CrossMark

We want it all: experiences from a survey seeking to capture social network structures, lifetime events and short-term travel and activity planning

Chiara Calastri[1] · Romain Crastes dit Sourd[1] · Stephane Hess[1]

Abstract Recent work in transport research has increasingly tried to broaden out beyond traditional areas such as mode choice or car ownership and has tried to position travel decisions within the broader life context. However, while important progress has been made in terms of how to capture these additional dimensions, both in terms of detailed tracking of movements and in-depth data collection of long term decisions or social network influences, surveys have tended to look at only a handful (or often one) of these issues in isolation, especially at the data collection end. Making these links is the key aim of the data collection described in this paper. We conducted a comprehensive survey capturing respondents' travel, energy and residential choices, their social environment, life history and short-term travel patterns. The survey is composed of a detailed background questionnaire, a life-course calendar and a name generator and name interpreter. Participants were also required to use a smartphone tracking app for 2-weeks. We believe that this is an unprecedented effort that joins complexity of the survey design, amount of information collected and sample size. The present paper gives a detailed overview of the different survey components and provides initial insights into the resulting data. We share lessons that we have learned and explain how our decisions in terms of specification were shaped by experiences from other data collections.

Keywords Travel survey · GPS tracking · Social networks · Life-course events · Data collection

✉ Stephane Hess
s.hess@leeds.ac.uk

Chiara Calastri
tscca@leeds.ac.uk

Romain Crastes dit Sourd
r.crastesditsourd@leeds.ac.uk

[1] Institute for Transport Studies and Choice Modelling Centre, University of Leeds, Leeds, UK

Introduction

Alongside major theoretical innovations, the field of travel behaviour research has over the last two decades been characterised by a fundamental re-evaluation of what is driving people's travel decisions. This has been accompanied by work into how to capture information on factors going beyond traditional level-of-service characteristics and socio-demographic information, with an increasing embrace of new data collection techniques, in particular based on mobile technologies.

A number of distinct strands can be identified. Perhaps the most prominent area has been the research into the role of attitudes, perceptions and plans (Molin et al. 2016; Daly et al. 2012; Choudhury et al. 2010). Notwithstanding the criticisms leveled at this area of research over the last two years (Vij and Walker 2016; Chorus and Kroesen 2014), capturing information on such *soft* factors remains a very active area of research.

A different area of research has encouraged analysts to look at the role of other people, both in terms of joint decision making (Arentze and Timmermans 2009) and the influence that someone's social and professional network may have on that person's decisions, even if travelling alone (Dugundji and Walker 2005; Maness and Cirillo 2016). This has again motivated extensive research into how to capture data on the interactions between people in their travel behaviour (Silvis and D'Souza 2006; Lin and Wang 2014; Axhausen 2005), where a particular focus has also been on the formation and maintenance of social networks and the way in which people interact with those in their network (Kowald et al. 2010; Calastri et al. 2017c, d).

While much of the above work has looked at the role of such non-traditional (from a travel behaviour research perspective) influences on choices, other research has questioned the wisdom of treating individual decisions in isolation. Much has been made of research looking at the inter-dependencies of travel and activity choices at the day-level (in activity based as well as time use modelling) (Bhat and Singh 2000; Arentze and Timmermans 2005), but it is clearly conceivable that interactions and influences cover a much broader time horizon. Paleti et al. (2013) have simultaneously analysed different short, medium and long term choice dimensions such as residential location choice, car ownership, work location choice as well as commuting mode and distance and find strong interdependencies in the choice continuum. They conclude that ignoring these correlations is not reflective of the true relationships that exist across these choice dimensions. Not only are there likely to be influences by past decisions, often in a seemingly unrelated context (e.g. residential location during childhood may drive mode choice decisions as an adult), but there is also scope for forward looking, e.g. making a commuting mode choice decision now with a view to changing car ownership next year. Some research has looked at inter-generational influences in travel behaviour. For example, Döring et al. (2014) analysed commute behaviour over different generations, finding that attitudes and residential location of the younger members of a family are associated with the same influencing factors as their parents. These are in turn determining the choice of commuting mode.

The final piece of the puzzle has been a very rapid uptake of mobile data collection approaches as well as a growing interest in longitudinal data sources that can help to understand some of the longer term influences. For the former, especially GPS surveys have grown in popularity, and they are rapidly replacing traditional travel diary surveys. The latter, while appealing from the point of view of following the same person over many days, is often beset by poor retention rates as well as a lack of data on short term decisions.

The various developments above are reasons for great excitement in the field. However, as is all too often the case in research, these are individual research efforts by separate communities within the field, with often little or no interaction amongst them, especially at the data collection end. If we accept the notion of a role for social networks impacts as well as attitudes and perceptions, then there is clearly scope for some interaction between the two. The same goes for the interplay between different choice dimensions, both within and across different time horizons. Capturing detailed data of a person's travel decisions using GPS tracking has reduced appeal if we do not at the same time understand the influence of past choices and life events for the same person.

Making these links was the key aim of the survey described in this paper, driven by our own experience of working with datasets from surveys focussing on just one aspect. We conducted a comprehensive survey capturing respondents' travel, energy and residential choices, their social environment and life history and a 2-week smartphone app travel survey. Guided by recent efforts in the literature, we try to capture multiple aspects that jointly play a role in shaping travel behaviour, and consider the possibility of interaction with other decisions, such as energy use. At the same time, we also make refinements to the individual components of this overall survey, again driven by our recent insights gained in work using GPS data Calastri et al. (2017a), name generators (Calastri et al. 2017c, d) and time use data (Calastri et al. 2017a, e).

While individual research contributions making use of this data will follow, this paper serves as a resource for other researchers who wish to break away from more one-dimensional surveys. We share lessons we have learned and also explain how our decisions in terms of specification were shaped by experiences with other datasets.

The remainder of this paper is organised as follows. The next section describes the different survey components in details. We then discuss the survey protocol and conduct, including the channels used for recruitment, the incentives used and the timing of the data collection. We will also discuss some of the challenges we faced and acknowledge the potential biases implied by the survey tools adopted. In the fourth section, we will describe the sample by providing some descriptive statistics related to the different survey parts. Finally, we will conclude by giving some recommendations to scholars interested in carrying out similar data collections and outline the next steps for the present project.

Survey structure

Our survey was made up of a number of separate components, which we will now look at in turn[1]. The first three components of the survey were completed using an online portal while the final part relied on a smartphone app. These first three components are hereafter referred to as BQ for Background Questionnaire, LCC for Life-course Calendar and NG for Name Generator.

Online survey components

In the first three components of the survey, respondents answered basic background questions, gave an overview of life events and provided a snapshot of their social network. The

---

[1] A complete survey script is available at http://stephanehess.me.uk/papers/615596-DECISIONS_survey_script.pdf.

average time respondents took for completing this part of the survey was 37 minutes (s.d. 42.4).

The background questionnaire

The BQ first collected essential socio-demographic information such as sex, age, education history, personal and household income, country of origin, occupation and marital status.

We next focussed on travel, residential location and energy, as well as susceptibility to influence by other people.

Respondents first provided data on transport mode ownership and usage (including frequency and mode split for commuting), as well as the availability of parking at home, work and other locations. This information is not only later used to prepopulate available options in the smartphone travel app (see "rMove" section), but is also crucial in understanding the trip patterns observed there.

They then gave detailed information on residential location and dwelling type, with extensive questions about energy sources (e.g. heating) used, as well as energy saving interventions such as double-glazing and insulated walls. People were also asked about what temperature they set their heating too (if under their control), how frequently they used different home appliances and what waste they recycled. Going beyond simple income questions, we also collected detailed information on monthly expenditure, with categories covering rent/mortgage, grocery, childcare, transport, communication, as well as money spent on utility bills.

The final component of the BQ included questions related to respondents' susceptibility to interpersonal influence, probing for agreement with ten carefully worded attitudinal statements. We used a series of six-points Likert scales ranging from "Strongly disagree" to "Strongly agree". They covered the level of importance that respondents attribute to other people's behaviour, especially in the domain of active travel and environmental friendliness, as well as the importance they attach to how others perceive their behaviour.

The life-course calendar

One important aspect of our survey concerns collecting information about respondents' past choices and behaviour. While such information is often collected in longitudinal surveys, these are affected by poor response rates and fail to make the link with short-term activities. We instead rely on a life-course or life-history calendar to retrospectively obtain data about events and activities occurred during the life of respondents (Caspi et al. 1996). Life-course calendars have been used to collect data on many different life events, such as education, employment, family events. Examples of research questions pursued using this type of data are timing of employment, receipt of welfare, marriage, cohabitation and children's schooling (Furstenberg et al. 1987) and timing of work and migration (Anderson and Silver 1987).

Figure 1 shows the LCC used in the present study. The list of events shown on the left-hand side is pre-populated on the basis of BQ information provided by participants about their education and employment history, their past home location, important relationships, children as well as cars currently and previously owned. We implemented an easy-to-use tool in which respondents simply need to click on any point on the timeline and then drag and drop the blue line to indicate start and end points. Respondents could change the occurrence and duration of events even after creating them in the tool.

Fig. 1 Life-course calendar: example

While all education *events* were presented, only up to and including the three most recent employments were shown, while, for residential location, only those after moving out of the parents' home were presented in the calendar grid (with others covered in the BQ). The number of years shown in the grid depends on the age bracket of the respondent.

As explained by Freedman et al. (1988), there are two main advantages of this tool. First of all, the visual and mental relation of different kinds of events provided by the LCC improves the quality of the data by providing reference points and preventing time inconsistencies. Secondly, the visual aid of the calendar tool eases the task of listing a potentially high number of different short events, a task that would be more difficult in traditional surveys. Especially the first point is apparent from Fig. 1. In our example, the year 2007 marked a change in employment, address and relationship. Remembering that these changes occurred at the same time will aid the completion of the survey.

A number of past studies have linked changes in life events to mobility decisions making use of data collected via life-course calendars. Beige and Axhausen (2012) use of a 20-years longitudinal retrospective survey, showing how turning points in life such as relocation or marriage are connected to one another as well as to long term mobility decisions. Similarly, Schoenduwe et al. (2015) observe that half of the changes in the travel mode for work trips coincide with a key life event (in the same year), while this is true for only around 30% of changes in the numbers of cars per household. Our survey allows us to capture such links too.

The name generator and name interpreter

A key area of research activity has looked at the role of social networks in shaping travel decisions (van den Berg et al. 2013; Carrasco et al. 2008b; Lin and Wang 2014; Calastri et al. 2017a). This requires a snapshot of the composition and influence of a respondent's network of social contacts.

A NG is a tool originating from the sociology literature used to collect information about egocentric social networks. It takes the form of a table that participants are asked to fill in with the names, nicknames or initials of the members of their social network. A specific survey question is presented, and the specific formulation is instrumental to the scope of the study. Studies making use of NGs to subsequently estimate the number of contacts have suggested an overall network size of around 1500 (Killworth et al. 1990; Freeman and Thompson 1989). In order to have a manageable list of social contacts, most studies do not ask the generic question "Who do you know?" but more specific questions depending on the research objectives, such as "Who are your three closest friends?" or "Who would watch your home if you were out of town?" (Campbell and Lee (1991)).

The approaches to define the boundaries of the network differ depending on the part of social network that researchers aim to capture. For example, some studies ask respondents to report the names of those with whom they interact during a certain period of time [e.g.] (Pike 2014), while others capture those who are emotionally close to the ego, i.e. an affective approach [e.g.] (Carrasco et al. 2008c). Some studies also focused on contacts providing/receiving a specific form of support [e.g.] (Burt 1984).

NGs are generally followed by *name interpreters*, which gather additional information about the named contacts and their relationship with the respondent, such as socio-demographic characteristics, time they have known the respondent and circumstances where they met him/her, frequency of interaction. A known challenge is represented by finding a

balance for the amount of information asked for each social contact, especially as in most surveys the number of social contacts that can be reported is not limited (Marsden 1990).

Travel studies have recognised that social networks are likely to be more relevant for travel for social and leisure purposes and have consequently prompted respondent to report the members of their social network who could most likely have an influence on decisions regarding this type of travel, i.e. emotionally closer contacts (Carrasco et al. 2008c and people with whom respondent spend their free time with (Kowald et al. 2010). Our recent work (Calastri et al. 2017d, c), in line with what has been done by e.g. Frei and Axhausen (2007), has used data capturing using either of these approaches. We see value in combining the two approaches to more fully capture people's network in our survey.

We first adopt the Kowald et al. (2010) approach and ask respondents to report those people with whom they interact outside of work. The exact wording of the NG question is "Please list the people with whom you choose to regularly interact outside of work, either in person or via phone or digital media." Respondents were also told to report just the name and the first letter of the surname (for privacy reasons). In the next line, they could also read "We provide you with 30 spaces below, but please feel free to use just as many as you need, and focus on who you normally stay in contact with". An example is shown in Fig. 2. In this first screen, respondents were also asked to specify the type of relationship with each of their social contacts from a drop-down menu. The possible options were *Partner/spouse*, *Parent*, *Sibling*, *Friend*, *Other relative*, *Colleague*, *Other acquaintance*. Thirty lines were available in this screen for respondents to list their social contacts, but if needed, they could add ten additional spaces by clicking on a "I want to add more people" button at the end of the page.

Upon selecting the "All relevant people entered" option, an additional prompt was used to ensure respondents had not omitted anyone by asking "Is there any person with whom you spend time or you have an affective relationship that you did not include in the list?", thus seeking to also capture the angle covered by Carrasco et al. (2008c).

After completing the NG part, the survey moved to the name interpreter questions. For each social contact, the respondent was required to select the sex, age range, city in the United Kingdom (or state whether the person lived abroad) and frequency of interaction in person, by phone (separately by call and text), by email and by online social networks.

Fig. 2 Name generator: example

Seven different categories were possible for the frequency of interaction, ranging from "Multiple times per day" to "Never".

Smartphone travel and activity app

After completing the three online survey components, respondents were invited to download and install a smartphone app which would collect data on their travel and activities for a period of two weeks.

rMove

The use of mobility apps as an alternative to paper based diaries has increased over recent years. They have the advantage of automatic and accurate collection of location and duration information for trips, and the potential of reduced respondent burden by learning about regular trips over time. This can potentially reduce problems of respondent retention over a longer survey period. The collection of multi-day data has great advantages in that it allows to observe habits and patterns in time use and travel that give a complete picture of real-life behaviour (Minnen et al. 2015; Jara-Díaz and Rosales-Salas 2015). A growing number of smartphone based GPS apps are available, where our work made use of a heavily customised version of rMove Resource Systems Group (2017), an app previously used in several US-based household surveys [e.g.] (Greene et al. 2016). The app was adapted for the specific needs of this survey, both in terms of suitability for the UK context,[2] and to meet our specific requirements, in particular the links with other survey components.

Participants were invited to download the app from GooglePlay[TM] or iTunes[TM] immediately after completing the online survey. They then received a personal activation code by email approximately 24 h after completing the survey, which ensured that the app interface was customised for each specific respondent. This made use of BQ and NG data on vehicle availability and social network members, helping survey engagement by making the app relevant to each person while also improving the accuracy of the data we collected.

rMove passively records travel data such as position, speed and route using multiple sensors, including GPS. For each completed trip, participants are asked to fill in short surveys, asking them questions about their travel mode, trip cost (if the trip was by taxi or public transport) and trip purpose. Over time, the app starts *learning* about regular trips and prepopulates the answers and only asks the respondent to confirm. Maps are used to help recall but also to allow users to add stops or correct a trip, if not displayed correctly. Figure 3 shows screenshots from the iOs version of the app where a trip is first displayed, then the respondent is asked to confirm whether the trip is correct and then to answer questions about the trip mode and purpose.

Respondents were asked to indicate who was with them during the trip, and, in the case of a leisure or social activity, whether someone joined them at destination. They could select contacts that they had listed in the NG as well as indicate that other people were present. The time burden required by the app varied depending on individual levels of travel activity, but decreased over time as the app started learning about frequent trips and destinations.

The app also contained a specific feature to collect information about in-home activities. Each day, respondents were asked to indicate which activities they performed while at home during the previous day. The list included cooking, cleaning, do-it-yourself (DIY),

---

[2] To meet EU data protection requirements, a server based in Ireland was used for data storage.
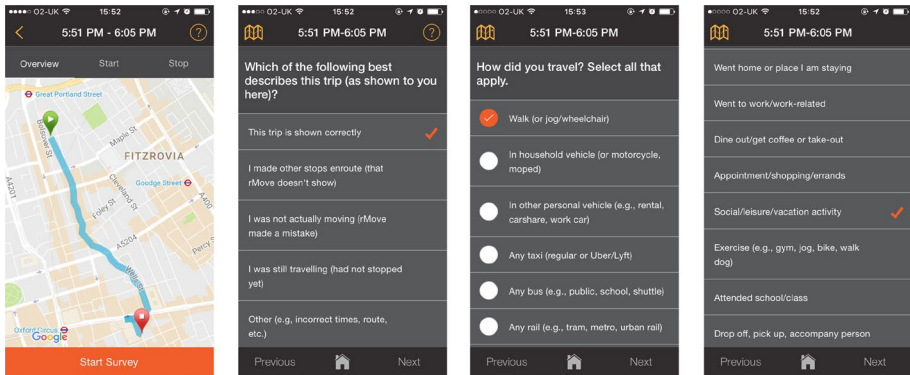
Fig. 3 rMove smartphone app screenshots

eating, exercising, playing games, listening to music, online activities, talking on the phone, reading, socialising, watching TV, working. Twice a week, on a weekday and on a weekend day, respondents were also asked to indicate the time intervals during which such activities were performed, as well as the home appliances used.

The feedback system

As mentioned above, one of the rising trends in travel behaviour research looks at understanding social influences on travel and activity behaviour. We specifically looked at the role of feedback in this context, with a view to understand how respondents could be *nudged* to change their behaviour.

The sample of respondents was split into three different groups (or treatments). The first (control) group used the app for two weeks without receiving any feedback information. The second (treatment 1) group used the app for one week and then received a digital feedback sheet on their activities. This contained a breakdown of the time spent travelling per day by each mode, and the resulting CO_2 emissions as well as calories burnt by active travel. An example of the feedback received by this group is shown in the left panel of Fig. 4.

The third group (treatment 2) received a similar feedback sheet (see right panel of Fig. 4), with the difference that the respondents in this group could compare their performance with other respondents taking part in the study. In particular, the comparison was made with people who were "similar" to the respondent in term of socio-economic status. This approach was inspired by studies such as the Quantified Traveller (QT) experiment (Jariyasunant et al. 2014), which combined the use of a tracking app and the online social feedback system. The QT study was focused on testing the *Theory of Planned Behaviour* and therefore largely focused on attitudinal questions, while we aimed at investigating changes in behaviour occurring during the second week. Moreover, we included a control group and work with a larger sample, overcoming some of the limitations of that study.

Data collection and post-pilot changes

The data collection for the present survey started in November 2016 and ended in April 2017, after two pilot studies aimed at testing the different survey components, one in
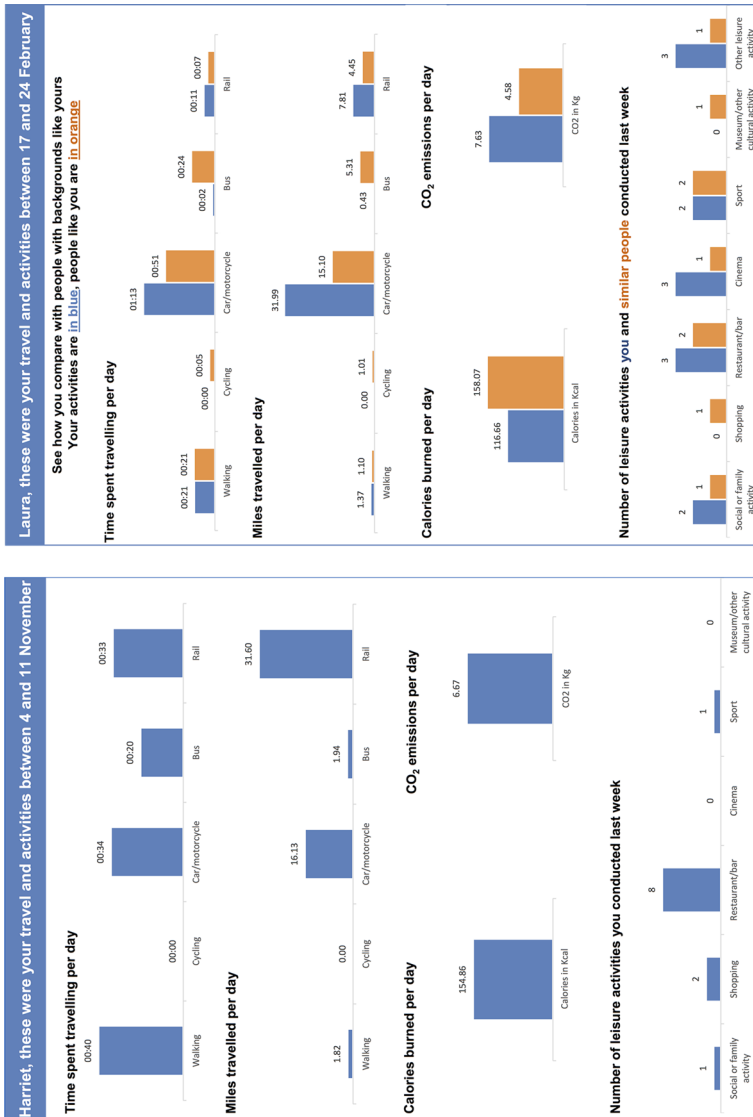
Fig. 4 Treatment 1 (left) and Treatment 2 (right) user feedback examples

January 2016, mainly involving University staff and students, and one in April–May 2016, for which a sample of 70 people in the city of Leeds were recruited. Different incentive structures were also tested and respondents were asked to provide feedback on this specific point. A detailed description of the recruitment strategy for the main survey is given below, after we cover some of the changes made after the pilot survey.

Insights from pilot survey and subsequent changes

Our survey represents a very rich and complex data collection effort, where a vast amount of information relating to different choice domains are gathered from respondents, and important trade-offs between the amount and quality of information needed to be considered. As expected, changes and modifications had to be implemented with respect to the initial survey design, and several parts of the BQ and LCC were cut or shortened to minimise respondents' burden and focus on aspects deemed more important for our research objectives. The initial survey design differed from the final implementation, and we wish to share some of the challenges we faced, hoping that this might help researchers overcome certain obstacles in future efforts. In addition, the survey tools adopted in the present data collections imply potential sources of bias that we wish to acknowledge, explain how we addressed these and make suggestions about areas where methodological contributions are most needed in the future.

Sampling technique

Our initial sampling plan focussed on the use of "snowball" technique, which can help capture population-wide social networks. In this sampling method, initial respondents or "seeds" are asked to name their social contacts who are subsequently invited to take part in the survey themselves and then asked to also recruit the members of their network. This procedure can be repeated for a given number of iterations, until the desired sample size is reached (Kowald and Axhausen 2014).

Due to ethical constraints, we could not collect the contact information of the alters and therefore we could not invite them to take part in the survey, an approach at least partly applied in the Swiss study described by Kowald and Axhausen (2014) where respondents could choose to send out the invitation cards or rely on the research team for this. We encountered strong resistance from respondents to agree to send the email to their social contact at the end of the survey. Focus groups and interviews carried out after the pilots revealed that the main worry of respondents was to disturb their social contacts by sending them the survey and making them feel morally obliged to complete it. We believe that a cultural difference between the UK and Switzerland may be at play in this case, as the respondents' motivations seemed to go beyond personal burden or privacy concerns, but relate to social norms.

It is important to notice that while snowball sampling has been applied in the UK, this has been done with the aim of investigating hidden populations, such as criminal gangs (Patrick 1973), non-heterosexual women (Browne 2005) and lone mothers (Duncan and Edwards 1999). To our knowledge, there is no study that was not aimed at capturing such a hard-to-reach population which successfully collected a snowball sample in this country. The main phase of the survey thus relied on a single level egocentric sampling approach.

Incentives

A study requiring such a level of involvement needs to provide adequate incentives to make sure that respondents will sign up and complete the survey. Several different incentive structures have been attempted in the pilot studies. Similarly burdensome research studies provided monetary incentives to each participant (Kowald et al. 2010; Montini et al. 2014; Greaves et al. 2014). In particular, Kowald and Axhausen (2014) argue that a monetary reward is the best way to incentivise participants due to its universally understandable nature. We rewarded all respondents who completed the entirety of the survey with a £25 shopping voucher.

Number of contacts reported in the name generator

The NG provides respondents with a limited number of spaces to report their network. As specified in "The name generator and name interpreter" section, respondents in the final version of the survey are initially provided with 30 lines to input the names of their social contacts, and they can use 10 extra lines if needed. Such upper limits were designed on the basis of survey testing and findings from previous studies. Participants in a study in Toronto (Carrasco et al. 2008a) reported a mean of 23.76 alters; a study in Chile, surveying affective networks including both very close and somewhat close social contacts, reports a mean network size of 22.24 (Carrasco et al. 2013), another study in Eindhoven, based on a 2-day interaction diary and a social network survey, reports an average network size of 23.28 alters (van den Berg et al. 2009). Two different studies from Switzerland found slighty lower mean number of alters, namely 12.35 in an egocentric survey (Frei and Ohnmacht 2016) 14 in a snowball survey (Kowald et al. 2010).

Given these figures, in our pilot we presented respondents with 15 slots that they could fill in with names of their social contacts (so that the layout could be easily accommodated in a single screen of an internet browser). If needed, they could add 10 additional slots, up to a maximum of 60 spaces, a number rarely reached in previous studies. We observed that the tool we provided had an effect on the number of contacts that participants were naming. Figure 5 shows that when 15 slots where shown, a large number of respondents listed 10 or 15 contacts. While resulting from a small sample, this statistics gives us an indication of the fact that respondents might have felt like they needed to fill the grid they were provided with, or provide a round number of contacts. In a number of pilots studies, we observed that respondents never named more than 30 contacts, therefore the final version of the survey offered 30 spaces, with the possibility to add an additional 10, for a maximum of 40. As visible from the orange line in Fig. 5, in the final survey we again observe a peak at 30 named contacts. We believe that this is a bias due to the instrument in question that could not easily addressed in the present study, but indeed requires further attention in the future. In both the pilot and the final survey, several respondents also only list one contact. While this will require further investigations, it could be due to fatigue effects.

Smartphone app issues

A well-known issue with surveys making use of smartphone apps is that of battery drainage. The reported travel activity could be affected by issues such as flat batteries or users turning off localisation services to preserve power, especially during days involving substantial travel activities. Unfortunately, the app does not allow us to check whether this
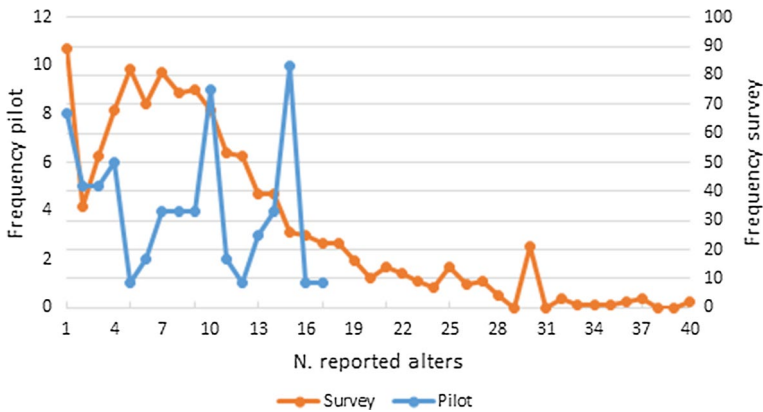
Fig. 5 Number of social network members in pilot and final study

is the case. Participants could have also actively prevented tracking for privacy concerns while performing specific activities. These issues are common to all studies making use of travel apps, and our decision to use such a tool is motivated by the fact that we concluded that smartphone apps were a better option than traditional travel diaries and GPS loggers. The former have indeed been shown to imply under-reporting in the number of trips (Wolf 2006; Zmud and Wolf 2003). Bricka and Bhat (2006) report descriptive results from a study comparing the number of trips reported by Kansas City drivers in a CATI survey versus those recorded by the GPS devices. They found that 71% of respondents accurately reported all their trips, 29% missed at least 20% two trips, 10% three trips, 5.5% four and 14.5% missed 5 or more trips.

Traditional diaries have also been shown to result in overestimation of trip durations (Kelly et al. 2013) and other inconsistent or missing trip information as well as missing route choice data (Bhat et al. 2005). Studies using GPS loggers underlined the fact that participants found it hard to remember to carry the device with them Bohte and Maat (2009), while personal smartphones are rarely forgotten. We address the issues resulting from the data collection process at the data processing stage, censoring those participants who are likely to have not properly used the app.

Recruitment strategy

Participants were mainly recruited in the greater Leeds (UK) area, although people living elsewhere in the UK were also invited to take part. Participants were recruited through a number of different means. Our most reliable source came in the form of the *Leeds Citizens' Panel*, a repository of Leeds residents willing to engage in surveys and administered by *Leeds City Council*. A related group of people consisted of *Leeds Council* staff. We also made use of a number of commercial and community mailing lists as well as paper-based flyers and letters, manually distributed in the Leeds area. Flyers were mainly distributed in the city centre of Leeds, while letters were posted in a representative range of residential areas. A limited number of students were invited to participate by distributing flyers on campus at the University of Leeds.

As the present project is research-driven, the sampling strategy aimed at collecting a sample entailing a good variety in terms of socio-demographic characteristics as well as mobility and activity behaviour, rather than achieving representativeness of the population. The mere fact that respondents were supposed to use their own smartphones certainly implies a selection bias, probably leading to over-sampling of younger and high-income people. This is confirmed by statistics for the UK showing that older individuals from lower social grades are substantially less likely to own a smartphones than younger people from higher social grades (Local Level UK 2014).

Previous experience suggests that long duration surveys can only succeed if the research team can gain respondents' trust (Kowald and Axhausen 2014; van den Berg et al. 2009). In order to achieve this, the recruitment material (e-mail/flyer/letter) contained detailed information about the research team, the importance of the project for scientific research, the type of commitment implied by taking part and the incentive provided. Details were also given on the ethical clearance obtained for the survey. Respondents were able to contact the research team via e-mail, phone (both by call or text message) at any time and the team would respond promptly. Respondents needed to fulfil a number of criteria to be suitable to take part in the study. They needed to be aged 18 and over and own an Android or iOs smartphone with a phone data plan (for the smartphone app).

Table 1 shows the total number of people contacted and recruited as well as a breakdown of completion rates by survey component. As expected (given the highly burdensome nature of the study), the overall response rate is low, 2%. A low response rate should not come as a particular surprise for a survey of this nature, both given the amount of data collected, the type of information collected off respondents, and the time horizon over which respondents need to stay involved in the survey.

The number varies across recruitment sources, with very poor performance of flyers and letters and commercial mailing lists and good performance of the *Leeds City Council* mailing list and *Leeds Citizens' Panel*. In some ways, these differences are expected, given that respondents who have signed up to community mailing lists and citizen panels are likely to be more open to participating in surveys, where the same also applies to council staff who have a greater interest in local affairs. These are cases where the message was sent by a source deemed trustworthy by the recipients, and this probably helped respondents believe that the survey was worth their attention and that there were guarantees that they would be rewarded upon completion.

The average retention rate throughout the survey is 26%, and we observe more stability in this statistic across recruitment sources.

The work of Axhausen et al. (2015) seeks to establish a relationship between response burden and response rates. The approach used is to compute a response burden index (RBI) as a function of the number and types of questions asked. A comparison across numerous studies can then be used to establish a quantitative relationship between this response burden score and the response rate.

This approach is particularly appropriate for surveys with a uniform format across respondents and with specific types of questions, such as stated choice. In these surveys, the response burden is driven by the survey design and it is then natural to expect lower response rate from more burdensome surveys. In our survey, this rationale particularly applies to the BQ, for which we obtain a RBI score of 173. The response rate for this part of the survey is equal to 81.1%, which is slightly lower than the one predicted using the tool described in Axhausen et al. (2015).

However, in the NG and LCC, the response burden is driven by the respondent rather than the survey itself. More importantly, higher survey engagement leads to increased

Table 1 Recruitment sources and survey completion statistics

	Contacted	Signed-up	BQ	LCC	NG	rMove (install)	rMove (2 weeks)
Total							
N	27,500	1747	1416	1,290	1109	643	452
% contacted	–	6%	5%	5%	4%	2%	2%
% signed-up	–	–	81%	74%	63%	37%	26%
Flyers and letters							
N	10,000	117	101	94	83	44	31
% contacted	–	1%	1%	1%	1%	0%	0%
% signed-up	–	–	86%	80%	71%	38%	26%
Commercial mailing list							
N	10,000	155	131	123	114	52	36
% contacted	–	2%	1%	1%	1%	1%	0%
% signed-up	–	–	85%	79%	74%	34%	23%
Leeds University Students and Staff							
N	2500	163	146	138	122	82	54
% contacted	–	7%	6%	6%	5%	3%	2%
% signed-up	–	–	90%	85%	75%	50%	33%
Community mailing lists							
N	2500	283	229	211	176	92	69
% contacted	–	11%	9%	8%	7%	4%	3%
% signed-up	–	–	81%	75%	62%	33%	24%
Leeds Citiziens Panel							
N	2000	465	382	334	275	157	117
% contacted	–	23%	19%	17%	14%	8%	6%
% signed-up	–	–	82%	72%	59%	34%	25%
Leeds Council Staff							
N	500	285	227	212	189	126	84
% contacted	–	57%	45%	42%	38%	25%	17%
% signed-up	–	–	80%	74%	66%	44%	29%
Unknown							
N	–	279	200	178	150	90	61

Table 2 Average response time for online survey and BQ only

	Online survey response time	BQ response time
All	36.99	16.3
Finished online survey only	34.96	16.32
Finished online survey and used rMove for 2 weeks	40.01	16.25

burden. As an example, a respondent who names more alters will also need to answer more questions about those alters. The same goes for the LCC, where more reported events leads to a greater burden. Finally, respondents who use the smartphone app as required will again face more questions and need to provide more data. For the components other than the BQ,

there is thus no easily justifiable way to determine the response burden given that the number of questions asked is a direct function of survey engagement.

Table 2 shows that increased burden correlates positively with increased participation for those components of the survey that are not uniform across respondents (the LCC, NG and rMove). In the table, we show the average completion time for the entire online survey ("Response time") and for the BQ only ("BQ response time"). This table shows that there are no substantial differences in the time taken to complete the BQ between the people who only complete the online survey and those who also use the smartphone app. This makes sense, as the burden is constant for this part of the survey. But when looking at the time taken to complete the entire online survey (which also includes the LCC and NG) we see that those who then also use the app (and therefore are more involved with the survey) take longer also to complete the online survey, showing that there is a link between the respondent driven burden and involvement.

As a further example, we see that respondents who complete the online survey and RMove name on average 10.98 alters in the NG, while for those who do not complete RMove, this figure is only 9.48. The number of life-course events is also higher for those that complete RMove, at an average of 13.81 compared to 13.25.

Nevertheless, if our interest was to purely predict response rates for the different components of the online survey, we believe that we could make reasonable assumptions based on previous work or available statistical data and compare the predicted response rates with those obtained. Therefore, we computed the RBI for the LCC drawing on available UK statistics about the number of residential locations, cars, relationships and for the NG making the assumption that people would name 14 alters (like in Kowald et al. (2010)). We obtained an RBI for the LCC and for the NG, respectively, of 128 and 392, corresponding to predicted response rates of 80.3 and 66%, just slightly higher than the real percentages we observed in the survey work.

Preliminary sample analysis

In this section we present overall sample characteristics, with particular attention paid to the differences between the groups of people who reached different levels of completion of the survey. This will allow us to assess whether survey engagement can be related to differences in participants' characteristics.

Background questionnaire

Table 3 reports the main socio-demographic characteristics of respondents collected in the BQ, split by the level of advancement in the survey, i.e. for people who only completed the BQ, for those who also completed the LCC, the NG and finally for those who completed the survey in all its parts, including using the smartphone app for two weeks.

Table 3 shows that the sample includes a high share of women, young people and people with a high level of education. The average age of the sample, computed by using the midpoint of each category, is 39.5. Many surveys tend to have a higher share of women than men, and we believe the low number of elderly people was to be expected given the format of the research, including an online survey and a smartphone app.

We find that the percentage share of the different socio-demographic characteristics remain fairly stable across different people who have achieved different levels of survey

Table 3 Socio-demographic characteristics: comparison across survey completion stages

Total	Completed BQ 1416		Completed BQ and LCC 1290		Completed BQ, LCC and NG 1109		Completed online survey and used rMove for 2 weeks 452	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Gender								
Female	807	56.99	736	57.05	623	56.18	262	57.96
Male	609	43.01	554	42.95	486	43.82	190	42.04
Age								
18–24	248	17.51	234	18.14	217	19.57	67	14.82
25–29	152	10.73	143	11.09	129	11.63	44	9.73
30–39	358	25.28	331	25.66	292	26.33	134	29.65
40–49	307	21.68	276	21.4	227	20.47	93	20.58
50–59	219	15.47	191	14.81	158	14.25	82	18.14
60–65	76	5.37	69	5.35	51	4.6	19	4.2
66 and above	56	3.95	46	3.57	35	3.16	13	2.87
Education								
No education	30	2.12	29	2.25	24	2.16	6	1.33
O level	100	7.06	87	6.74	73	6.58	24	5.31
A level	128	9.04	115	8.91	97	8.75	28	6.19
Vocational school	219	15.47	200	15.5	165	14.88	66	14.6
Undergraduate	624	44.07	569	44.11	498	44.91	204	45.13
Masters	230	16.24	213	16.51	183	16.5	87	19.25
Ph.D.	85	6	77	5.97	69	6.22	37	8.19
Marital status								
Single	468	33.05	441	34.19	392	35.35	119	26.33
Married	584	41.24	523	40.54	439	39.59	217	48.01
Cohabiting	275	19.42	248	19.22	215	19.39	91	20.13
Divorced	76	5.37	69	5.35	54	4.87	21	4.65
Widowed	13	0.92	9	0.7	9	0.81	4	0.88
Car availability								
No	453	31.99	416	32.25	369	33.27	124	27.43
Yes	963	68.01	874	67.75	740	66.73	328	72.57
Bicycle availability								
No	783	55.3	711	55.12	615	55.46	231	51.11
Yes	633	44.7	579	44.88	494	44.54	221	48.89
Household income								
Below 10k	79	5.58	75	5.81	68	6.13	21	4.65
10–20 k	122	8.62	111	8.6	94	8.48	32	7.08
20–30 k	183	12.92	167	12.95	146	13.17	55	12.17
30–40 k	206	14.55	180	13.95	147	13.26	66	14.6
40–50 k	211	14.9	191	14.81	167	15.06	63	13.94
50–75 k	277	19.56	257	19.92	224	20.2	106	23.45
75–100 k	110	7.77	99	7.67	83	7.48	48	10.62
100–125 k	28	1.98	24	1.86	18	1.62	8	1.77
125–150 k	16	1.13	16	1.24	13	1.17	4	0.88

Table 3 (continued)

Total	Completed BQ 1416		Completed BQ and LCC 1290		Completed BQ, LCC and NG 1109		Completed online survey and used rMove for 2 weeks 452	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Above–150	9	0.64	8	0.62	8	0.72	5	1.11
Do not know	90	6.36	83	6.43	76	6.85	22	4.87
Prefer not to say	85	6	79	6.12	65	5.86	22	4.87
Location								
Leeds	899	63.49	821	63.64	704	63.48	298	65.93
Elsewhere in W Yorkshire	161	11.37	146	11.32	126	11.36	45	9.96
Elsewhere in the UK	356	25.14	323	25.04	279	25.16	109	24.12

completion, except for the levels of car availability, where we see a higher share in the group who reach the final stage of the survey. There is clearly a possibility that this higher rate is a reflection of greater survey engagement with the car ownership question. Overall, the stability in socio-demographics characteristics is reassuring in that we see that there are no major underlying differences in respondents who drop out or stay in the survey for its whole duration.

Approximately 65% of the recruited participants were from the greater Leeds area, reflecting the fact that several local recruitment channels were used, as discussed in "Recruitment strategy" section. Approximately 10% of the sample lived elsewhere in West Yorkshire, while the rest of the sample was from elsewhere in the UK, with clusters in major cities such as London.

Life-course calendar

We now turn to a detailed descriptive analysis of the other parts of the survey data. In each case, we include in the statistics all the people who have completed up to that specific component, i.e. for example in the descriptive statistics of the LCC data, reported in Table 4, we use a sample of 1290. This table shows different life-course changes (listed as column headings) and the mean number of occurrences by age group and gender. We also report the shares the different age and gender groups in the sample. As expected, the highest percentage of respondents still living with their parents belongs to the age group 18–24, followed by the group 25–29. Somewhat surprisingly, about 13 respondents in the age range 60–65 stated that they live with their parents, but this might be the case of people who took elderly parents in their home.

We asked participants to state how many times they relocated while living with their parents. We observe that over 60 s report the highest means (with relatively high standard deviations) as well as people in the 25–29 group, but statistical testing shows that there are no significant differences across age groups in this specific event. As expected, a rising number of long-term partners (defined as partners with whom a person has lived with) is reported with increased age, except for the age group 60–65. The same is true for the number of children and the number of jobs, while the 60-65 age category reports the highest

Table 4 LCC descriptive statistics

	Share	Still living with parents		No of relocations while living with parents		No of long-term partners		No of children		No of homes		No of jobs		No of cars	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age															
18–24	18.14%	45%	0.50	1.72	1.90	0.36	0.64	0.14	1.09	1.60	1.93	1.96	1.88	0.39	0.85
25–29	11.09%	13%	0.34	2.17	2.39	1.00	1.27	0.19	0.60	3.80	2.71	3.45	2.23	1.20	2.27
30–39	25.66%	5%	0.21	1.97	2.43	1.41	1.86	0.91	1.09	4.63	2.89	4.68	3.73	2.02	1.95
40–49	21.40%	1%	0.12	1.81	2.09	1.53	1.01	1.35	1.29	4.92	2.99	4.62	3.45	3.75	3.20
50–59	14.81%	2%	0.12	1.74	1.80	1.50	1.31	1.54	1.19	5.16	3.00	5.37	3.87	5.34	4.55
60–65	5.35%	1%	0.12	2.48	2.41	1.42	0.86	1.49	1.15	4.96	2.92	4.83	3.04	7.64	6.45
66 and above	3.57%	0.00	0.00	2.30	2.26	1.61	0.95	1.63	1.20	5.07	2.90	6.26	5.57	6.41	4.71
Men	42.95%	0.12	0.33	2.01	2.13	1.29	1.76	1.01	1.34	4.12	3.05	4.21	3.89	3.60	4.55
Women	57.05%	0.11	0.31	1.85	2.21	1.17	0.97	0.88	1.16	4.20	3.02	4.19	3.25	2.46	2.96
All	–	0.12	0.32	1.92	2.18	1.22	1.37	0.94	1.24	4.16	3.03	4.20	3.54	2.95	3.77

Table 5 NG descriptive statistics

Ego	Alter		
	Female	Male	Total
Ego			
Female	7.21	3.51	10.72
	67.24%	*32.76%*	–
Male	3.92	5.37	9.28
	42.19%	*57.81%*	–

number of cars owned during the life course. Statistical testing shows that the differences across age groups are significant, as expected.

There are no significant differences in the distribution of life-course events across men and women, except in the case of car ownership, where women own significantly fewer cars than men over their life.

As in the case of the socio-demographics characteristics, we have prepared two additional versions of Table 4: one including the participants who have completed the entire online survey and one only included those who have in addition used the rMove for 2 weeks[3]. We also tested whether people who have completed different stages of the survey are different in terms of average number of the different life-course events. Our results show that, for all events except number of people still living with parents and number of relocations while living with parents, there are significant differences at the 0.05 level. We believe this depends on survey involvement, as the events where the differences are present are those that involved inputting more information in the LCC, a relatively demanding and time consuming task.

Name generator

Tables 5 and 6 report descriptive statistics for the NG and name interpreter data. In particular, Table 5 shows, depending on the gender of the *ego* (i.e. the participant whose social network is analysed), the number and share (in italics) of *alters* (i.e. social contacts) of each gender were listed in the NG. As expected, we observe that people are more likely to report alters of their own gender. This gender homophily was confirmed to be significant by statistical testing. The rightmost coloumn of the table represents the total number of contacts reported, showing that female egos reports slightly more contacts than male. The same type of information is displayed in Table 6 for the age of the egos and alters. We can see that there is a high age homophily, meaning that people report a high number of social contacts of their same age group. This finding was also confirmed by statistical tests of significance.

As in the previous cases, we compared these two tables with the corresponding ones including only those respondents who not only completed the online survey, but also used the smartphone app for 2 weeks (see footnote 3). We found significant differences in the number of contacts by gender and by all age categories except 40–49, 60–69 and over 70.

---

[3] Due to space limitations, we report those tables in an online appendix at http://www.stephanehess.me.uk/papers/Calastri_Crastes_Hess_2017_online_appendix.pdf.

Table 6 NG descriptive statistics

Age Alter	Under 18	18–24	25–29	30–39	40–49	50–59	60–69	70 and above	Total
Ego									
18–24	0.49	6.06	1.04	0.36	0.70	0.97	0.23	0.24	10.09
	4.9%	*60.0%*	*10.3%*	*3.6%*	*6.9%*	*9.6%*	*2.3%*	*2.3%*	*–*
25 - 29	0.24	1.26	4.34	1.99	0.46	1.12	0.51	0.26	10.19
	2.4%	*12.3%*	*42.6%*	*19.6%*	*4.5%*	*11.0%*	*5.0%*	*2.6%*	*–*
30–39	0.28	0.24	0.89	4.84	1.38	0.75	1.18	0.28	9.84
	2.8%	*2.4%*	*9.1%*	*49.1%*	*14.1%*	*7.6%*	*12.0%*	*2.9%*	*–*
40– 49	0.44	0.26	0.30	1.67	3.70	1.32	0.77	0.91	9.36
	4.8%	*2.8%*	*3.2%*	*17.8%*	*39.5%*	*14.1%*	*8.2%*	*9.7%*	*–*
50–59	0.35	0.45	0.56	0.93	1.99	4.24	1.24	1.06	10.83
	3.2%	*4.1%*	*5.2%*	*8.6%*	*18.4%*	*39.2%*	*11.5%*	*9.8%*	*–*
60–65	0.20	0.22	0.43	1.53	1.33	2.41	3.27	1.20	10.59
	1.9%	*2.0%*	*4.1%*	*14.4%*	*12.6%*	*22.8%*	*30.9%*	*11.3%*	*–*
66 and above	0.46	0.46	0.69	1.06	1.49	1.89	4.09	2.46	12.57
	3.6%	*3.6%*	*5.5%*	*8.4%*	*11.8%*	*15.0%*	*32.5%*	*19.5%*	*–*

Recall survey

As part of our ongoing work, we plan to survey our respondents again at different points in time to observe changes that may have occurred since the main data collection took place. At this stage, for illustration purposes, we have performed a pilot recall survey aimed at understanding changes in social network over time and limitations in the use of the NG to address such survey question. Studies such as Calastri et al. (2017c) and Sharmeen et al. (2016) have used network data collected at different points in time using name generators, to study the stability and changes in social networks. Especially when the period of time between the two surveys is extended, recall issues may emerge and it might be difficult to distinguish people who are no longer part of the network from those who have been forgotten. On the other side, showing the initial NG at the moment of the second survey and asking people to confirm that the network is unchanged may lead to bias.

For this reason, we adopted a different approach. In July 2017 we recontacted 10 people who had been surveyed in November 2016 and asked them to complete the NG from scratch. Once completed, we showed them the NG they filled out in November, and asked them to match the names of the people who were in both lists. If there was a perfect correspondence, the survey ended. This did not happen for any of the respondents. For the contacts who had only been reported in November but not in July, we asked if they were still in touch. If not, we asked for the reason why. For those people who were only reported in the July NG, we asked whether they were new social contacts, or if they were already known to the participants in November, but had been forgotten in the NG at the time.

The average network size of these 10 people was 12.9 in November and 11.2 in July (not statistically different) and they on average matched 8.4 contacts. On average, there were 4.2 contacts who had not been named in July but with whom they were still in touch and 0.2 who were not named and with whom they were actually not in touch any more. While the former were spread across the sample, the latter were 2 contacts of a single participant, lost

Table 7 Share of rMove trips by purpose

Trip purpose	%
Going home	28.47
Going to work	19.42
Social	11.98
Going shopping	10.23
Other[a]	10.19
Drop off/Pick up	6.62
Errands	4.56
Sports and exercise	4.52
Leisure	2.31
Going to School/ class/ University	1.15
Petrol	0.57

[a]Includes "change travel mode" and "Other (non specified) purpose"

because he/she went to university. In addition, there were on average 2.6 people who were only named in the July list but were actually known in November already, spread across the sample, while only 0.1 (1 alter from one ego) was a new contact. The present sample is of course very small, and a larger recall study is needed for any significant conclusion. Nevertheless, we find the results quite striking, as they highlight serious limitations of the NG tool for eliciting social networks, suggesting it suffers from severe recall issues.

Smartphone travel and activity app

As shown in Table 3, 452 respondents used the smartphone app rMove for two weeks. A total of 40, 672 trips were recorded, for which participants had to answer different questions, as explained in "rMove " section. 84.71% of trips were recorded correctly, and for 81.1% all details were provided by respondents. 12.34% of total trips were marked by users as GPS errors and 2.95% as other errors (e.g. the app recorded a movement when the person did not actually move). This implies an average number of daily trips of approximately 5.45, which is substantially larger than the figure reported by the National Household Travel Survey of 3.79 (U.S. Department of Transportation 2009). This difference comes as no surprise after multiple data collections taking place in different geographical locations have found trip under-reporting when data is collected via traditional diary methods (Bricka and Bhat 2006; Stopher et al. 2007; Wolf 2006).

Table 7 shows the share of rMove trips by purpose[4], where the most common destination is home or work, followed by social (all non-work activities that involve other people, such as going out and spending time with the family) and shopping (including both groceries and other shopping). Thanks to the link with the NG, we were able to determine that 55.63% of the trips were made by participants on their own, 15.62% were conducted only with people listed in the NG, 17.03% with both contacts from the NG and others, and the remaining 11.73% of trips was conducted with contacts unreported in the NG only. As discussed in "Recall survey" section above, these contacts could be people who are part of the

---

[4] The purposes listed in the tables are macro-categories to give an overview of the trips, but a more detailed distinction is made in the data.

	Share of trips (%)	Average distance per trip (km)
Walk	46.71	1.55
Car	37.57	11.88
Urban bus	6.95	10.49
Rail	2.40	54.33
Bicycle	2.27	7.85
Taxi	1.76	6.26
Urban rail	1.23	18.99
Other bus	0.65	42.13
Other modes	0.46	186.40

Table 8 Share of trips and average distance by mode

network but respondents have forgotten to name them, or someone they do not normally spend time with, and who therefore have not been reported in the NG.

Table 8 reports the share of trips by each mode, showing that most trips are performed by car and on foot. While the share is much lower, urban buses are used more than other public transport modes. This is to be expected in Leeds, as there is no metro/light rail service in the city. Rail trips are mainly non-urban, as the distance also shows. Modes other than those listed were grouped together, where the occasional presence of trips by plane pushes the mean distance (186.40 km) upwards.

Summary

Recent work in transport research has increasingly tried to broaden out beyond traditional areas such as mode choice or car ownership and has tried to position travel decisions within the broader life context. However, while important progress has been made in terms of how to capture these additional dimensions, both in terms of detailed tracking of movements and in-depth data collection on long term decisions or social network influences, surveys have tended to look at only a handful (or often one) of these issues in isolation. The ERC funded DECISIONS project set out to instead collect data in a unified approach by jointly looking at numerous dimensions. We believe that this is an unprecedented effort that joins complexity of the survey design, amount of information collected and sample size. The present paper has given an overview of the different components of this survey and provided initial insights into the resulting data. The full dataset will be made publicly available at the end of the DECISIONS project in 2020.

The survey was made up of two different components, an online survey and a smartphone tracking app. The online survey was in turn divided into three elements, a background questionnaire, a life-course calendar and a name generator & name interpreter. These different tools were used to capture different aspects of participants' lives, including travel and energy choices, life-course events, social networks and short term activity and mobility behaviour. The level of complexity goes beyond what is typically expected of survey respondents, and the rate of retention across different survey components is consequently lower than in many other studies. Here, it is worth noting that we not only faced the challenge of retaining respondents throughout a long and demanding survey, but that our

approach, by linking different parts of the survey (e.g. the social network and the smartphone application) to capture the complexity of real-life behaviour, asked respondents to disclose a vast amount of data about their life and habits. Nevertheless, we end up with a usable sample of respondents who completed all parts of the survey, and initial analysis of the data shows high quality information and crucially allows us to understand the links between short-term and long-term decisions for the same individual while also making links with his/her social network.

We believe that the experience gained by designing and implementing the present survey allows us to formulate a few recommendations that will hopefully be useful to other researchers who will attempt to collect similar datasets in the future. These are listed below:

- When presenting a long survey, it might be useful to build links between the different parts, and generate contents on the basis of the answers to previous questions. This gives a sense of continuity and fosters involvement.
- Using multiple recruitment approaches is advisable, as it helps minimise the risk of low retention rates. In particular, using sources that respondents trust or developing methods to gain respondents' trust (providing plenty of information, direct lines of support) is a recommended approach, especially when many questions about respondents' lives are presented.
- Monetary incentives are preferred to other forms of incentives (such as prize raffles). The incentives need to be adequate, but as suggested by others in the literature, do not need to pay for participation time, and the primary motivation should be communicated in terms of contribution to research/policy.
- When using a travel and activity app, extensive testing is recommended to avoid respondents facing malfunctioning, battery drainage and other issues that might lead them to abandon the survey. Moreover, a good level of customisation provides a more enjoyable experience and can help retention.

Author contributions CC: Literature search, protocol development and manuscript writing. R Crastes Dit Sourd: Coordination of data collection and contribution to protocol development. SHess: Contribution to the protocol development and manuscript editing.

References

Anderson, B., Silver, B.: The validity of survey responses: insights from interviews of married couples in a survey of soviet emigrants. Social Forces **66**(2), 537–554 (1987)

Arentze, T.A., Timmermans, H.J.: Representing mental maps and cognitive learning in micro-simulation models of activity-travel choice dynamics. Transportation **32**(4), 321–340 (2005)

Arentze, T.A., Timmermans, H.J.: A need-based model of multi-day, multi-person activity generation. Transp. Res. Part B Methodol. **43**(2), 251–265 (2009)

Axhausen, K.W.: Social networks and travel: some hypotheses. Soc. Dimens. Sustain. Transp. Transatl. Perspect. pp. 90–108 (2005)

Axhausen, K.W., Schmid, B., Weis, C.: Predicting response rates updated, Arbeitsberichte Verkehrs-und Raumplanung, p. 1063 (2015)

Beige, S., Axhausen, K.W.: Interdependencies between turning points in life and long-term mobility decisions. Transportation **39**(4), 857–872 (2012)

Bhat, C.R., Singh, S.K.: A comprehensive daily activity-travel generation model system for workers. Transp. Res. Part A Policy Pract. **34**(1), 1–22 (2000)

Bhat, C.R., Srinivasan, S., Bricka, S.: Conversion of volunteer-collected gps diary data into travel time performance measures: literature review, data requirements, and data acquisition efforts. Tech. rep, Texas Department of Transportation (2005)

Bohte, W., Maat, K.: Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: a large-scale application in the netherlands. Transp. Res. Part C Emerg. Technol. **17**(3), 285–297 (2009)

Bricka, S., Bhat, C.: Comparative analysis of global positioning system-based and travel survey-based data. Transp. Res. Rec. J. Transp. Res. Board **1972**, 9–20 (2006)

Browne, K.: Snowball sampling: using social networks to research non-heterosexual women. Int. J. Soc. Res. Methodol. **8**(1), 47–60 (2005)

Burt, R.S.: Network items and the general social survey. Soc. Netw. **6**(4), 293–339 (1984)

Calastri, C., Hess, S., Choudhury, C., Daly, A., Gabrielli, L.: Mode choice with latent availability and consideration: theory and a case study. Transp. Res. Part B (2017a)

Calastri, C., Hess, S., Daly, A., Carrasco, J.A.: Does the social context help with understanding and predicting the choice of activity type and duration? An application of the multiple discrete-continuous nested extreme value model to activity diary data. Transp. Res. Part A Policy Pract. **104**, 1–20 (2017b)

Calastri, C., Hess, S., Daly, A., Carrasco, J.A., Choudhury, C.: Modelling relationship loss and retention in social networks: the role of tie strength and intra-respondent heterogeneity. Paper presented at the Fifth International Choice Modelling Conference, Cape Town (2017c)

Calastri, C., Hess, S., Daly, A., Maness, M., Kowald, M., Axhausen, K.: Modelling contact mode and frequency of interactions with social network members using the multiple discrete-continuous extreme value model. Transp. Res. Part C Emerg. Technol. **76**, 16–34 (2017d)

Calastri, C., Hess, S., Pinjari, A.R., Daly, A.: Accommodating correlation across days in multiple-discrete continuous models for activity scheduling: estimation and forecasting considerations. In: Paper presented at the Fifth International Choice Modelling Conference, Cape Town (2017e)

Campbell, K.E., Lee, B.A.: Name generators in surveys of personal networks. Soc. Netw. **13**(3), 203–221 (1991)

Carrasco, J., Miller, E., Wellman, B.: How far and with whom do people socialize?: Empirical evidence about distance between social network members. Transp. Res. Rec. J. Transp. Res. Board **2076**, 114–122 (2008a)

Carrasco, J.A., Bustos, C., Cid-Aguayo, B.: Affective personal networks versus daily contacts: analyzing different name generators in a social activity-travel behavior context. In: Transport survey methods: best practice for decision making. Emerald Group Publishing Limited, pp. 409–426 (2013)

Carrasco, J.A., Hogan, B., Wellman, B., Miller, E.J.: Agency in social activity interactions: the role of social networks in time and space. Tijdschrift voor economische en sociale geografie **99**(5), 562–583 (2008b)

Carrasco, J.A., Hogan, B., Wellman, B., Miller, E.J.: Collecting social network data to study social activity-travel behavior: an egocentric approach. Environ. Plan. B Plan. Des. **35**(6), 961 (2008c)

Caspi, A., Moffitt, T.E., Thornton, A., Freedman, D., Amell, J.W., Harrington, H., Smeijers, J., Silva, P.A.: The life history calendar: a research and clinical assessment method for collecting retrospective event-history data. Int. J. Methods Psychiatr. Res. (1996)

Chorus, C.G., Kroesen, M.: On the (im-) possibility of deriving transport policy implications from hybrid choice models. Transp. Policy **36**, 217–222 (2014)

Choudhury, C.F., Ben-Akiva, M., Abou-Zeid, M.: Dynamic latent plan models. J. Choice Model. **3**(2), 50–70 (2010)

Daly, A., Hess, S., Patruni, B., Potoglou, D., Rohr, C.: Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail travel behaviour. Transportation **39**(2), 267–297 (2012)

Döring, L., Albrecht, J., Scheiner, J., Holz-Rau, C.: Mobility biographies in three generations—socialization effects on commute mode choice. Transportation Research Procedia 1 (1), 165 – 176, planning for the future of transport: challenges, methods, analysis and impacts—41st European Transport Conference Selected Proceedings (2014)

Dugundji, E., Walker, J.: Discrete choice with social and spatial network interdependencies: an empirical example using mixed generalized extreme value models with field and panel effects. Transp. Res. Rec. J. Transp. Res. Board **1921**, 70–78 (2005)

Duncan, S., Edwards, R.: Lone Mothers, Paid Work and Gendered Moral Rationalitie. Springer, Berlin (1999)

Freedman, D., Thornton, A., Camburn, D., Alwin, D., Young-DeMarco, L.: The life history calendar: a technique for collecting retrospective data. Sociol. Methodol, pp. 37–68 (1988)

Freeman, L.C., Thompson, C.R.: Estimating acquaintanceship volume. The small world, 147–158 (1989)

Frei, A., Axhausen, K.W.: Size and structure of social network geographies (2007)

Frei, A., Ohnmacht, T., 2016. Egocentric networks in zurich: Quantitative survey development, data collection and analysis. In: Kowald, M., Axhausen, K.W. (eds.) Social Networks and Travel Behaviour, 1st Edition. Ashgate Publishing, Ltd., Ch. 3, pp. 51–98

Furstenberg, F.F., Brooks-Gunn, J., Morgan, S.P.: Adolescent Mothers in Later Life. Human Development in Cultural and Historical Contexts. Cambridge University Press, Cambridge (1987)

Greaves, S.P., Ellison, A.B., Ellison, R.B., Rance, D., Standen, C., Rissel, C., Crane, M.: A web-based diary and companion smartphone app for travel/activity surveys. In: 10th International Conference on Transport Survey Methods (2014)

Greene, E., Flake, L., Hathaway, K., Geilich, M.: A seven-day smartphone-based gps household travel survey in indiana. In: Transportation Research Board 95th Annual Meeting (16-6274) (2016)

Jara-Díaz, S., Rosales-Salas, J.: Understanding time use: Daily or weekly data? Transp. Res. Part A Policy Pract. **76**, 38–57 (2015)

Jariyasunant, J., Abou-Zeid, M., Carrel, A., Ekambaram, V., Gaker, D., Sengupta, R., Walker, J.L.: Quantified traveler: Travel feedback meets the cloud to change behavior. J. Intell. Transp. Syst. ahead-of-print, 1–16 (2014)

Kelly, P., Krenn, P., Titze, S., Stopher, P., Foster, C.: Quantifying the difference between self-reported and global positioning systems-measured journey durations: a systematic review. Transp. Rev. **33**(4), 443–459 (2013)

Killworth, P.D., Johnsen, E.C., Bernard, H., Shelley, G.A., McCarty, C.: Estimating the size of personal networks. Soc. Netw. **12**(4), 289–312 (1990)

Kowald, M., Axhausen, K.W.: Surveying data on connected personal networks. Travel Behav. Soc. **1**(2), 57–68 (2014)

Kowald, M., Frei, A., Hackney, J.K., Illenberger, J., Axhausen, K.W.: Collecting data on leisure travel: the link between leisure contacts and social interactions. Procedia Soc. Behav. Sci. **4**, 38–48 (2010)

Lin, T., Wang, D.: Social networks and joint/solo activity-travel behavior. Transp. Res. Part A Policy Pract **68**, 18–31 (2014)

Local Level UK: http://www.local-level.org.uk/uploads/8/2/1/0/8210988/info_briefing_1_smartphone_ownership.pdf (2014)

Maness, M., Cirillo, C.: An indirect latent informational conformity social influence choice model: formulation and case study. Transp. Res. Part B Methodol. **93**, 75–101 (2016)

Marsden, P.V.: Network data and measurement. Annu. Rev. Sociol. **16**(1), 435–463 (1990)

Minnen, J., Glorieux, I., van Tienoven, T.P.: Transportation habits: evidence from time diary data. Transp. Res. Part A Policy Pract. **76**, 25–37 (2015)

Molin, E., Mokhtarian, P., Kroesen, M.: Multimodal travel groups and attitudes: a latent class cluster analysis of dutch travelers. Transp. Res. Part A Policy Pract. **83**, 14–29 (2016)

Montini, L., Prost, S., Schrammel, J., Rieser-Schüssler, N., Axhausen, K.W.: Comparison of travel diaries generated from smartphone data and dedicated gps devices. In: 10th International Conference on Survey Methods in Transport. Leura, Australia (2014)

Paleti, R., Bhat, C., Pendyala, R.: Integrated model of residential location, work location, vehicle ownership, and commute tour characteristics. Transp. Res. Rec. J. Transp. Res. Board **2382**, 162–172 (2013)

Patrick, J.: A Glasgow Gang Observed. Eyre Methuen, London (1973)

Pike, S.: Travel mode choice and social and spatial reference groups: comparison of two formulations. Transp. Res. Rec. J. Transp. Res. Board **2412**, 75–81 (2014)

Resource Systems Group, R.: rmove. http://rmove.rsginc.com/index.html (2017)

Schoenduwe, R., Mueller, M.G., Peters, A., Lanzendorf, M.: Analysing mobility biographies with the life course calendar: a retrospective survey methodology for longitudinal data collection. J. Transp. Geogr. **42**, 98–109 (2015)

Sharmeen, F., Chávez, Ó., Carrasco, J.-A., Arentze, T., Tudela, A.: Modeling population-wide personal network dynamics using two-wave data collection method and origin-destination survey. In: Transportation Research Board 95th Annual Meeting. No. 16-3387 (2016)

Silvis, Julia, D.N., D'Souza, R.: Social networks and travel behavior: report from an integrated travel diary. In: 11th International Conference on Travel Behaviour Research, Kyoto (2006)

Stopher, P., FitzGerald, C., Xu, M.: Assessing the accuracy of the sydney household travel survey with gps. Transportation **34**(6), 723–741 (2007)

U.S. Department of Transportation: Summary of travel trends. national household travel survey. Tech. rep., Federal Highway Administration (2009)

van den Berg, P., Arentze, T., Timmermans, H.: A path analysis of social networks, telecommunication and social activity-travel patterns. Transp. Res. Part C Emerg. Technol. **26**, 256–268 (2013)

van den Berg, P., Arentze, T.A., Timmermans, H.J.: Size and composition of ego-centered social networks and their effect on geographic distance and contact frequency. Transp. Res. Rec. J. Transp. Res. Board **2135**(1), 1–9 (2009)

Vij, A., Walker, J.L.: How, when and why integrated choice and latent variable models are latently useful. Transp. Res. Part B Methodol. **90**, 192–217 (2016)

Wolf, J.: Applications of new technologies in travel surveys. In: Travel Survey Methods: Quality and Future Directions. Emerald Group Publishing Limited, pp. 531–544 (2006)

Zmud, J., Wolf, J.: Identifying the correlates of trip misreporting-results from the california statewide household travel survey gps study. In: 10th International Conference on Travel Behaviour Research. pp. 10–15 (2003)

Chiara Calastri   is a Research Fellow at the Institute for Transport Studies and Choice Modelling Centre at the University of Leeds. Her work focuses on applying advanced choice models to different types of complex revealed preference data. She is mainly working with discrete-continuous models and on decisions related to time use and social interactions.

Romain Crastes dit Sourd   is a Research Fellow in the Institute for Transport Studies and Choice Modelling Centre at the University of Leeds. His main research interest are modelling rare and long-term choices and operationalising choice models for large Revealed Preferences data sets. He is originally an environmental economist and has also addressed issues related to the economic valuation of non-market goods.

Stephane Hess   is Professor of Choice Modelling in the Institute for Transport Studies and Director of the Choice Modelling Centre at the University of Leeds. He is also Honorary Professor in Choice Modelling in the Institute for Transport and Logistics Studies at the University of Sydney and has a director position at RSG, a leading North American consultancy company. His area of work is the analysis of human decision using advanced discrete choice models, and he is active in the fields of transport, health and environmental economics. Hess has made contributions to the state of the art in the specification, estimation and interpretation of such models. He is also the founding editor in chief of the Journal of Choice Modelling and the founder and steering committee chair of the International Choice Modelling Conference.