (8%), or 2.1 days per candidate; the interviewer-based criterion indicates 1574 fully verified days (11%), or 2.8 days per candidate (Figure 3; Table 3**Error! Reference source not found.**).

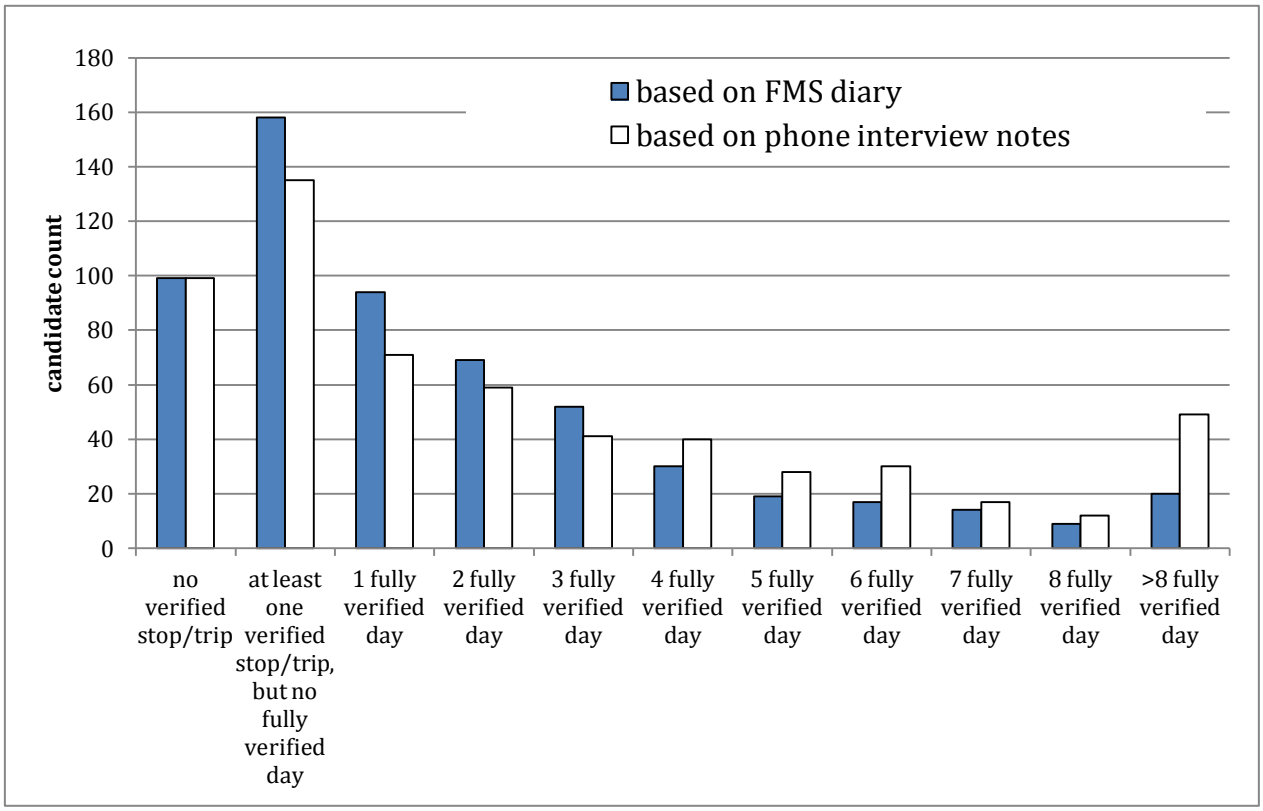

Figure 3 Distribution of FMS candidates according to their number of fully verified days

The rate of non-participation (candidates never verifying or only partially verifying) leads to concerns about the representativeness of the participants and the travel data collected. Does evidence of selective non-participation exist? We attempt to answer this question in three ways. First, we look for observable factors which explain the failure to participate at all (i.e., do not verify at least one day). Second, since the FMS design intended to collect verified data for multiple days (up to 3 weeks), we examine evidence which might explain an individual's total number of verified days. Finally, since theory suggests that characteristics of a specific day might

impact its likelihood of being verified (e.g., a more complex travel day increases verification burden), we explore individual and day-specific factors that predict day verification. The following section describes the methods used.

Table 3 Statistics on the number of fully verified days per person

	FMS diary-based		Interviewer-based	
	All participants	Participants with > 0 verified days	All participants	Participants with > 0 verified days
Mean	2.10	3.30	2.80	4.40
Median	1.00	2.00	2.00	3.50
Variance	6.73	6.59	11.76	11.28

4 METHODS

4.1 Modeling individual non-participation

We use the mixed logit technique (Hensher and Greene 2003) to model the likelihood of an individual never participating in verification (despite enrolling in the survey). The model includes random effects, for unobserved factors correlated within households (as a single household can have two participating individuals). The random effects framework follows statistical theory dealing with panel data (Cameron and Trivedi 2005; Greene 2003). We estimate the model with the glmmADMB package (Fournier et al. 2012; Skaug et al. 2012) in R. GlmmADMB uses the Laplace approximation. The model formulation for estimating an individual's likelihood of verifying at least one day is:

$$y_{ih} = \beta_{ih} S_{ih} + \beta_h S_h + \alpha_h + \varepsilon_{ih} \tag{1}$$

where y_{ih} is the binary choice of whether individual i from household h participated in FMS; S_{ih} and S_h are attributes associated with the individual i and household h respectively; the βs are vectors of estimable parameters; α_h is the random effects capturing the correlated error term within households; ε_{ih} is the independently and identically distributed (i.i.d.) noise.

4.2 Modeling an individual's number of verified days

Out of 484 participants who provided at least one verified trip/stop record, 135 (according to the interviewer-based criterion) or 158 (according to FMS diary-based criterion) have zero fully verified days. Given this censoring, ordinary count models may be inappropriate. As such, we employ a zero-inflated model, which allows zeros to remain in the count model, by estimating an individual's likelihood of being in the 'zero' group. Specifically, we use a zero-inflated Poisson (ZIP) model, with random effects, again to accommodate within-household correlation. In a ZIP model, the count response variable is assumed to be distributed as a mixture of a Poisson (λ) distribution and a distribution with point mass of one at zero, with mixing probability p (Hall 2000; Lambert 1992). We implement the model in Mplus 7.

4.3 Modeling the likelihood of an individual verifying a given day.

We also use mixed logit to model the likelihood of a day being fully verified. In this case, the model includes random effects for unobserved factors correlated within households and within individuals (since we are modeling many days among individuals). The model formulation for estimating whether an individual's given day is fully verified is:

$$y_{ihck} = \beta_h S_h + \beta_{ih} S_{ih} + \beta_{ihc} S_{ihc} + \lambda(T_{ki}) + \alpha_h + \alpha_{ih} + \varepsilon_{ihk} \quad (2)$$

where y_{ihck} is a binary variable or whether day k of individual i from household h is fully verified by interviewer call c; S_h, S_{ih}, S_{ihc} are attributes of household h, individual i, and call c; T_{ki} represents attributes related to an individual's day; the βs and λ are parameters to be estimated; α_h, α_{ih} are random effects for household and individual; and, ε_{ihk} is i.i.d. noise.

4.3.1 Measures of relative travel burden

As discussed, the raw sensor data collected via FMS are inherently noisy. This noise results in inaccurate inferences by the FMS algorithms. This might increase the verification burden for a given day (the model in equation 2, above). Drawing from Hanson and Hanson (1981), we derive four indicators to represent the expected relative burden, for each day of collected data. We calculated these variables for all user days, using the unverified inferences from the FMS algorithms. The unverified inferences reflect measurement error from the raw data and, consequently, relative verification burden.

The first indicator we calculated was the number of detected stops, a straightforward calculation direct from the FMS inferences. The second indicator, total distance traveled (km), uses the GPS traces recorded by FMS. For this, we filtered the raw GPS traces to consider only the high accuracy points (> 100 meters) for the travel distance calculation. The filtered traces were further smoothed such that only the points 30 meters away from a previous GPS point are included. If GPS traces were missing between two stops, then the total travel distance was calculated using the straight-line distance between two stops.

The third indicator, distance from home to activities' centroid (km), attempts to represent the how far from home an individual travels to perform a daily set of activities. We used the stops inferred by FMS for a given user on a given day to calculate the centroid of the activities. Let $S_i(x_i, y_i) \forall i = 1,2, \dots n$, be the n number of stops inferred by FMS for a given day and x_i, y_i be the latitude and longitude of the stops respectively. The activity centroid $S_c(x_c, y_c)$ is obtained by calculating the centroids: $x_c = \frac{\sum_i x_i}{n}, y_c = \frac{\sum_i y_i}{n}$. We then calculate the straight-line distance from the respondent's self-reported home location to her activity centroid.

The last indicator of burden we calculated, standard radius of activities (km), provides a measure of the extent of scattering of an individual's activities in space. The standard radius of activities is obtained by calculating the straight-line distance between activity centroid $S_c(x_c, y_c)$ and activity $S_j(x_j, y_j)$ where S_J is the farthest inferred activity from the centroid.

4.4 Spatial autocorrelation test: Moran's I

Except for possible unobserved factors correlated within households and individuals, spatial autocorrelation (lag or error) may exist. Incorporating both spatial and non-spatial random effects into generalized linear mixed models, however, adds considerable complexity. Instead, we conduct post-estimate tests for spatial autocorrelation, using Moran's I (Moran 1950) of the model residuals. Values of Moran's I range from -1 to +1. Negative values indicate negative spatial autocorrelation and positive values indicate positive spatial autocorrelation.

5 FINDINGS

5.1 Factors influencing individual non-participation

Do observable socioeconomic and demographic factors predict an individual's likelihood of participating in FMS? To answer this question we use three different definitions of non-participants in the model: (1) individuals who do not generate any verified record (99 observations); (2) individuals who do not have any diary-based fully verified days (227 observations); (3) individuals who do not have any interviewer-based fully verified days (252 observations). We tested a range of independent variables including available individual- and household-level socio-demographic and economic indicators (e.g., gender, education, age, job type, marriage status, index for wealth per household member, an individual's depression score, vehicle ownership, working hours) and relative location (distance from home to CBD). The final model estimation, presented in Table 4 excludes variables with $p>.40$ (in at least one of the models).

The results (Table 4) reveal little bias in likelihood of participating across most individual and household characteristics, irrespective of how we define non-participation. The only characteristics of potential concern are education and certain job types. With all three definitions, more years of education are correlated with a higher likelihood of participating, perhaps implying that lack of familiarity with technology could potentially be a barrier to scaling up smartphone-based methods among less-educated individuals. On the other hand, people employed in the public sector (government job or public company) are more likely to be non-participants, an interesting finding, perhaps representing some fear public employees may have of being 'tracked.' This result warrants further investigation. The Moran's I test of the residuals reveals no significant autocorrelation (non-participant by definition 1) or marginally significant autocorrelation (definitions 2 and 3).

Table 4 Results of Binary Logit Model Showing the Socio-demographic Factors Influencing the Likelihood of Non-participation

	Percent	Non-participant Definition (1): individuals have no verified record			Non-participant Definition (2): individuals have no fully verified day (FMS-based)			Non-participant Definition (3): individuals have no fully verified day (interviewer-based)		
		Estimate	Pr(>|z|)		Estimate	Pr(>|z|)		Estimate	Pr(>|z|)	
(Intercept)	-	-3.493	0.000	***	0.111	0.700		0.511	0.149	
Female	49%							-0.212	0.094	.
Male (base)	51%									
Education years	-	-0.123	0.029	*	-0.037	0.083	.	-0.043	0.028	*
ln(distance from home to CBD)	-							-0.106	0.216	
Type of primary job: Government	3%	2.131	0.038	*	1.059	0.019	*	0.768	0.039	*
Type of primary job: Private company	12%	0.676	0.334		0.218	0.480		0.016	0.946	
Type of primary job: Private individual(s)	11%	1.108	0.101		0.556	0.076	.	0.021	0.930	
Type of primary job: Public company	2%	1.853	0.047	*	0.265	0.598		0.241	0.568	
Type of primary job: self-employed	52%	0.589	0.301		0.261	0.295		-0.144	0.415	
Type of primary job: no job (base)	*20%*									
Total working hours	-				-0.032	0.068	.			
n		581			581			581		
sigma (household random effect)		5.236			0.626			0.362		
log-likelihood		-174.94			-343.66			-336.26		
AIC		365.90			705.30			692.50		
Moran's I of the residuals		-0.004	0.924		-0.045	0.101		-0.054	0.047	*

5.2 Factors influencing an individual's number of fully verified days

While only modest evidence of bias emerges regarding observable characteristics and individuals' likelihood of participating, what about for an individual's total number of verified days (recall that up to 3 weeks of data were to be collected)? We use two versions of the dependent variable: the FMS diary-based number of fully verified days; and the interviewer-based number of fully verified days. As discussed above, the latter represents an arguably less conservative definition of verified day, as interviewers may tolerate some incompleteness in the verification process (see, also, Table 3).

We use the same set of independent variables as in the previous model. The results (Table 5) indicate some biases in the total number of verified days, with some differences depending on the definition of day verification (diary-based versus interviewer-based). The probability of having zero fully verified days does not differ much across different socio-demographic groups, consistent with the participation model above. However, the count of verified days for the non-zero group is related with several socio-demographic factors.

With both definitions of the dependent variable, the likely number of verified days goes down for married and widowed individuals, individuals with a car driver's license, and people who work for government or for private individuals or are self-employed. The number goes up with age. Married people may have other prioritoes (although this effect is not statistically significant for polygamous cases); the widowed might have more housework burdens. People with a car driver's license may have more complex trip chains, dissuading verification. People who work in government may be concerned about being 'tracked" and the self-employed may be preoccupied with their own business. Older people may have more time for, or interest in, verifying days.

Other results are less consistent across the two definitions of the dependent variable: gender; working for a public company; living together with partners but not married; wealth per household member; monthly sales of family business. The fact that the results vary by definition of the dependent variable suggests a possible interviewer effect, although the differences are for the most part marginal in terms of statistical significance. This is an area worth further examination. For interviewer-based fully verified days, the small, but significant, negative Moran's I indicates slightly negative spatial autocorrelation, but the value close to zero should have little impact on the model results.

Table 5 Results of Zero-inflated Poisson Model Showing the Socio-demographic Factors Influencing the Number of Fully Verified Days

	Number of fully validated days (FMS diary-based)			Number of fully validated days (interviewer-based)		
	Estimate	Pr(>|z|)		Estimate	Pr(>|z|)	
Count part for non-zero group						
Female	-0.243	0.062	`	-0.088	0.408	
Male (base)						
Age	0.01	0.042	*	0.011	0.025	*
ln(distance to CBD)	0.091	0.314		0.114	0.14	
Type of primary job: Government	-1.059	0	***	-0.41	0.073	`
Type job: Private company	-0.176	0.383		-0.032	0.835	
Type job: Private individual(s)	-0.339	0.094	`	-0.458	0.004	**
Type of primary job: Public company	0.016	0.958		0.307	0.049	*
Type of primary job: self-employed	-0.305	0.034	*	-0.269	0.017	*
Type of primary job: no job (base)						
Married (monogamous)	-0.388	0.003	**	-0.273	0.039	*
Married (polygamous)	-0.352	0.361		0.013	0.961	
Living with partner, not married	-0.404	0.146		-0.582	0.006	**
Divorced	0.364	0.177		0.099	0.742	
Separated	-0.054	0.88		-0.07	0.828	
Widowed	-0.643	0.082	`	-0.421	0.038	*
Never married (base)						
Wealth per HH member	0.149	0.239		0.23	0.038	*
Depression scores	0.011	0.112				
Driving license - car	-0.232	0.097	`	-0.42	0.005	**
Driving license - motorbike	0.284	0.268		0.083	0.767	
Driving license - none (base)						
Monthly sales of family business (0 if the person is not self-employed)	0.001	0.083	`			
Probability to be in zero group						
Female				-0.353	0.088	`
Male (base)						
Education years				-0.068	0.025	*
ln(distance from home to CBD)	-0.217	0.12		-0.141	0.312	
Type of primary job: Government				1.127	0.062	`
Type job: Private company				0.014	0.969	
Type job: Private individual(s)				-0.113	0.767	
Type of job: Public company				0.317	0.629	
Type of job: self-employed				-0.292	0.289	
No job (base)						
Wealth per HH member	-0.366	0.183				
Intercepts						
Probability to be in zero group	-0.037	0.913		0.775	0.159	
Count part for non-zero group	0.818	0.033	*	1.31	0	***
N	581			581		
log-likelihood	-1014.724			-1184.589		
AIC	2075.448			2423.179		
BIC	2173.374			2538.136		
Moran's I of the residuals	-0.019	0.525		-0.065	0.017	*

5.3 Factors influencing the verification of a day

Finally, we turn to the analysis of factors influencing the likelihood of a given user day being verified. Beyond attributes of the individual and the household, the models also include the day of the week and characteristics of the verification call (whether the interviewer's verification call gets through and previous call's duration; the latter assumes that a prior long call will dissuade an individual from participating in the subsequent call). We also include the four indicators of day complexity in an attempt to estimate whether estimated relative travel burden for a given day influences the likelihood of that day being verified.

5.3.1 Measures of relative travel burden

Figure 6 shows the distribution of the four day-complexity indicators, derived from the unverified FMS output. Among 7,285 days with FMS sensor data, 767 days have only one stop detected. The distribution of total distance traveled, the distance from home to activities' centroid, and the standard radius of equivalent activity circle all concentrate close to zero. Except for the impact from one-stop days, the apparent predominance of short-distance travel also contributes to the shapes of the distribution.

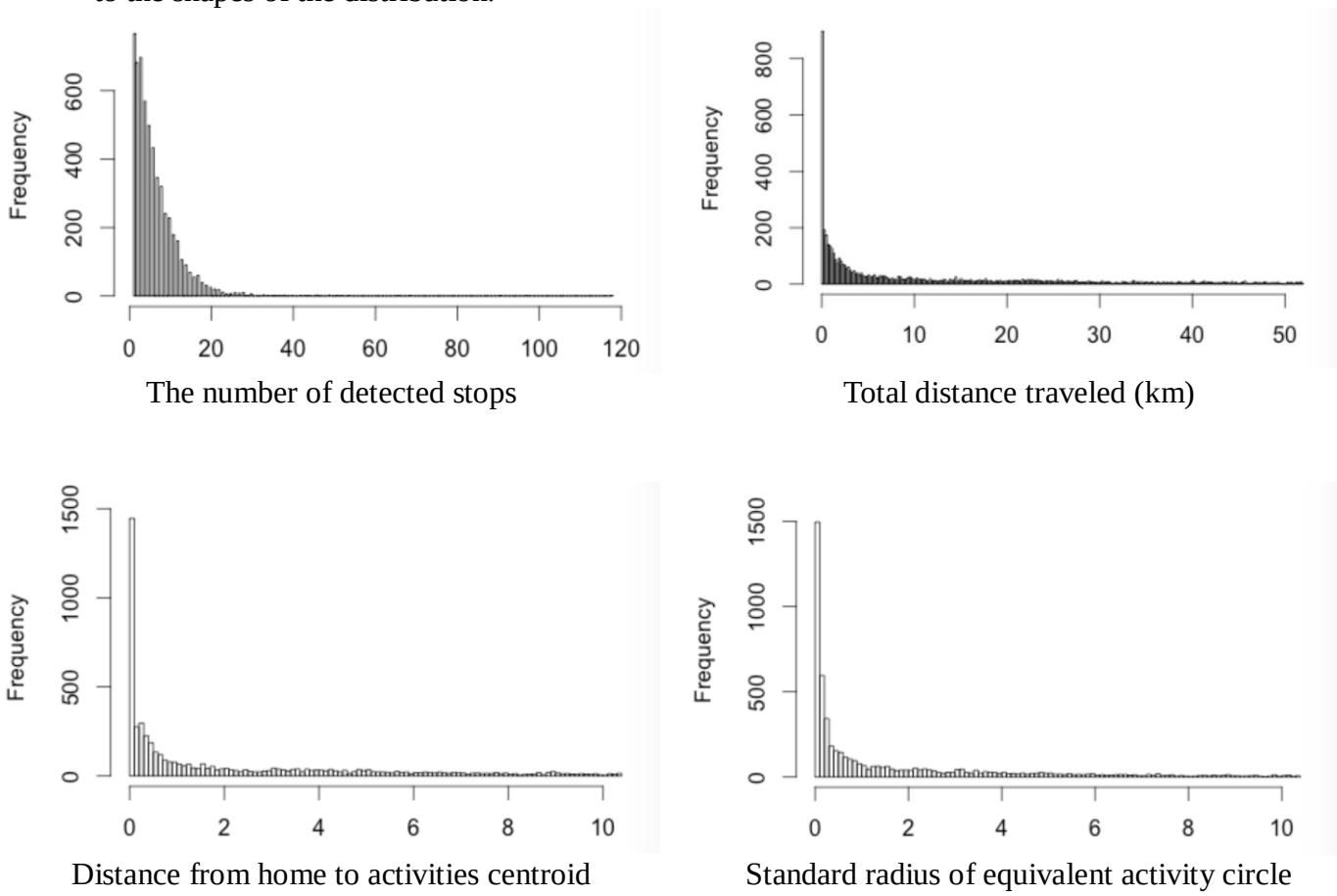

Figure 6 Distribution of four day-complexity indicators

5.3.2 Model Results

The day complexity indicators may introduce endogeneity in the model. Most basically, an individual's low level of interest in participation can hypothetically influence both the day verification status and the number of stops detected by FMS. For example, if a person is not very interested in the program and often leaves the phone at home or fails to charge the phone, a low number of stops will be detected. This low level of interest may simultaneously lead to non-cooperation in verifying the records. Unfortunately, no simple way to correct for endogeneity exists in this case. As a simplified method to examine the potential for endogeneity, we report model results with and without the day complexity indicators as independent variables; those models generate similar results. More sophisticated techniques such as a two step control-function and multiple indicator solution would be needed to formally test the endogeneity concern (Guevara 2015), but these require instrumental variables that are hard to obtain in this case.

We can draw several findings from the results (Table 6). First, counter-intuitively, whether the verification call gets through does not significantly influence the likelihood of a day being fully verified. Two related factors might be playing a role: the participants' actual willingness to verify and the experimental design which did not allow interviewers to add stops. If stops were not detected by FMS, interviewers would leave the activity diary intact and label the record as not verified. The model estimation results imply that the interviewer effect dominates the participant's effect. In other words, the day verification status is influenced more by the severity of the data gap than by participants' willingness to cooperate. Consistent with this implication, a long duration of the previous call does not significantly increase people's impatience to verify. This interpretation implies that interviewees were more willing to participate than interviewers were willing or able to invest in the verification effort.

Second, the number of detected stops highly significantly increases the likelihood of the day being fully verified. Ignoring possible endogeneity, this suggests that the more complete a travel diary detected by FMS for a given day, the less likely that day will be labeled as not verified by interviewers because of missing stops. Third, the negative influence of distance from home to activities centroid may imply that participants have more chances to charge their phone (and/or do not use their phone battery as much) if their destinations are close by. Fourth, the results for the socio-economic and demographic variables are similar to those relating to the number of days validated as estimated in the ZIP count model. Once more, we observe a significant negative effect for government employees (for both the FMS-diary and interviewer-based definitions of the dependent variable) and significant negative effect for private individuals and self-employed persons (for interviewer-based definition only). A consistently negative effect is also revealed for people with a car driver's license. Again, this may be due to relative travel complexity of car drivers. Those working long hours, on the other hand, have a higher likelihood of verifying a given day, perhaps representing relatively mundane (easy to verify) days, with fewer total trips. No apparent relationship emerges based on the day of the week. Finally, the intercept is very significant, indicating that many unobserved factors influence the day verification.

Table 6 Results of Binary Logit Model of Likelihood of a Day Being Fully Verified

		%	Day being fully verified (based on FMS diary)						Day being fully verified (based on phone notes)					
			Full model			Basic Model (w/o potentially endogenous variables)			Full model			Basic Model (w/o potentially endogenous variables)		
			Coef.	Pr(>|z|)		Coef.	Pr(>|z|)		Coef.	Pr(>|z|)		Coef.	Pr(>|z|)	
	(Intercept)	-	-1.710	1.30E-10	***	-1.610	1.60E-09	** *	-1.620	1.80E-08	***	-1.560	6.70E-08	** *
day mobility indicators (*potentially endogenous variables*)	The number of detected stops		0.0242	3.90E-08	***				0.012	0.005	**			
	Total distance traveled		-0.0007	0.223					-0.001	0.171				
	Distance from home to activities centroid		-0.0013	0.055	.				-0.001	0.039	*			
	standard radius of activity circle		0.0026	0.068	.				0.000	0.838				
day characteristics	Day of week: Tuesday	16%							-0.071	0.319		-0.072	0.312	
	Day of week: Wednesday	16%							-0.077	0.288		-0.080	0.267	
	Day of week: Thursday	16%							0.066	0.349		0.064	0.362	
	Day of week: Friday	12%							0.159	0.023	*	0.150	0.033	*
	Day of week: Saturday	12%							0.067	0.346		0.063	0.375	
	Day of week: Sunday	12%							0.065	0.364		0.056	0.432	
	Day of week: Monday (base)	15%												
call characteristics	Last call duration	-	0.000	0.225		0.000	0.218		0.000	0.558				
	Whether the call gets through		0.051	0.438		0.035	0.589		0.052	0.365				
socio-economic variables	Education years	-	0.019	0.107		0.017	0.143		0.018	0.145		0.017	0.180	
	Age	-	0.008	0.049	*	0.009	0.029	*	0.009	0.049	*	0.009	0.049	*
	ln(distance from home to CBD)	-	0.089	0.104		0.091	0.096	.	0.110	0.076	.	0.106	0.089	.
	Type of primary job: Government	3%	-0.664	0.006	**	-0.557	0.020	*	-0.623	0.021	*	-0.589	0.029	*
	Type of primary job: Private company	13%	-0.161	0.204		-0.115	0.366		-0.305	0.064	.	-0.308	0.062	.
	Type of primary job: Private individual(s)	11%	-0.197	0.154		-0.153	0.267		-0.364	0.033	*	-0.358	0.036	*
	Type of primary job: Public company	2%	0.039	0.876		0.191	0.453		0.030	0.917		0.083	0.771	
	Type of primary job: self-employed	54%	-0.097	0.342		-0.069	0.501		-0.275	0.043	*	-0.275	0.043	*
	Type of primary job: no job (base)	19%												
	Marriage: Married (monogamous)	66%	-0.285	0.007	**	-0.301	0.004	**	-0.213	0.066	.	-0.218	0.060	.
	Marriage: Married (polygamous)	2%	-0.187	0.486		-0.216	0.422		-0.166	0.552		-0.181	0.516	
	Marriage: Living together with partner, but not married	5%	-0.257	0.193		-0.227	0.249		-0.261	0.230		-0.256	0.241	
	Marriage: Divorced	1%	0.364	0.232		0.307	0.315		0.440	0.176		0.425	0.192	
	Marriage: Separated	2%	-0.239	0.369		-0.194	0.467		-0.061	0.816		-0.045	0.863	
	Marriage: Widowed	3%	-0.417	0.073	.	-0.446	0.055	.	0.092	0.692		0.085	0.715	
	Marriage: never married (base)	21%												
	Driving license - car	11%	-0.206	0.074	.	-0.136	0.236		-0.343	0.006	**	-0.324	0.009	**
	Driving license - motorbike	2%	0.326	0.173		0.409	0.089	.	-0.139	0.601		-0.133	0.618	
	Driving license - none (base)	87%												

	Total working hours	-			0.019	0.052	.	0.020	0.036	*
goodness-of-fit	**N**		7285	7285	7285			7285		
	sigma (random effect for household)		0.298	0.290	0.441			0.441		
	sigma (random effect for individual)		0.425	0.435	0.462			0.467		
	log-likelihood		-1945.6	-2402.6	-2989.4			-3186.2		
	AIC		4816.1	4847.1	6419.0			6428.3		
	Moran's I of the residuals		N/A because one person has multiple observations in the same location							

6 DISCUSSION AND CONCLUSIONS

The smartphone-based HTS remains a nascent practice. This is certainly the case in the Global South. The Dar es Salaam experiment with FMS reveals some of the associated potential social and technical difficulties of a prompted-recall smartphone-based HTS. Unlike previous FMS implementations, which used a self-completed web-based verification protocol, the Dar pilot retained interviewers as a way of minimizing expected challenges related to technological literacy and internet access. Such an approach seems likely to continue to be necessary for some time, at least in some places and/or among some cohorts, or until sensor data quality and backend intelligence enables fully automated detection of behaviors of interest.

We find modest evidence of bias in observable characteristics influencing an individual's likelihood of not fully verifying any days. Several socioeconomic and demographic characteristics apparently do influence, however, the number of days fully verified per individual. Similar apparent biases emerge when predicting the likelihood of a given day being verified. Sensor-based estimates of the day's travel burden were also associated with the likelihood of a day being validated, although these variables also run the risk of introducing endogeneity in the models. Overall, non-random, non-participation and non-response seems to exist. Possible solutions include enhancing mobile phone training of participants and/or improving the interviewers' verification procedure.

One shortcoming in the analysis, due largely to the experimental design, is that we cannot know if the lack of verification is due to the soft refusal or selective non-response of the interviewee or the interviewer and/or interactions among the two. A related problem was that interviewers were, by design, not allowed to add missing stops to the backend-generated travel/activity diary. This appears to have a larger influence on the data quality than the participants' willingness to participate. Well-trained and well-supervised interviewers remain important to HTS success, as does training users on the phones and the application. Better contextual information (e.g., on land uses, public transport routes) together with enhanced backend intelligence would improve inferences and should, thus, reduce respondent burden. But, obtaining quality contextual data in a city like Dar, where informal land uses and semiformal transit predominate, remains a challenge.

Various possible research extensions exist. The data collected in this case could be analyzed to assess implications of the behavior revealed (e.g., for travel equity outcomes) and how, for example, the behavior compares to that derived from other HTS for the city. Another interesting possibility would be to fuse the data from this purposeful collection approach with passive "big" data (e.g., from CDRs). Such data fusion would allow the imputation of refined individual and trip-based characteristics for massive CDR datasets, based on a small, smartphone-based HTS. Qualitative research could also lend insights into some of the reasons for apparent non-participation, non-response. Technological savviness and/or privacy concerns may play a role. For example, the lower participation likelihood of people who work for government potentially reflects concerns about being "tracked." Further research on instrument design (e.g., the cultural appropriateness of FMS' timeline-based diary) and technique (e.g., phone-based CAWI), or best combination of techniques, would also be valuable. Digital technologies do hold some promise

to reduce HTS burdens and improve the accuracy of the data collected. Nonetheless, they also introduce new problems and challenges which we are only beginning to understand.

REFERENCES

Alexander, L., Jiang, S., Murga, M., & González, M. C. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies*.

Ampt, E.S. and Ortúzar, J.D.D., 2004. On best practice in continuous large- scale mobility surveys. *Transport Reviews*, *24*(3), pp.337-363.

Bachu, P., Dudala, T. and Kothuri, S., 2001. Prompted recall in global positioning system survey: Proof-of-concept study. *Transportation Research Record: Journal of the Transportation Research Board*, (1768), pp.106-113.

Behrens, R., L. Diaz-Olvera, D. Plat, and P. Pochet. 2006. Collection of passenger travel data in Sub-Saharan African cities: Towards improving survey instruments and procedures. *Transport Policy*, *13*(1), pp.85-96.

Behrens, R., and E. Masaoe. 2009. "Towards Improved Travel Diary Survey Instruments in African Cities: Findings of a Comparative Experimental Application of Trip-Based, Placebased and Activity-Based Diaries in Cape Town and Dar Es Salaam." In Southern African Transport Conference, Pretoria. http://www.fut.se/download/18.778e24d112a169fd1c180006122/Behrens+-+Towards+improved+travel+diary+survey+instruments.pdf.

Behrens, R., D. McCormick, D. and D. Mfinanga, eds., 2015. *Paratransit in African cities: Operations, regulation and reform*. Routledge.

Bricka, Stacey G., Sudeshna Sen, Rajesh Paleti, and Chandra R. Bhat. 2012. "An Analysis of the Factors Influencing Differences in Survey-Reported and GPS-Recorded Trips." Transportation Research Part C: Emerging Technologies 21 (1): 67–88.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. Microeconometrics: Methods and Applications. Cambridge University Press.

Cottrill, C., F. Pereira, F. Zhao, I. Dias, H. B. Lim, M. Ben-Akiva, and C. Zegras. 2013. "Future Mobility Survey: Experience in Developing a Smart-Phone-Based Travel Survey in Singapore", *Transportation Research Record 2354*: 59-67.

de Leeuw, E. D., & de Heer, W. (2002). Trends in household survey nonresponse: A longitudinal and international comparison. In R. M. Groves, A. D. Dillman, J. Eltinge, & R. Little (Eds.), *Survey nonresponse*. New York, NY: John Wiley & Sons, pp. 41-54.

Demissie, M. G., Phithakkitnukoon, S., Sukhvibul, T., Antunes, F., Gomes, R., & Bento, C. (2016). "Inferring passenger travel demand to improve urban mobility in developing countries

using cell phone data: A case study of Senegal." *IEEE Transactions on Intelligent Transportation Systems*, *17*(9): 2466-2478.

Economic and Social Research Foundation (ESRF). 2015. Tanzania Human Development Report 2014. United Nations Development Programme and Government of the United Republic of Tanzania, Dar es Salaam.

Fournier, David A., Hans J. Skaug, Johnoel Ancheta, James Ianelli, Arni Magnusson, Mark N. Maunder, Anders Nielsen, and John Sibert. 2012. "AD Model Builder: Using Automatic Differentiation for Statistical Inference of Highly Parameterized Complex Nonlinear Models." Optimization Methods and Software 27 (2): 233–249.

Greene, William H. 2003. Econometric Analysis. Pearson Education India.

GSMA Intelligent. 2016. "The Mobile Economy Africa 2016." https://www.gsmaintelligence.com/research/?file=3bc21ea879a5b217b64d62fa24c55bdf&download.

Guevara, C. Angelo. 2015. "Critical Assessment of Five Methods to Correct for Endogeneity in Discrete-Choice Models." Transportation Research Part A: Policy and Practice 82: 240–254.

Hall, Daniel B. 2000. "Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study." Biometrics 56 (4): 1030–1039.

Hanson, Susan, and Perry Hanson. 1981. "The Travel-Activity Patterns of Urban Residents: Dimensions and Relationships to Sociodemographic Characteristics." Economic Geography 57 (4): 332–47.

Hensher, David A., and William H. Greene. 2003. "The Mixed Logit Model: The State of Practice." Transportation 30 (2): 133–176.

Jiang, S., Ferreira, J. and González, M.C., 2017. Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. *IEEE Transactions on Big Data*, *3*(2), pp.208-219.

Kagerbauer, M., W. Manz, D. Zumkeller. 2013. Analysis of PAPI, CATI, and CAWI Methods for a Multiday Household Travel Survey. In J. Zmud, M. Lee-Gosselin , M. Munizaga, J. A. Carrasco (ed.) Transport Survey Methods: Best Practice for Decision Making, pp.289 - 304

Kiunsi, Robert. 2013a. "A Review of Traffic Congestion in Dar Es Salaam City from the Physical Planning Perspective." Journal of Sustainable Development 6 (2): 94-103.

Kiunsi, Robert. 2013b. "The Constraints on Climate Change Adaptation in a City with a Large Development Deficit: The Case of Dar Es Salaam." Environment and Urbanization. http://eau.sagepub.com/content/early/2013/07/08/0956247813489617.abstract.

Lambert, Diane. 1992. "Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing." Technometrics 34 (1): 1–14.

Madre, J.-L., K.W. Axhausen, W. Brög. 2007. Immobility in travel diary surveys. *Transportation*, 34: 107-128.

Masaoe, Estomihi N., Romano F. Del Mistro, and George Makajuma. 2011. "Travel Behaviour in Cape Town, Dar Es Salaam and Nairobi Cities." SATC 2011. http://www.repository.up.ac.za/handle/2263/17373.\

McDonald, N., C. Porter, & R. Gorham. 2003." Availability of travel forecasting methods and data for air quality planning in developing countries." *Transportation Research Record 1859*, 110-116.

McGuckin, N., Keyes, M.A. and Liss, S., 2004. Hang-ups: Looking at non-response in telephone surveys. US Department of Transportation: https://www.fhwa.dot.gov/ohim/hang_ups.htm.

Melbye, Dea Christine, Lasse Møller-Jensen, Manja Hoppe Andreasen, Juma Kiduanga, and Anne Gravsholt Busck. 2015. "Accessibility, Congestion and Travel Delays in Dar Es Salaam–A Time–distance Perspective." Habitat International 46: 178–186.

Mkalawa, Charles Cosmas, and Pan Haixiao. 2014. "Dar Es Salaam City Temporal Growth and Its Influence on Transportation." Urban, Planning and Transport Research 2 (1): 423–446.

Moran, Patrick AP. 1950. "Notes on Continuous Stochastic Phenomena." Biometrika 37 (1/2): 17–23.

National Bureau of Statistics (NBS). 2014. Basic Demographic and Socio-Economic Profile Report: Tanzania Mainland. United Republic of Tanzania, Ministry of Finance.

Nawaranthne, K., F. Zhao, F. C. Pereira, C. C. Zegras, and M. Ben-Akiva. 2014. "The Impact of GPS Based Outdoor Activity Detection on Smartphone Battery Life." In 10th International Conference on Transport Survey Methods, Australia.

Ndibatya, Innocent, Justin Coetzee, and Thinus Booysen. 2017. "Mapping the informal public transport network in Kampala with smartphones: international." *Civil Engineering= Siviele Ingenieurswese* 25.1 (2017): 35-40.

Nkurunziza, Alphonse, Mark Zuidgeest, Mark Brussel, and Martin Van Maarseveen. 2012. "Modeling Commuter Preferences for the Proposed Bus Rapid Transit in Dar-Es-Salaam." Journal of Public Transportation 15 (2): 5.

Ortúzar, Juan de Dios. 2006. Travel Survey Methods in Latin America, in Peter Stopher , Cheryl Stecher (ed.) Travel Survey Methods: 1-18

Pinelli, F., Nair, R., Calabrese, F., Berlingerio, M., Di Lorenzo, G., & Sbodio, M. L. 2016. "Data-Driven Transit Network Design From Mobile Phone Trajectories." IEEE Transactions on Intelligent Transportation Systems, 17(6): 1724-1733.

Poushter, Jacob. 2016. "Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies." Pew Research Center. http://www.diapoimansi.gr/PDF/pew_research%201.pdf.

Rieser-Schüssler, N. and K.W. Axhausen. 2014. “Self-tracing and reporting: state of the art in the capture of revealed behaviour.” In *Handbook of Choice Modelling*, Edward Elgar Publishing Cheltenham, UK: 131-151.

Saddier, S., Patterson, Z., Johnson, A., & Chan, M. (2016). Mapping the Jitney Network with Smartphones in Accra, Ghana: The AccraMobile Experiment. Transportation Research Record: Journal of the Transportation Research Board 2581: 113-122.

Schafer, A. 2000. “Regularities in travel demand: an international perspective.” Journal of transportation and statistics, 3(3): 1-31.

Shen, L. and Stopher, P.R., 2014. Review of GPS travel survey and GPS data-processing methods. *Transport Reviews, 34*(3), pp.316-334.

Skaug, H., D. Fournier, A. Nielsen, A. Magnusson, and B. Bolker. 2012. “Generalized Linear Mixed Models Using AD Model Builder.” R Package Version 0.7 2.

Stopher, P., FitzGerald, C. & Xu, M. Assessing the accuracy of the Sydney Household Travel Survey with GPS. Transportation (2007) 34(6): 723-741.

Weiner, E. 2013. Urban Transportation Planning in the United States: History, Policy, and Practice. Fourth Edition. Springer, New York.

Wolf, J., Wilhelm, J., Casas, J. and Sen, S., 2013. A case study: Multiple data collection methods and the NY/NJ/CT regional travel survey. In *Transport Survey Methods: Best Practice for Decision Making* (pp. 71-90). Emerald Group Publishing Limited.

Zhao, Fang, Francisco Câmara Pereira, Rudi Ball, Youngsung Kim, Yafei Han, Christopher Zegras, and Moshe Ben-Akiva. 2015. “Exploratory Analysis of a Smartphone-Based Travel Survey in Singapore.” Transportation Research Record No. 2494: 45–56.

APPENDIX

Following Filmer and Pritchett (2001), we construct a linear wealth index for the surveyed households using asset ownership indicators. We apply principal components analysis (PCA) and calculate the wealth index based on the normalized variables and their respective loadings (Table A1) on component 1.

Table A1
First Nine Components from Principal Components Analysis on Household Durable Goods

Variable	C1	C2	C3	C4	C5	C6	C7	C8	C9
Telephone (land line)	0.201	0.240	-0.176	0.070	0.458	0.323	0.277	-0.194	0.114
Telephone (mobile)	0.097	0.006	0.003	-0.039	-0.111	-0.541	0.444	0.255	0.261
Television	0.747	-0.430	0.078	-0.179	0.153	-0.059	-0.024	-0.110	0.033
TV decoder (satellite)	0.732	-0.303	0.043	-0.127	-0.011	-0.020	0.025	-0.190	-0.067
Music System	0.543	-0.060	-0.120	-0.257	0.075	0.032	0.070	-0.046	-0.109
Computer	0.560	0.303	-0.194	-0.026	-0.151	0.082	0.054	0.057	-0.148
Air Conditioner	0.335	0.279	-0.186	-0.019	0.010	0.307	-0.053	0.034	0.531
Refrigerator	0.732	-0.034	-0.067	-0.057	-0.129	-0.028	-0.031	0.040	-0.068
Motor Vehicle	0.541	0.289	-0.132	0.042	-0.375	0.150	0.066	0.007	-0.058
Motor Cycle	0.275	0.115	-0.119	-0.057	0.557	-0.364	-0.032	0.100	-0.042
Bajaji (3-wheel MV)	0.153	0.095	-0.139	-0.048	0.261	0.371	0.527	0.123	-0.263
Radio	0.143	-0.120	0.264	0.488	0.087	0.105	-0.057	0.301	0.202
Video	0.669	-0.409	0.149	-0.164	0.165	-0.052	-0.050	-0.037	0.155
Tape Recorder	0.164	-0.044	0.184	0.167	0.312	0.323	-0.331	0.345	0.024
Fan	0.599	-0.399	0.138	-0.138	0.038	0.060	-0.114	0.033	0.015
Sewing Machine	0.229	0.221	0.058	0.252	0.034	-0.234	-0.026	-0.536	-0.074
Watches	0.451	0.038	-0.001	0.181	-0.250	0.075	-0.007	-0.011	-0.043
Beds	0.163	-0.147	0.145	0.181	-0.035	-0.060	0.440	-0.095	0.405
Iron	0.568	-0.098	0.103	0.236	-0.123	0.034	-0.025	-0.044	-0.101
Stove 1	0.618	0.311	-0.125	-0.055	-0.207	-0.082	-0.084	0.167	-0.002
Stove 2	0.015	-0.319	0.312	0.382	-0.197	0.279	0.201	-0.253	-0.081
Water Heater	0.511	0.268	-0.074	0.067	0.005	-0.043	-0.152	0.139	-0.116
Books	0.393	0.047	0.132	0.375	-0.027	-0.193	0.134	0.399	-0.098
Bicycle	0.210	0.249	0.075	0.368	0.300	-0.196	0.066	-0.116	-0.333
Carts	0.138	0.272	0.242	0.243	0.092	-0.148	-0.271	-0.235	0.160
Boat	0.085	0.312	0.603	-0.323	-0.077	0.068	0.085	0.042	-0.124
Outboard Engine	0.003	0.321	0.617	-0.263	0.019	0.036	0.097	0.046	-0.077
Water Pump	0.367	0.475	-0.091	-0.066	0.029	0.014	-0.077	-0.171	0.311
Trailer Tractors	0.002	0.254	0.605	-0.175	0.054	0.021	0.021	-0.021	0.103
Eigenvalue	5.25	1.938	1.679	1.326	1.214	1.173	1.082	1.067	1.023
Proportion Variation	0.18	0.067	0.058	0.046	0.042	0.041	0.037	0.037	0.035

Filmer, D. and Pritchett, L.H., 2001. Estimating wealth effects without expenditure data—or tears: an application to educational enrollments in states of India. *Demography*, *38*(1), pp.115-132.