WORKING PAPER

ITLS-WP-23-23

Who stays and who plays? Participant retention and smartphone app usage in a longitudinal travel survey.

By
Stephen P. Greaves[a], Alec Cobbold[b], Oli Stanesby[c], Mel Sharman[c], Kim Jose[c], Jack Evans[c], Verity Cleland[c]

[a] Institute of Transport and Logistics Studies (ITLS),
[b] The University of Sydney, Australia
[c] School of Public Health, University of Sydney
Menzies Institute for Medical Research, UTAS

December 2023

ISSN 1832-570X

INSTITUTE of TRANSPORT and LOGISTICS STUDIES

The Australian Key Centre in
Transport and Logistics Management

The University of Sydney

Established under the Australian Research Council's Key Centre Program.

NUMBER:	Working Paper ITLS-WP-23-23
TITLE:	**Who stays and who plays? Participant retention and smartphone app usage in a longitudinal travel survey.**

ABSTRACT:

Longitudinal studies have become increasingly popular for investigating changes in behaviour, but present additional challenges around participant recruitment, retention, compliance, and ultimately data quality. Personal technologies, particularly smartphones, have become integral to tackling these challenges but come with their own caveats around user acceptance and compliance. The current paper investigates these issues in the context of a longitudinal investigation of interventions designed to encourage use of public transport and increase associated physical activity in Tasmania, Australia. The study comprises multiple waves of data collection over a seven-month period in which travel data were collected using a smartphone app and supplemented with user experience surveys. Evidently attrition is lower for older participants, those engaging with the app more, and those responding to the research/environmental/health messaging of the survey as well as the potential for financial gain. App usage is lower among older participants while app engagement is stronger for males, those recording less travel and those indicating environmental reasons as a motivator for completing the study. Experiences with the app were mixed, participants reported positive sentiments about the ease of use, hedonic motivation, and help in recalling travel; however, concerns were raised over the accuracy of trip recording, the associated burden of correcting trips, and reductions in smartphone battery-life. Despite the unplanned coincidence with the COVID-19 restrictions, outcomes provide important guidance around recruitment, retention and post-hoc analysis of results from longitudinal studies.

KEY WORDS: ***Longitudinal travel surveys; smartphone apps; participant burden;***

AUTHORS: **Greaves, Cobbold, Stanesby, Sharman, Jose, Evans, Cleland**

ACKNOWLEDGEMENTS: The analysis conducted in this paper uses data collected as part of NHMRC-APP1152999, Health by Stealth: Developing strategies to increase active and public transport.

CONTACT:

INSTITUTE OF TRANSPORT AND LOGISTICS STUDIES (H04)

The Australian Key Centre in Transport and Logistics Management

The University of Sydney NSW 2006 Australia

Telephone: +612 9114 1813

E-mail: business.itlsinfo@sydney.edu.au

Internet: http://sydney.edu.au/business/itls

DATE: December 2023

1. Introduction/Background

Longitudinal panel or cohort studies in which the same individual or household are surveyed more than once, have intrinsic appeal for studying behaviour over an extended period and/or following an intervention or policy change (Stopher and Greaves, 2010; Greaves et al. 2015). They provide rich insights into the dynamics of changes in travel, health and other outcomes at the individual level, accounting for changes in both personal circumstances and the broader environment. However, the challenges of participant recruitment, retention, compliance and ultimately data quality are magnified the longer the data collection period. Study results may be compromised or misleading if participants completing the survey exhibit different characteristics from those dropping out, resulting in attrition bias. Equally problematic is variation in individual-level data quality over the study duration, compromising conclusions around intra-personal change.

Personal technologies have become increasingly integral to tackling some of these challenges, particularly through developments in smartphones and associated apps. The locational and interactive potential of smartphones combined with their ubiquity and pragmatism as something people are more likely to keep charged and with them, is particularly appealing for travel behaviour research. However, in addition to a lack of concordance on how travel survey apps should be designed, challenges remain related to the smartphone technology itself (battery life, dealing with the myriad of makes, models and operating systems, quality of Global Positioning System [GPS] receiver), tightening restrictions around default background tracking and personal privacy[1], and user willingness to 'play ball' in terms of downloading and engaging with an app designed to track personal mobility with greater precision.

While attrition in panel surveys (e.g., Greaves et al., 2015) and factors impacting the use of apps in surveys (e.g., Assemi et al., 2018) are reasonably understood, attrition and the level of app engagement when participants are asked to use an app repeatedly in a longitudinal study remains unclear. The current paper investigates the primary factors related to participant retention and engagement with a customised smartphone app designed to collect travel information (trips, mode and purpose) at four timepoints of a longitudinal evaluation of travel and physical activity in Tasmania, Australia (Sharman et al., 2020). User experience surveys provide additional information around participation, with specific questions designed to capture perceptions of general usability, usefulness, and concerns about the app.

2. Literature Review

2.1: Attrition in longitudinal panel studies

Attrition in longitudinal panel studies occurs when participants become ineligible for subsequent waves of data collection either by choice or circumstance, such as moving away. Attrition results in two potential problems: a loss of sample and subsequent lack of statistical power to detect change, and an increasingly biased sample as attrition tends to be greater for some segments of the sample than others (Ellison et al., 2017). In terms of sample loss, evidence from the transport literature suggests attrition rates between waves are generally around 25-30%, but this varies substantially dependent on the complexity of the survey (Ortúzar et al., 2011). Similar levels of attrition (32% per annum) are reported by Stopher et al. (2013) from a five-year annual GPS-based panel study of travel behaviour change initiatives in Australia.

[1] https://www.theage.com.au/technology/new-apple-software-brings-slight-solid-improvements-across-devices-20190607-p51veo.html. Accessed June 12th, 2019.

In a multi-year survey more closely aligned with that under investigation here in terms of use of GPS/smartphone technology, (Ellison et al., 2017) report that 43% of the original sample completed all three waves. That study also found the sample to become increasingly unrepresentative, with younger participants and males dropping out at greater rates. Similar issues around the loss of young and male participants are reported in longitudinal health intervention investigations (Van der Mispel et al., 2017) with lower education and lower socio-economic status also factors impacting drop-out (Otahal et al., 2021, Nguyen et al., 2023). In addition to socio-demographics, travel behaviour itself may impact attrition rates. Those making more trips might be expected to be less likely to return and indeed less likely to complete all the survey tasks given the additional burden (Ellison et al., 2017). However, the reverse has also been reported with the highest attrition rates reported for those with the lowest trip rates (La Paix Puello et al., 2017). The study purpose also has an impact on attrition. In their study of social networks during the COVID-19 pandemic, Nguyen et al. (2023) reported a relatively low attrition rate of 15% over a 12-month period, which may in part have been due to a surge in community interest in social cohesion during the COVID-19 lockdowns.

Various strategies have been used to retain participants in longitudinal studies including incentives, reminders, and re-sending surveys (Booker et al., 2011). While incentives (e.g., cash or gifts) are thought to be effective, there are questions about the optimal types, timing, and amount (Scheepers and Hoogendoorn-Lanser, 2018). Further, while incentives may increase retention rates, they may do so unevenly across demographic characteristics, increasing the potential for attrition bias (Singer and Ye, 2013). Other methods of participant retention (reminders and re-sending surveys) may be less effective when dealing with more contemporary methods of data collection in longitudinal studies, such as web surveys, smartphone apps, and wearable technologies. Instead, it is suggested that reducing participant burden through flexible data collection methods (i.e., barrier-reduction) is the most effective strategy of participant retention in longitudinal studies (Teague et al., 2018). However, there has been some success with offering alternative incentives, such as rideshare tokens (Leavens et al., 2019). It is likely the most successful retention strategies are those that employ several methods together.

2.2 Smartphones and longitudinal panel studies

Since the turn of the millennium, technology has become integral to longitudinal studies of travel and health. Advancements in GPS and smartphone technology have facilitated the automated tracking/recording of people's movements, which can be used to infer travel and prompt users for additional information about their travel and/or health (Shen and Stopher, 2014, Prelipcean et al., 2018). As of the early 2020s, applications generally remain in trial phase or niche applications developed for specific projects (Gadziński, 2018), with some notable large-scale applications (Zhao et al., 2015, Thomas et al., 2018, Hong et al., 2021, Faghih Imani et al., 2020). Delving into the reasons for this, on the one hand, there is inertia in changing conventional data collection methods, while on the other, there is resistance from potential participants to using their own smartphones for non-personal use. For instance, evidence suggests only half of people are willing to download a travel survey app, citing reasons such as a privacy and battery drain (Verzosa et al., 2018). Compounding this are continuing technical challenges, particularly over the reliable identification of stops (Zhao et al., 2015), mode and trip purpose and the extent to which participant interaction is required to confirm or provide trip information. Advances in data processing functions may be able to alleviate some of this burden using machine learning or other inference algorithms. However, as smartphone apps have been rolled out to reduce the burden on participants of recording travel diaries by traditional methods, this issue presents a continued barrier to their wider uptake (Harding et al., 2021).

While the proliferation of smartphones and associated apps for all manner of purposes has uncovered much about what engages users to download and use them, there is a relative paucity of knowledge in the travel/health survey context. Evidently, app-based surveys tend to appeal, initially at least, to users who are younger and more tech-savvy with associated learnings (potentially) limited to those who already carry a more positive view (Assemi et al., 2018, Eisenmann et al., 2019). In terms of gender, the evidence is conflicting, suggesting this may be down to the specific instrument being tested (Verzosa et al., 2021). It has also been suggested that intrinsic motivations (interest, enjoyment, etc.) are more indicative than extrinsic motivations (incentives, obligations, etc.) of willingness to participate in smartphone travel surveys (Assemi et al., 2018, Verzosa et al., 2021, Bürbaumera et al., 2022). In their evaluation of the *ATLAS II* travel app in Australia, Assemi et al. (2018) revealed facets of a smartphone app's design, namely 'ease of use' and 'usefulness', that were associated with user satisfaction, which in turn had associations with participants' intentions to continue with the survey. There are also suggestions that gamification of travel data collection may appeal to different users, with younger participants more likely to engage with gamified rewards over simple monetary incentives (Verzosa et al., 2018). This highlights the need to investigate how self-reported experiences with the app might impact acceptability and ultimately the quality of data collected.

When it comes to longitudinal travel/health surveys, as with their GPS precursors, smartphone capabilites have the potential to collect more accurate data over multiple days and years. However, there is still a requirement for the participant to download and engage with the app and the survey itself, which may discourage ongoing participation if this is percieved as too onerous. To date, while we have a reasonable body of evidence around attrition in longitudinal surveys and the relative usefulness of strategies to try to addess this, we have little knowledge of attrition associated with re-use of smartphone apps in this context.

3. Materials & Methods

3.1 The trips4health study

This analysis draws from a longitudinal cohort study of adult infrequent bus users (<=2 bus trips/week on average) in Greater Hobart, Tasmania, Australia, known as *trips4health* (Sharman et al., 2020, Jose et al., 2022). The planned study comprised four distinct phases: a baseline assessment of travel and health information (T1), a 16-week randomised controlled trial (RCT) of financial interventions to increase bus use, a post-intervention assessment (T2) and a final assessment a further 3 months later (T3). Each assessment stage (T1, T2 and T3) involved several components including the completion of surveys, wearing of an accelerometer, and downloading and using a smartphone travel app for one week. To facilitate recruitment and retention, all participants were to receive compensation of $AU5 for completing the T1 assessment, $AU10 for the T2 assessment and $AU15 for the T3 assessment. The aim was to recruit 350 participants, which (allowing for attrition) was anticipated to realise 300 completing all phases of the study allowing sufficient power for statistical analyses.

3.2 The trips4health smartphone app

The *trips4health* travel app was developed by a third-party vendor, Ionata Digital, based on their *TourismTracer* app (Hardy et al., 2019). Following download and registration, the app prompted users for home, work and study (if relevant) address information. The app then operated passively in the background, tracking travel and inferring trip-ends, with trips available to classify 10 minutes after completion. This invoked the trip classification screen, which required users to provide three pieces of information; the reason/purpose of travel, the mode/method of travel and the destination (Figure 1). Participants were able to delete spurious trips, split trips and manually add missing trips. Users who had not classified all trips by 6.30pm in the evening received a push notification on their phone prompting them to classify the day's trips.

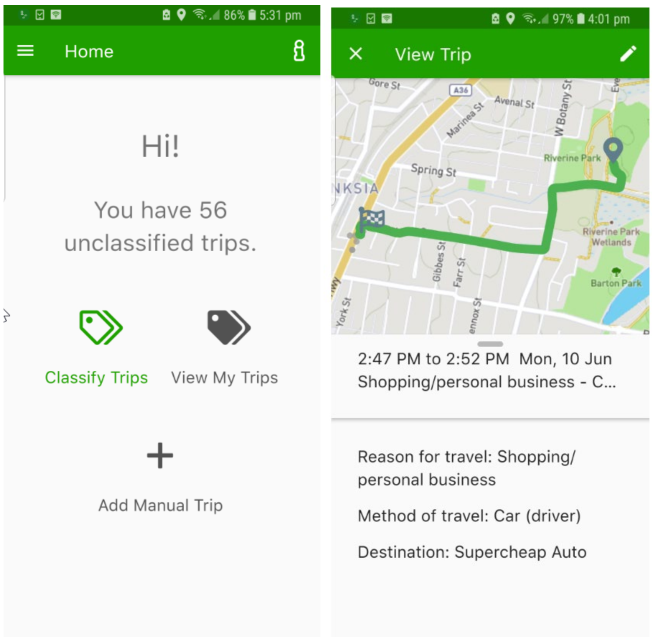

Figure 1: Trips4Health App Summary and Classification Screens

3.3 App experience measures

As part of the study process evaluation, participants completed an app experience survey at T2 (Appendix 1). The experience survey was designed to capture dimensions of app acceptability informed by theoretical constructs for technology acceptance first proposed by (Davis, 1989) and later updated by Venkatesh et al. (2012) with their unified theory of acceptance and use technology (UTAUT2) framework. Further, acknowledging findings of more recent studies around smartphone acceptability, issues pertaining to privacy and risk were incorporated into the survey (Assemi et al., 2018). Participants were asked to report to what extent they agreed (from seven options: 1 strongly disagree, 2 disagree, 3 slightly disagree, 4 neutral, 5 slightly agree, 6 agree, 7 strongly agree) with a series of statements related to the following dimensions of acceptability:

Performance Expectancy – the extent to which participants perceive benefits from using the app, such as reminding them of trips, together with quicker and more accurate recording of travel.

Perceived Ease of Use/Effort Expectancy – the effort required to use the app, encompassing the ease of installation, how easy it was to use and the extent to which it fitted with their lifestyle.

Hedonic Motivation - the fun/pleasure associated with using the app, which the digitisation of survey instruments has the potential to improve, particularly for surveys conducted over several days.

Feasibility - the potential for the app to impact smartphone battery life and data usage.

Perceived Risks - the perceived privacy risks associated with using smartphones to monitor travel and trust with how the data were to be used.

Continuance Intention was designed to gauge whether people stated they would continue to use a similar app outside this study and recommend it to others.

Moderators of Behavioural Intention: Self-reported age, gender, education, employment, student status.

App Usage – open-ended questions around what they liked most/least about the app.

3.4 Sampling and Recruitment

Recruitment began in September 2019. Participants meeting the eligibility criteria were invited to a study clinic, where they were given face-to-face instructions by trained research assistants around the many components of the study, including the downloading and use of the app (Sharman et al., 2020). 'Trips' were defined as moving 'from one place to another' and it was noted that 'Moving around your home or workplace is not a trip but traveling from home to the shops/your workplace/a friend's house by any means would count as a trip'. Participants were encouraged to classify their trips in the app daily for seven days, after which the app would provide a 'Finish Program' icon for the participant to confirm that tracking was complete. Participants were encouraged to keep location/GPS and 'mobile data' settings on, take their phone with them on all trips, charge their phone regularly, and to ask for help if needed. Participants were also asked not to delete the app until the seven assessment days had passed, even if they missed days or were having trouble using the app.

In March 2020, in response to COVID-19 restrictions, the study was paused before ultimately being abandoned in May 2020 (Stanesby et al., 2023). At the point of abandonment, 110 participants had completed the T1 assessment, 64 of whom had completed the T2 assessment, with none having completed the T3 assessment. While the intervention component of the study was discontinued, ongoing assessments of travel behaviour during the COVID-19 pandemic were opportunistically continued. Participants of the abandoned trial were invited back in July 2020 to complete two additional assessments (one at T3, then another three months later at T4). In total, 70 participants completed T3 and 67 completed T4. The app experience survey was completed following the T2 assessment period by 56 participants, with 41 participants completing this before the trial was paused, and 15 completing it after. Figure 2 provides a timeline of the assessment periods within the context of COVID-19 statistics and restrictions in Tasmania.

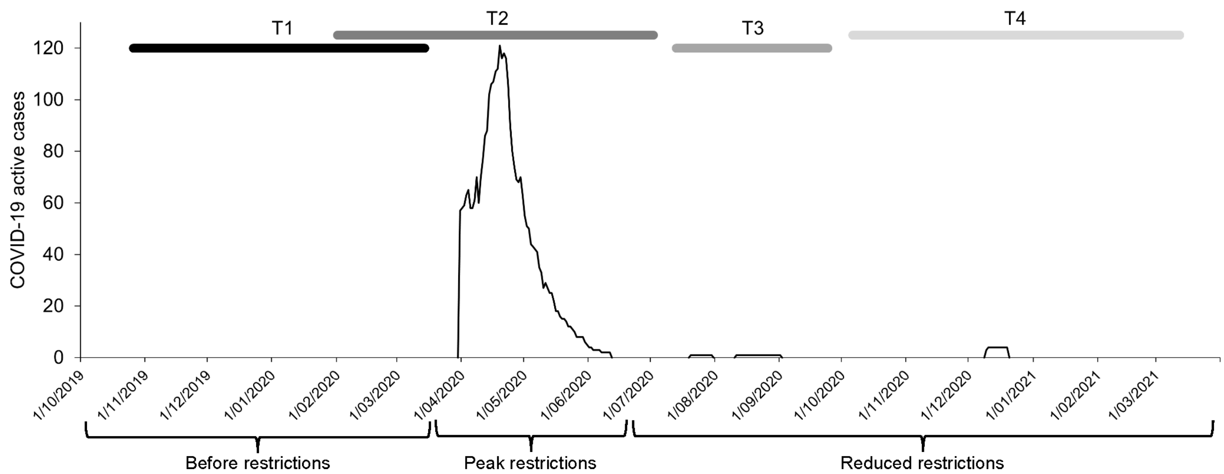

Figure 2: ***Trips4Health*** **Study Timeline (adapted from Stanesby et al., (2023)**

3.5 Analytical approach

Given the many potential confounders impacting the intervention period, the focus here is on establishing the characteristics, app usage and app engagement of those returning for the T4 assessment. Initially, we compare participant socio-demographics, group allocation (control or treatment), study motivations and whether they downloaded and used the app across the assessment periods using simple group percentages. App usage is assessed by comparing high level metrics of app-inferred trip information across the assessment periods using means and standard deviations. These metrics include days with recorded trip segments (indicating the app was used at all), days with moving trip segments (indicating the person made at least one trip), and segments by mode by day (indicating mode choice). Trip-level comparisons are not made as they are not relevant to the current investigation and the trip inference algorithm proved overly sensitive, resulting in many potential spurious trips, which had to be manually checked and either deleted or combined with other trips. App engagement/interaction is evaluated by capturing the percentage of trip segments checked/edited and then confirmed as correct by participants. This provides what we term an *assessment rate* for each participant, which ranged from 0 (no interaction with the app) to 100 (confirmed every trip segment). App user experiences are evaluated using Confirmatory Factor Analysis to examine the reliability of the app experience survey questions in loading onto the pre-defined dimensions of app acceptability detailed previously. We then explore the combined effects of these potential influences using multivariate binary logistic regression techniques. Two models are developed; Model (1) predicts the odds of returning for the T4 assessment for all study participants in T1, while Model (2) predicts the odds of using the app in the T4 assessment for all study participants in T1. App engagement is assessed using simple correlation analysis, given insufficient sample sizes for a multivariate analysis.

4. Results

4.1 Sample characteristics

In total, 110 participants were recruited, completed the clinical assessment and were randomised for the study before it was paused in March, 2020. Characteristics of the T1 and T4 samples together with whether they downloaded and used the app are shown in Table 1. The T1 sample ranged in age from 18-80 years (average age 45) with females (69%) and those with tertiary education (55%) over-represented compared to the Greater Hobart population (52% females and 17% tertiary educated)[2]. Sixty-seven participants (61%) completed the T4 assessment, with higher attrition of younger participants and to a lesser extent those not working. Being assigned to the control or treatment group for the original trial made no difference to the decision to return for T4. In terms of study motivation, after the trial period, 74 participants provided one or more reasons for participating. 'Participating in Research' was the most common reason selected by 53 (72% of participants), followed by 'Increase Exercise' (34%), 'Environmental Reasons' (24%) and 'Financial Reasons' (15%). Retention rates were primarily influenced by whether participants gave a reason at all with attrition lowest for those citing 'Environmental Reasons'.

[2] Australian Bureau of Statistics. Census of Population and Housing: Reflecting Australia - Stories from the census, 2016. Cat no. 2071. Canberra: ABS; 2017.

Table 1: Participant Characteristics and App Usage

	T1-Participants	T1-App Users	T4-Participants	T4-App Users
Total (% of T1-Participants)	**n = 110**	**n = 90 (82%)**	**n = 67 (61%)**	**n = 36 (33%)**
Age				
18-34	40 (36%)	34 (38%)	17 (25%)	11 (31%)
35-54	34 (31%)	29 (32%)	25 (37%)	16 (44%)
55+	36 (33%)	27 (30%)	25 (37%)	9 (25%)
Gender				
Female	76 (69%)	61 (68%)	45 (67%)	23 (64%)
Male	34 (31%)	29 (32%)	22 (33%)	13 (36%)
Education*				
Low	22 (20%)	19 (21%)	12 (18%)	5 (14%)
Medium	27 (25%)	24 (27%)	17 (25%)	8 (22%)
High	61 (55%)	47 (52%)	38 (57%)	23 (64%)
Employment Status				
Full-time (35+ hrs/wk)	29 (26%)	24 (27%)	21 (31%)	13 (36%)
Part-time (20-34 hrs/wk)	24 (22%)	20 (22%)	14 (21%)	9 (25%)
Part-time (<20 hrs/wk)	22 (20%)	17 (19%)	13 (19%)	6 (17%)
Not working	35 (32%)	29 (32%)	19 (28%)	8 (22%)
Student Status				
Non-Student	70 (64%)	58 (64%)	45 (67%)	24 (67%)
Student	40 (36%)	32 (36%)	22 (33%)	12 (33%)
Treatment Group#				
Control	55 (50%)	41 (46%)	33 (49%)	18 (50%)
Treatment	55 (50%)	49 (54%)	34 (51%)	18 (50%)
Study Motivation**	**n=74**	**n=49**	**n=60**	**n=36**
Participating in Research	53 (72%)	41 (84%)	43 (72%)	22 (61%)
Financial Reasons	11 (15%)	10 (20%)	10 (17%)	5 (14%)
Environmental Reasons	18 (24%)	17 (35%)	18 (30%)	13 (36%)
Increase Exercise	25 (34%)	22 (45%)	22 (37%)	14 (39%)
Other Reason	4 (5%)	3 (6%)	3 (5%)	2 (6%)

**Education level, Low: Year 12 or less; Medium: Trade, apprenticeship, certificate, diploma; High: tertiary.*
***Asked after the end of the trial period (T2). Participants were able to select as many of these options, including none at all. Numbers represent those who remained at each stage who had completed this question.*
#Participants were assigned into control or treatment samples as part of the abandoned RCT.

4.2 App usage

Of the 110 baseline participants, 96 agreed to download the app, while 14 refused. Six of those agreeing were found to have no app data recorded, leaving 90 (82%) of participants with recorded app data (Table 1). Evidently, there was little to differentiate those taking up the app in T1 versus those who did not, other than whether they were assigned to the control or intervention groups for the (ultimately abandoned) RCT trial and whether they provided a reason for participating in the study. At T4, just over half of those returning had recorded app data. The percentage of participants who continued to use the app at T4 was higher for 35-54 year-old participants, those with higher education levels, those working more than 20 hours/week and those indicating 'Environmental Reasons' for participation.

Table 2 compares the high-level metrics app-inferred statistics of app usage and engagement across the four waves of data collection – the T2 and T3 waves are provided for completeness. There was a reasonable consistency of app usage across the timepoints based on the number of days of data. Likewise, there was consistency in the proportion of days where moving trip segments were detected (around 90%). There was an increase in app-measured car and bicycle trips during the study period, and a decline in app-measured walking and 'other' trips. The metrics of user engagement with the app (assessment rate) generally improved over the course of the study, suggesting those who continued to use the app were generally more engaged and/or became more used to the app.

Table 2: App-inferred statistics across the four waves

		T1	T2	T3	T4
Number of Participants		90	46	41	36
Days with recorded trip segments (mean, SD)		7.5, 2.1	7.9, 1.8	7.9, 2	7.3, 1.8
Days with moving trip segments (mean, SD)		6.7, 2.2	7.2, 2	7.1, 2.3	6.7, 1.9
Segments per day by Mode (mean, SD)	**Car**	1.8, 1.7	2.3, 1.7	2.2, 1.5	3, 1.7
	Bus	0.2, 0.4	0.3, 0.5	0.2, 0.3	0.2, 0.3
	Bicycle	0.1, 0.2	0.2, 1.1	0.1, 0.4	0.3, 1.2
	Walk#	8.9, 11.5	6.4, 5.7	8.1, 14.1	7.4, 6
	Taxi/Uber	1.5, 1.6	1.4, 1.3	0.9, 1.1	0.8, 1.5
	Total	12.5, 12.7	11.3, 8.6	11.6, 15.3	11.7, 7.4
	Total (excl walk)	3.7, 2.3	4.9, 4.7	3.5, 2.1	4.3, 2.5
Assessment Rate 25/50/75 percentile		17%/37%/74%	27%/57%/83%	32%/48%/76%	40%/63%/84%

#Walk trips were likely over-estimated given the over-sensitivity of the app-based trip detection algorithms.

4.3 App experience survey

The app experience survey was completed by participants who took the app at T2 or T3 and provided 56 usable responses. Confirmatory factor analysis/reliability analysis (Table 3) shows the items generally loaded well onto the pre-defined dimensions (Cronbach's alpha >0.7), the only exception being those under the 'Feasibility' dimension. In this study, battery life emerged as the biggest practical barrier (corroborated by several of the open-ended comments), with mixed sentiments around impacts on data usage. Overall, participants found the app easy to use, useful for recalling their travel, visually appealing and to a lesser extent enjoyable to use. However, there was mixed sentiment over the accuracy of trip recording and how much time was involved having to manually correct trips. Overall, while sentiment was again mixed, there was a suggestion more participants would not continue to use a similar app outside of this study or recommend it to their friends.

Table 3: Dimensions of App Experiences

Dimension	Question	Mean, SD	Cronbach's Alpha	Corrected Item-Total Correlation
Perceived Ease of Use	Installing and registering the app was simple.	6.25, 1.05	0.75	0.46
	The app was easy to use.	4.68, 1.86		0.70
	Using the app fitted with my lifestyle.	5.20, 1.58		0.66
Feasibility	My battery seemed to drain faster when using the app**	2.82, 1.57	0.51	0.36
	My mobile data usage increased when using the app**	3.82, 1.13		0.36
Perceived Usefulness	The app recorded my trips accurately.	3.41, 2.00	0.86	0.73
	The app helped jog my memory about my travel.	5.21, 1.71		0.71
	The app made it quicker to record my travel.	4.68, 1.97		0.75
	I had to spend a lot of time manually adding or changing trips**	3.16, 1.70		0.64
Hedonic Motivation	The app was visually appealing.	4.88, 1.39	0.84	0.58
	The app was enjoyable to use.	3.96, 1.68		0.86
	I felt a sense of fulfillment using the app.	3.95, 1.59		0.72
Perceived Risks	I trust the app with my data.	4.75, 1.64	0.71	0.55
	I was concerned about privacy when using the app**	5.02, 1.54		0.55
Continuance Intention	I would continue to use a similar app outside of this study.	3.36, 1.73	0.83	0.71
	I would recommend the app to my friends.	3.07, 1.62		0.71

*n=56; 7 item-scale (Strongly disagree – strongly agree); **reverse framing*

While experiences with the app were driven by a range of factors, correlation analysis between the dimensions and socio-demographic variables suggests that in general older participants perceived the app to be less useful ($r(56) = -0.54$, $p < 0.001$) and less easy to use ($r(56) = -0.48$, $p < 0.001$) than younger participants. Females appeared more hedonically motivated ($r(56) = 0.35$, $p = 0.008$) and marginally more concerned about the perceived risks of the study ($r(56) = 0.31$, $p = 0.021$) compared to males. None of the other socio-demographic factors listed in Table 1 were significantly correlated. Evidently, there was significant heterogeneity in responses (Figure 3), particularly those around 'Perceived Usefulness', which suggests these experiences were quite participant-specific.

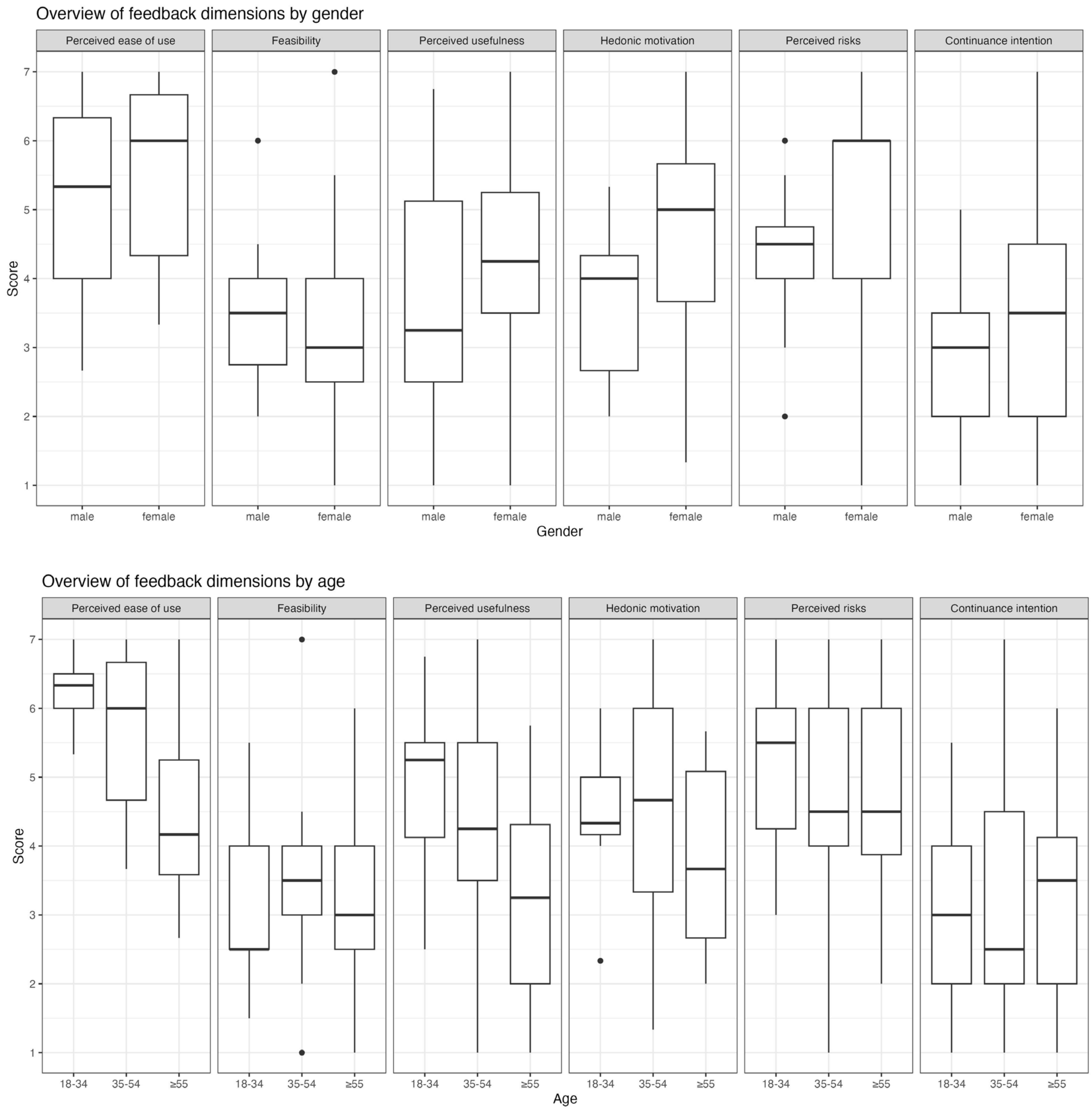

Figure 3: Feedback Dimensions by Gender and Age

Open-ended comments

Participants were given the opportunity to provide open-ended comments about what they liked and disliked about the app. In total, 27 participants provided positive comments, while 39 provided negative comments – these have been assimilated into the major issues, shown in Figure 4. The comments largely re-enforced the insights from the empirical results, suggesting the experience was somewhat participant-specific. Positive comments focused on the accuracy of trip recording and simplicity of using the app, and to a lesser extent the help it provided in recalling travel. The main negative comments were around inaccuracy of trip recording and the resultant burden on participants to correct erroneous trips – the main issue here was an oversensitivity to trip inference, resulting in (some cases) hundreds of spurious trips. Battery drain was the other main negative comment with privacy concerns mentioned by only one participant.

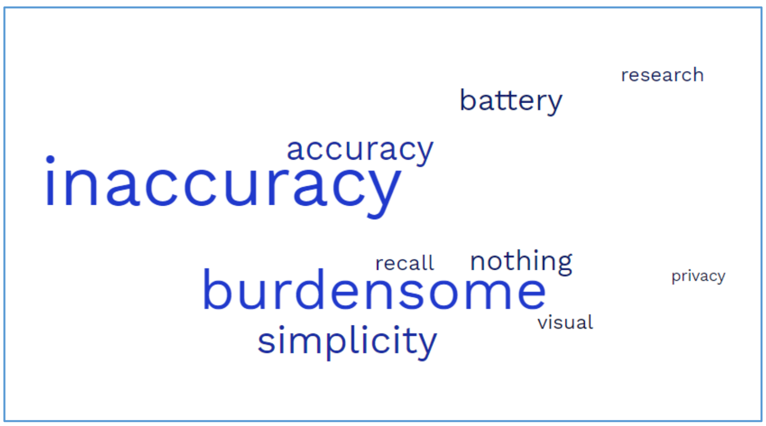

Figure 4: Main issues raised in open-ended comments

4.4. Regression Analysis

The results presented so far suggests that factors pertaining to the participant, their motivation for participating, experiences with the app, and their travel may be factors in both their decision to return in T4 and use/interact with the app. Binary logistic regression models were used to unpack the relative importance of these factors.

Model (1) results presented in Table 4 provide a measure of retention for T1 participants who returned for the T4 assessment. Age was the most significant socio-demographic predictor of returning with older participants between 3 and 4 times more likely to return than the 18–34 age-group. While the number of moving trip segments recorded in T1 were not associated with the decision to return, a single unit increase (1%) in proportion of trips assessed increased the odds of returning by around 3%. All of the stated study participation motivations were associated with greater odds of returning in T4, although the high confidence intervals and small sample size suggest caution be drawn over the strength of this conclusion. None of the app experience survey dimensions were significant in the prediction of participants returning in T4. However, returners were marginally less concerned about privacy *($r(56) = -0.299$, $p = 0.025$)* which, given the correlation with gender, may partially contribute to why proportionately more males came back than females.

Table 4: Model (1) Factors impacting participant retention in T4

Predictors		*aOR [95% CI]*	*p-value*
Gender			
Female		*ref*	
Male		1.46 [0.61, 3.64]	0.408
Age-group (years)			
18 – 34		*ref*	
35 – 54		**4.02 [1.45, 11.94]**	**0.009**
55+		**3.13 [1.23, 8.45]**	**0.020**
Education level			
Low		*ref*	
Medium		0.94 [0.26, 3.30]	0.918
High		0.97 [0.32, 2.85]	0.956
Treatment group			
Control		*ref*	
Intervention		1.12 [0.50, 2.55]	0.784
*Trip metrics (T1)**			
Moving trip segments (*n*)		1.01 [1.00, 1.01]	0.105
Moving trip segments assessed (%)		**1.03 [1.01, 1.05]**	**0.005**
Study participation motivations†			
Participation in research	No	*ref*	
	Yes	**6.24 [1.83, 24.21]**	**0.005**
Financial reasons	No	*ref*	
	Yes	**13.09 [1.59, 318.05]**	**0.043**
Environmental reasons	No	*ref*	
	Yes	**18.93 [2.84, 397.9]**	**0.011**
Increase PA or exercise	No	*ref*	
	Yes	**7.29 [1.90, 39.28]**	**0.009**

aOR = adjusted odds ratio; 95% CI = 95% confidence interval;

** These metrics are included individually in separate models that adjusted for gender, age, and education level.*

† Participants were able to select as many of these four options, including none at all. These are treated as individual binary options in separate models that adjusted for gender, age, and education level.

Model (2), presented in Table 5, provides a measure of diligence based on T1 participants who both returned and used the app in T4. Noting the small sample size, none of the socio-demographic indicators were significant, despite the unadjusted analyses indicating older participants were less likely to use the app again. As with returners (shown in model (1)), while the number of moving segments recorded in T1 was not significantly associated with re-using the app, both the number and proportion of assessed trips were, increasing the odds by 4% and 2% respectively. Among study motivations, only environmental reasons were significantly associated with greater odds of using the app in T4. App feedback dimensions were insignificant across the board, but large confidence intervals were observed (results not tabulated).

Table 5: Model (2) Factors impacting participant re-use of the app in T4 (n = 67)

Predictors	*aOR [95% CI]*	*p-value*
Gender		
Female	*ref*	
Male	1.27 [0.42, 3.90]	0.675
Age-group (years)		
18 – 34		
35 – 54	1.00 [0.24, 4.15]	0.997
55+	0.34 [0.08, 1.26]	0.112
Education level		
Low		
Medium	1.12 [0.21, 6.16]	0.898
High	1.80 [0.44, 7.83]	0.415
Treatment group		
Control	*ref*	
Intervention	1.00 [0.35, 2.84]	0.999
*Trip metrics (T1)**		
Moving segments (*n*)	1.00 [1.00, 1.00]	0.143
Moving segments assessed (%)	**1.02 [1.00, 1.04]**	**0.046**
Study participation motivations†		
Participation in research	1.70 [0.25, 12.12]	0.582
Financial reasons	0.83 [0.16, 4.50]	0.824
Environmental reasons	**12.34 [2.41, 94.78]**	**0.006**
Increase PA or exercise	2.11 [0.59, 8.15]	0.260

aOR = adjusted odds ratio; 95% CI = 95% confidence interval;
** These metrics are included individually in separate models that adjusted for gender, age, and education level.*
† Participants were able to select as many of these four options, including none at all. These are treated as individual binary options in separate models that adjusted for gender, age, and education level.

A third model was developed to assess app engagement in T4, using the proportion of assessed trips as a proxy for app engagement, with participant characteristics and app engagement in T1 as predictor variables. However, the sample size (n = 36) precluded a multivariable analysis with sufficient power so insights were drawn from correlation analysis. Among T1 app users, trip assessment rates were higher for males than females ($r(89) = -0.221, p = 0.037$) and higher for those motivated to participate in the study for environmental reasons ($r(60) = 0.348, p = 0.006$) and for exercise ($r(60) = 0.315, p = 0.014$). Among T4 app users, the 'Feasibility' dimension was marginally significant in predicting assessed trips ($r(33) = 0.345, p = 0.049$)¸ suggesting issues around battery drain and data usage were still important issues among those who had continued to use the app, but did not necessarily interact with it as much. Having a higher number of moving segments was associated with a lower proportion of assessed trips in both T1 ($r(89) = -0.247, p = 0.02$) and T4 waves ($r(36) = -0.614, p < 0.001$).

5. Discussion

Longitudinal studies provide a mechanism for exploring the dynamics of behavioural change and new technologies offer opportunities to enhance our ability to measure these changes. Using data from a longitudinal study of travel and physical activity involving a travel app and repeated surveys, this study shows a high initial uptake is important but app engagement is not necessarily sustained and participant attrition is considerable but non-random. We discuss the characteristics and motives associated with initial and sustained travel app engagement and outline the barriers and some potential strategies to ensure sustained engagement with travel apps in longitudinal research studies.

First, it is evident that attrition in longitudinal studies is non-random, with challenges around the retention of young participants most pronounced. This broadly mirrors what has been found elsewhere although the loss of males, which is typically greater, is not so evident here (Greaves et al., 2013; Greaves et al., 2015). Second, the proportion of participants who took the app initially was higher than is typically experienced in similar studies – this could have been a function of the opportunities and sense of obligation provided by the face-to-face instruction at the study clinic, which included help with downloading and installing the app. However, evidently, many participants subsequently failed to engage with the app as hoped, which appears to be a construct of challenges around the app itself, partially exacerbated by the amount of travel made. Similar outcomes are reported in (Greaves et al., 2015). Third, participants who returned in T4 were less likely to take the app again, but for those that did, they were generally more engaged with the app. This implies caution for users of longitudinal data that samples generally become less representative and we may need to be cautious about treating data from the same individual as a constant.

What learnings can we take away as designers of long-term evaluations and travel apps? First, it is clear that achieving buy-in up-front and engagement at various points through the study is important as has been frequently reported elsewhere (Nguyen et al. (2023). Second, while we may feel a need to compensate participants for their time/effort, the findings here suggest other factors may be equally important in encouraging participation, related to clearly elucidating the importance of the research and stressing the wider importance of the results – in this study, the messaging around health/environment was evidently important in encouraging participation, complementing evidence around the importance of using messaging that appeals to people's broader values rather than direct personal benefit (Brüggen et al., 2011). Third, while smartphone apps offer intrinsic appeal to overcome many of the challenges associated with recall, many obstacles remain. We cannot compel people to 'play ball' in terms of downloading an app, give permissions, charge and carry their phone, and retain use of the app. Perceptions around impacts of tracking apps on battery life remain a problem, although it is interesting here that mobile data usage did not emerge as a major concern. A plausible explanation is that while battery life can be heavily compromised by tracking apps, they are not data hungry as most of the process involves geolocation which does not require data. Added to this, the price of mobile data has gone down substantially in Australia (Brüggen et al., 2011). Compounding these challenges are factors outside our control such as the myriad of makes/models of phones, operating systems and frequent system upgrades, privacy settings etc. One option is to furnish participants with devices, but we then run into issues around participants remembering to keep them with them and charged (Geurs et al., 2015). We also need to think about the user experience and an 'acceptable' level of user involvement. While eliminating user involvement might be perceived as the 'end goal', evidently fully passive apps are not ready for deployment (Harding et al. 2021). In our case, the app scored highly on hedonic motivation and ease of use, suggesting users liked the interface and generally enjoyed using it. However, they clearly did not like, what turned out to be a highly onerous task for many, having to check/correct trips.

We acknowledge several limitations to this study. First, it is highly likely that participants who signed up for this study were likely more motivated than the general population to complete the various requirements including taking the app. Second, events beyond our control impacted sample size and timing of some of the attitudinal surveys. While the app experience survey was conducted on a subsample of participants who had already used the app twice and therefore probably had more positive views than those not using it again, it captures the views of the majority of returning participants in T2 or T3. This still provides useful insights from those who returned and equally those who did not, to inform future surveys of this nature. Third, there was evidently a range of app experiences among participants. However, we argue this is more a realistic outcome of the investigation as opposed to a limitation.

6. Conclusions

The current paper adds to our understanding of the use of smartphone-based travel survey apps as part of a longitudinal methodology used to investigate the temporal dynamics of behaviour change post-intervention. Caveats associated with the unanticipated disruption of this study via the COVID-19 pandemic aside, it is evident that attrition and app engagement are non-random and should be addressed ideally through recruitment and engagement strategies. In this study, we identified potential conundrums for survey designers to ponder, such as that while older participants are more likely to stick with the survey, they are less likely to engage with the app on which we largely pinned our data collection hopes. Likewise, those recording less travel are more likely to engage with the app, evidently more travel is synonymous with more burden. Clearly, getting the messaging right for recruiting and retaining participants is important along with the financial rewards for what was a highly burdensome and lengthy commitment. Interestingly, while financial reasons were not deemed significant for those returning for T4 (when the trial itself and chance to earn public transport credits had gone), this comes with the caveat this was asked at the beginning of the study. Smartphone apps unquestionably have the potential to make the experience more engaging and useful for some, but equally more frustrating and burdensome for others - it seems likely that other options for completing travel surveys (independent and supplementary to smartphone apps) will continue to be needed for the foreseeable future.

7. References

Assemi, B., Jafarzadeh, H., Mesbah, M. & Hickman, M. 2018. Participants' perceptions of smartphone travel surveys. *Transportation Research Part F: Traffic Psychology and Behaviour,* 54**,** 338-348.

Booker, C. L., Harding, S. & Benzeval, M. 2011. A systematic review of the effect of retention methods in population-based cohort studies. *BMC Public Health,* 11**,** 249.

Brüggen, E., Wetzels, M., De Ruyter, K. & Schillewaert, N. 2011. Individual Differences in Motivation to Participate in Online Panels: The Effect on Reponse Rate and Reponse Quality Perceptions. *International Journal of Market Research,* 53**,** 369-390.

Bürbaumera, M., Kirchbergera, C., Berger, M. & Dorner, J. 2022. Recruitment process, participant motivation and response of a smartphone-based travel survey: Learnings from the Mobility Panel in aspern Seestadt. *12th International Conference on Transport Survey Methods.* Lisbon, Portugal: Transportation Research Procedia.

Davis, F. D. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly,* 13**,** 319-340.

Eisenmann, C., Chlond, B., Minster, C., Jödden, C. & Vortisch, P. 2019. Assessing the effects of a mixed-mode design in a longitudinal household travel survey. *Transportation,* 46**,** 1737-1753.

Ellison, A., Verzosa, N., Greaves, S. & Ellison, R. 2017. Who stays in? An analysis of participation and data quality in longitudinal transport surveys. *11th International Conference on Transport Survey Methods.* Quebec, Canada.

Faghih Imani, A., Harding, C., Srikukenthiran, S., Miller, E. J. & Nurul Habib, K. 2020. Lessons from a Large-Scale Experiment on the Use of Smartphone Apps to Collect Travel Diary Data: The "City Logger" for the Greater Golden Horseshoe Area. *Transportation Research Record,* 2674**,** 299-311.

Gadziński, J. 2018. Perspectives of the use of smartphones in travel behaviour studies: Findings from a literature review and a pilot study. *Transportation Research Part C: Emerging Technologies,* 88**,** 74-86.

Geurs, K. T., Thomas, T., Bijlsma, M. & Douhou, S. 2015. Automatic Trip and Mode Detection with Move Smarter: First Results from the Dutch Mobile Mobility Panel. *Transportation Research Procedia,* 11**,** 247-262.

Greaves, S., Ellison, A., Ellison, R., Rance, D., Standen, C., Rissel, C. & Crane, M. 2015. A Web-Based Diary and Companion Smartphone app for Travel/Activity Surveys. *Transportation Research Procedia,* 11**,** 297-310.

Harding, C., Faghih Imani, A., Srikukenthiran, S., Miller, E. J. & Nurul Habib, K. 2021. Are we there yet? Assessing smartphone apps as full-fledged tools for activity-travel surveys. *Transportation,* 48**,** 2433-2460.

Hardy, A., Aryal, J. & Wells, M. 2019. Comparing techniques for tracking: the case of Tourism Tracer in Tasmania, Australia. *E-review of Tourism Research,* 16**,** 84-94.

Hong, S., Zhao, F., Livshits, V., Gershenfeld, S., Santos, J. & Ben-Akiva, M. 2021. Insights on data quality from a large-scale application of smartphone-based travel survey technology in the Phoenix metropolitan area, Arizona, USA. *Transportation Research Part A: Policy and Practice,* 154**,** 413-429.

Jose, K. A., Sharman, M. J., Stanesby, O., Greaves, S., Venn, A. J., Blizzard, L., Palmer, A., Cooper, K., Williams, J. & Cleland, V. J. 2022. Incentivising public transport use for physical activity gain: process evaluation of the COVID-19 disrupted trips4health randomised controlled trial. *International Journal of Behavioral Nutrition and Physical Activity,* 19**,** 157.

La Paix Puello, L., Olde-Kalter, M.-J. & Geurs, K. T. 2017. Measurement of non-random attrition effects on mobility rates using trip diaries data. *Transportation Research Part A: Policy and Practice,* 106**,** 51-64.

Leavens, E. L. S., Stevens, E. M., Brett, E. I., Molina, N., Leffingwell, T. R. & Wagener, T. L. 2019. Use of Rideshare Services to Increase Participant Recruitment and Retention in Research: Participant Perspectives. *J Med Internet Res,* 21**,** e11166.

Nguyen, T., Thomas, A. J., Kerr, P., Stewart, A. C., Wilkinson, A. L., Nguyen, L., Altermatt, A., Young, K., Heath, K., Bowring, A., Fletcher-Lartey, S., Lusher, D., Hill, S., Pedrana, A., Stoové, M., Gibney, K. & Hellard, M. 2023. Recruiting and retaining community-based participants in a COVID-19 longitudinal cohort and social networks study: lessons from Victoria, Australia. *BMC Medical Research Methodology,* 23**,** 54.

Ortúzar, J. D. D., Armoogum, J., Madre, J. L. & Potier, F. 2011. Continuous Mobility Surveys: The State of Practice. *Transport Reviews,* 31**,** 293-312.

Otahal, P., Blizzard, L., Hosmer, D. W., Stankovich, J. & Venn, A. 2021. Characterising attrition from childhood to adulthood in a 20-year cohort: which baseline factors are influential, and can bias be corrected? *Longitudinal and Life Course Studies,* 12**,** 227-240.

Prelipcean, A. C., Susilo, Y. O. & Gidófalvi, G. 2018. Collecting travel diaries: Current state of the art, best practices, and future research directions. *Transportation Research Procedia,* 32**,** 155-166.

Scheepers, E. & Hoogendoorn-Lanser, S. 2018. State-of-the-art of incentive strategies – Implications for longitudinal travel surveys. *Transportation Research Procedia,* 32**,** 200-210.

Sharman, M. J., Ball, K., Greaves, S., Jose, K. A., Morse, M., Blizzard, C. L., Wells, G., Venn, A. J., Palmer, A. J., Lester, D., Williams, J., Harpur, S. & Cleland, V. J. 2020. trips4health: Protocol of a single-blinded randomised controlled trial incentivising adults to use public transport for physical activity gain. *Contemporary Clinical Trials Communications,* 19**,** 100619.

Shen, L. & Stopher, P. R. 2014. Review of GPS Travel Survey and GPS Data-Processing Methods. *Transport Reviews,* 34**,** 316-334.

Singer, E. & Ye, C. 2013. The Use and Effects of Incentives in Surveys. *The ANNALS of the American Academy of Political and Social Science,* 645**,** 112-141.

Stanesby, O., Greaves, S., Jose, K., Sharman, M., Blizzard, L., Palmer, A. J., Evans, J., Cooper, K., Morse, M. & Cleland, V. 2023. A prospective study of the impact of COVID-19-related restrictions on activities and mobility upon physical activity, travel behaviour and attitudes. *Journal of Transport & Health***,** 101624.

Stopher, P. R., Moutou, C. J. & Liu, W. 2013. Sustainability of voluntary travel behaviour change initiatives: A 5-year study. *Australian Transport Research Forum 2013 Proceedings.* Brisbane.

Teague, S., Youssef, G. J., Macdonald, J. A., Sciberras, E., Shatte, A., Fuller-Tyszkiewicz, M., Greenwood, C., McIntosh, J., Olsson, C. A., Hutchinson, D., Bant, S., Barker, S., Booth, A., Capic, T., Di Manno, L., Gulenc, A., Le Bas, G., Letcher, P., Lubotzky, C. A., Opie, J., O'Shea, M., Tan, E., Williams, J. & the, S. L. S. T. 2018. Retention strategies in longitudinal cohort studies: a systematic review and meta-analysis. *BMC Medical Research Methodology,* 18**,** 151.

Thomas, T., Geurs, K. T., Koolwaaij, J. & Bijlsma, M. 2018. Automatic Trip Detection with the Dutch Mobile Mobility Panel: Towards Reliable Multiple-Week Trip Registration for Large Samples. *Journal of Urban Technology,* 25**,** 143-161.

Van der Mispel, C., Poppe, L., Crombez, G., Verloigne, M. & De Bourdeaudhuij, I. 2017. A Self-Regulation-Based eHealth Intervention to Promote a Healthy Lifestyle: Investigating User and Website Characteristics Related to Attrition. *J Med Internet Res,* 19**,** e241.

Venkatesh, V., Thong, J. Y. L. & Xu, X. 2012. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly,* 36**,** 157-178.

Verzosa, N., Greaves, S., Ellison, R., Ellison, A. & Davis, M. 2018. Eliciting preferences for 'gamified' travel surveys: a best-worst approach. *Transportation Research Procedia,* 32**,** 211-223.

Verzosa, N., Greaves, S., Ho, C. & Davis, M. 2021. Stated willingness to participate in travel surveys: a cross-country and cross-methods comparison. *Transportation,* 48**,** 1311-1327.

Zhao, F., Pereira, F. C., Ball, R., Kim, Y., Han, Y., Zegras, C. & Ben-Akiva, M. 2015. Exploratory Analysis of a Smartphone-Based Travel Survey in Singapore. *Transportation Research Record,* 2494**,** 45-56.