Journal of Transport & Health 1 (2014) 190–201

Contents lists available at ScienceDirect

Journal of Transport & Health

journal homepage: www.elsevier.com/locate/jth

High group level validity but high random error of a self-report travel diary, as assessed by wearable cameras

P. Kelly [a,*], A. Doherty [a], A. Mizdrak [a], S. Marshall [b], J. Kerr [b], A. Legge [b], S. Godbole [b], H. Badland [c], M. Oliver [c], C. Foster [a]

[a] British Heart Foundation Health Promotion Research Group, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK
[b] Department for Family and Preventive Medicine, University of California San Diego, La Jolla, CA, USA
[c] Centre for Physical Activity and Nutrition, Auckland University of Technology, New Zealand

ARTICLE INFO

Article history:
Received 11 March 2014
Received in revised form
24 April 2014
Accepted 29 April 2014
Available online 5 June 2014

Keywords:
Travel measurement
Active travel
Self-report
Travel diaries
Wearable cameras

ABSTRACT

Self-report remains the most common method for collecting epidemiological evidence of the links between travel and health outcomes. This study assesses the validity and reliability of a self-reported travel diary (a modified version of a well-established UK travel diary; The National Travel Survey (NTS)) by comparison with wearable camera data.

Across four locations (Oxford, UK; Romford, UK; San Diego, USA; and Auckland, New Zealand) we collected 3–4 days of SenseCam (wearable camera) and travel diary data from 84 adult participants (purposive sample). Compliance with the data collection protocol was high and inspection of the crude results suggests acceptable agreement between measures for total days of data collected (diary=278; SenseCam=274), daily journey frequency (diary=4.78; SenseCam=4.64) and average journey duration in minutes (diary=17:46; SenseCam=15:40). Once these data were examined for total daily time spent travelling in minutes agreement was poorer (diary=84:53; SenseCam=72:35).

Analysis of matched pairs of journey measurements (n=1127) suggests a positive bias on self-reported journey duration of 2:08 min (95% CI=1:48–2:28; 95% limits-of-agreement=−9:10 to 13:26). Similar analysis of diary days matched to complete SenseCam days (n=201) showed a very small positive bias with a very large limits-of-agreement (1:41 min; 95% CI=−2:00 to 5:24; 95% limits-of-agreement=−50:29 to 53:41).

These results suggest self-reported journey and daily travel exposure data are relatively valid at a population level, though corrections according to reported bias could be considered. The large limits of agreement for matched journey and diary summary analysis suggest self-report diaries may be unsuitable for assessment of an individual's travel behaviour.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Active travel (walking and cycling for transport) is an important domain of physical activity (Department of Health, 2004; World Health Organisation, 2007), and can confer substantial health benefits (Bull et al., 2010; Department of Health, 2004; Oja et al., 2011; Woodcock et al., 2011; World Health Organisation, 2007). Travelling by active modes can also lead to environmental benefits when replacing vehicle travel (Ogilvie et al., 2007; Oliver et al., 2010; Woodcock et al., 2013). Sedentary travel by motor vehicle has been associated with adverse health outcomes (Florez Pregonero et al., 2012) and contributes to carbon emissions (Woodcock et al., 2013).

Valid and reliable measures of travel behaviour are required to support epidemiological research, health and transport modelling, and intervention evaluation (Sallis et al., 2000). Despite technological advances with objective measures such as GPS (Global Positioning Systems) devices and accelerometers, self-reported (subjective) measures remain more commonly used in population travel research for reasons of feasibility, cost, comparability, and ease of data handling (Armstrong and Welsman, 2006). Indeed, recent systematic reviews of

* Corresponding author. Tel.: +44 1865 617 782.
E-mail addresses: paul.kelly@dph.ox.ac.uk (P. Kelly), aiden.doherty@dph.ox.ac.uk (A. Doherty), anja.mizdrak@dph.ox.ac.uk (A. Mizdrak), sjmarshall@ucsd.edu (S. Marshall), jkerr@ucsd.edu (J. Kerr), legge.amanda@gmail.com (A. Legge), sgodbole@ucsd.edu (S. Godbole), hannah.badland@unimelb.edu.au (H. Badland), melody.oliver@aut.ac.nz (M. Oliver).

http://dx.doi.org/10.1016/j.jth.2014.04.003
2214-1405/© 2014 Elsevier Ltd. All rights reserved.

the health benefits of active travel found that most studies used self-report measures (diary or questionnaire) (Hamer and Chida, 2008; Lubans et al., 2011; Oja et al., 2011; Woodcock et al., 2011) and many countries employ some form of travel diary to monitor national transport trends (Stopher and Greaves, 2007).

Self-reported travel diaries may have limited validity and reliability, due to issues with recall, social desirability, and comprehension (Krizek et al., 2009; Welk, 2002; Yang et al., 2010). However, the validity and reliability of travel diaries are rarely quantified beyond correlation coefficients or test-retest reliability, limiting a clear understanding of quality of data collected by these methods. Understanding these parameters will allow better assessment of the suitability of travel diaries for different study designs and of the uncertainty on conclusions drawn from self-reported travel data.

The objective of this study is to investigate the validity (in this case systematic bias) and reliability (random error) of self-reported travel behaviour data using wearable cameras (SenseCam). Random error reveals the degree of reliability (Carmines and Zeller, 1979; Weisberg, 2009); however this study will not investigate other established forms of reliability such as test–retest or inter-rater reliability (Medical Research Council). SenseCam automatically records 2000–3000 first-person point-of-view images per day which have been used to objectively observe, classify, and quantify certain health-related behaviours. It has previously been used in pilot studies for travel (Kelly et al., 2011, 2012), sedentary behaviour (Kerr et al., 2013), physical activity (Doherty et al., 2013b), and diet (Gemming et al., 2013). A full description of the device is available elsewhere (Doherty et al., 2013a). The use of coded images enables the objective classification of travel behaviour dimensions (mode, duration and frequency) (Kelly et al., 2011, 2012). This study has two aims: to assess the validity and reliability of (1) self-reported journey durations; and (2) self-reported daily estimates of total travel time.

2. Methods

2.1. Ethics approval

This study received ethical approval from all participating institutions; University of Oxford (IDREC: SSD/CUREC1A/10-092); Auckland University of Technology (AUTEC 11/114); and University of California San Diego (#071243). We also developed ethical guidance for wearable camera research which guided the study protocols (Kelly et al., 2013b).

2.2. Study locations

Participants were recruited in four locations to give a range of High Income Country travel environments and contexts. The locations were selected from the sites of the three collaborating institutions; Oxford, Oxfordshire, UK and Romford, Essex, UK; Auckland, New Zealand and San Diego, California, USA.

2.3. Recruitment

All participants were volunteers. Recruitment in Oxford, Romford, and Auckland was active in nature; participants were approached from existing networks by email or in person. In San Diego, a more passive approach was adopted and adverts were placed (email and flyers on the university campus), with a particular request for cycle commuters. In San Diego a $50 compensation for the participants' time was also offered, in line with institutional policy.

The sample size estimate of 85–90 participants was based on pilot data (Kelly et al., 2011). This estimate was designed to provide a minimum meaningful uncertainty (95% confidence interval) around the bias between methods (at a journey level); in this case approximately 60 s. Pilot data were used to estimate expected number of journeys per participant per day. No pilot data existed for estimates based on daily summary travel duration.

2.4. Study protocol and study visit

Having volunteered to take part, participants were provided with information sheets, and given the opportunity to ask questions before providing informed consent. Post-consent, participants attended a group or individual session where the study protocol was presented and participants completed a short demographic survey. These sessions took between 15 and 20 min.

2.5. Travel diary protocol

Participants were given a three or four day travel diary (four days if collection period was over a weekend for scheduling reasons) which was a modified, un-validated version of a travel diary used in the UK National Travel Survey (NTS) (Department for Transport, 2010). The diary asked participants to record journey mode, purpose, and duration. In accordance with the NTS protocol, participants were asked to only include travel time, and not other activities such as waiting for public transport or sitting in a stationary car waiting to collect a passenger. Participants were asked to record any travel between two locations with an estimated journey time of greater than three minutes and less than 120 min. Recording journeys less than three minutes for 3–4 days was considered too burdensome for participants, and journeys over 120 min were considered long, irregular journeys, and therefore not part of our research question. These criteria are similar to the NTS protocols (Rofique et al., 2011). As with the NTS protocols, participants were also provided with a pocket sized travel log (memory jogger) so they could record their journeys throughout the day and transfer this to the survey when appropriate.

2.6. SenseCam (wearable camera) protocol

SenseCam devices were worn around the neck on a lanyard during waking hours. Participants were told that they could remove the device at any time (e.g., if feeling uncomfortable, asked to do so by a colleague or family member, or at a setting (e.g., bank, airport) where camera use was inappropriate). Participants were also instructed how to stop image recording for a seven minute period by using a privacy button. To minimise reactivity in reporting (more "accurate" diary reporting as a result of knowing this was being examined and studied), participants were told the study aimed to combine images and travel diary data. They were not explicitly informed the images would be used to assess "accuracy" of reporting. The participants were sent an optional reminder email or SMS text message to remind them to charge the device each day. The device was configured to scramble images, which could only be unscrambled on the laptops of the researchers administering this study. All protocols were in accordance with the ethical framework developed for wearable camera research (Kelly et al., 2013b).

2.7. Return of devices and data

Return study visits were conducted by a researcher, in person, and one-on-one, at which time SenseCam images were downloaded by the researcher and the participant was given the option to delete any images they wished (in private if preferred). Images were viewed on a laptop computer using standard SenseCam software (Doherty et al., 2011). Travel diaries were also collected at this time. Return study visits took between 25 and 30 min per participant.

2.8. Data annotation and coding

SenseCam image coding was conducted according to a protocol objectively validated against direct observation (Kelly, 2013, pp. 86–122) and shown to have high inter-rater reliability (Kelly et al., 2012). The following data and variables were recorded for both the travel diary and SenseCam journeys: location, date, mode, duration, purpose, whether journey was part of a trip-chain, and reason trip was missed (if applicable). In some instances, journeys were noted in the travel diary but analysis of the images showed that no such journey occurred on that day. These journeys were recorded as "phantom journeys". A second "coding pass" was made to check for segmentation fidelity and missed journeys.

2.9. Journey extraction and data analysis

Eligible journeys were those meeting the 3–120 min reporting criteria. For individual journey analysis (Study aim 1), we identified matched-pairs of journeys (using times of day and journey sequence). For daily summary analysis (Study aim 2) we calculated summaries for the following:

(A) All (crude) SenseCam data – summary daily travel time for all days of image data.
(B) SenseCam criterion data – summary daily travel time for days with full image data.
(C) All (crude) travel diary data – summary daily travel time for all days of diary data.
(D) Travel diary data corresponding to criterion SenseCam data – summary daily travel time on diary days corresponding to (B) SenseCam criterion days.

Agreement (bias) between methods was examined using the paired-*t* statistic providing the mean difference between methods, its 95% confidence interval, and its standard deviation. We plotted Bland–Altman 95% limits of agreement to distinguish systematic error (bias) and random error (limits of agreement) (Bland and Altman, 2007). Upper and lower 95% limits of agreement around the bias were calculated from $\pm 1.96 \times$ (standard deviation of the bias) where the bias is determined from the paired-*t* statistic.

The correlation coefficient was calculated for the relationship between the journey durations estimated from the two methods using both within-subject (Bland and Altman, 1995a) and between-subject methods (Bland and Altman, 1995b).

We conducted the following checks on the data set: assessment of whether SenseCam and travel diary differences were normally distributed (visual inspection of kernel density plot and Shapiro–Wilk test), tests for heteroscedasticity or proportional bias (visual inspection and Kendal's Tau), and checks on coding and data entry on a randomly selected 10% of the data set. All statistical analyses were conducted using *PASW* Statistics 18.0 (SPSS Inc., Chicago, IL, USA) software package.

3. Results

3.1. Data and participants

3.1.1. Data collection and coding

Data collection was phased over four locations; Auckland in June 2011 (winter); San Diego in July 2011 (summer); Oxford from May 2012 to October 2012 (summer–autumn) and Romford in September 2012 (autumn). Data coding took place between October 2012 and January 2013. The initial "coding pass" took one researcher approximately eight weeks, or 300 h (corresponding to 1–1.5 h per participant day of image data). A second "coding pass" to check coding fidelity took a further four weeks. Data analysis took place in January and February 2013.

3.1.2. Participants and compliance

Overall, 99 people were engaged, and 88 consented to the study. Between Oxford and Romford, 11 potential participants (11.1%) declined to take part after receiving the participant information. Their reasons were concerns about cameras in workplace ($n=6$); invasion of privacy ($n=2$); too burdensome or insufficient reward ($n=2$), and inconvenient timing of study ($n=1$). Four participants dropped out after consenting without giving any reason, resulting in a final sample of 84 participants (84.8%). Findings presented relate to these 84 participants.

Table 1 shows the participant demographics. San Diego provided the largest proportion of participants. The educational status of the sample was relatively high, probably due to high university-based recruitment. Participant ages ranged from 19 to 60 years, and there were almost equal proportions of males and females.

Table 1
Participant demographics for $n=84$ who completed the study protocol.

Characteristics	Overall		Auckland, New Zealand		San Diego, USA		Oxford, UK		Romford, UK	
	n	%	***n***	%	***n***	%	***n***	%	***n***	%
Participants	**84**	100	15	17.9	37	44.0	25	28.9	7	8.3
Male	**44**	52.4	3	3.6	26	31.0	11	13.1	4	4.8
Female	**40**	47.6	12	14.3	11	13.1	14	16.7	3	3.6
Educational status										
Completed high school	**13**	15.5	0	0.0	3	3.6	5	6.0	5	6.0
Current undergraduate student	**19**	22.6	0	0.0	2	2.4	16	19.0	1	1,2
Completed undergraduate degree	**16**	19.0	0	0.0	15	17.9	0	0.0	1	1,2
Completed graduate degree	**36**	42.9	15	17.9	17	20.2	4	4.8	0	0.0
Age (mean in years); range	**33.3**		38.3		37.7		26.0		25.3	
Standard deviation	**19–60**		21–57		21–60		19–51		19–49	
	12.3		10.3		12.6		8.5		10.8	

Table 2
Eligible SenseCam and travel diary journeys.

	SenseCam	**Travel diary**
Days of data returned	274	278
Total number of journeys recorded	1629	1369
Greater than 120 min	5 (0.3%)	6 (0.4%)
Less than 3 min	354 (21.7%)	35 (2.6%)
Total eligible journeys	1270 (80.0%)	1328 (97.0%)

3.1.3. *Days of travel data*

Table 2 displays the travel data collected by the two measures. From an expected 307 days of data, SenseCam returned 274 days (89.3%) and the travel diary 278 days (90.6%). There was lower compliance in both UK locations (Oxford: SenseCam=82.5%, travel diary=90.0%; Romford: SenseCam=70.0%, travel diary=56.0%), compared with Auckland (SenseCam=91.3%, travel diary=82.6%) and San Diego (SenseCam=96.2%, travel diary=98.7%).

Days of SenseCam data ($n=33$) were missed due to device malfunction, or participants failing to charge the device or forgetting to wear the device. Days of travel diary data ($n=29$) were missed due to the participant failing to record any data in the returned diary (i.e., returned blank diary when SenseCam data for same day showed journeys occurring). One participant returned full SenseCam data, but had lost their travel diary.

After application of the 3–120 min criteria, the SenseCam data yielded 1270 journeys and the travel diary data 1328 journeys. These data show the travel diary slightly outperformed the SenseCam in term of days of data collected ($n=4$; 1.4%) and in terms of number of eligible journeys collected ($n=58$; 4.6%). However, it was later shown that many travel diary journeys were less than 3 min and therefore incorrectly reported, or in some cases never occurred (phantom journeys).

3.1.4. *Eligible travel data*

Table 3 shows journey frequency and average duration of eligible journeys by travel mode. There was broad agreement between the two measures. Two-thirds of trips were active (walking or cycling); on average, walking trips had the shortest durations. Average travel diary journey durations for all modes were greater by 2:06 min per journey (12.1% of average reported time). Data collected via travel diaries yielded 0.12 journeys per day more than the SenseCam data, though fewer walking journeys were captured from travel diary data than the SenseCam data.

3.2. *Study aim 1 – assessing the validity and reliability of self-reported journey durations*

In total 1127 journey pairs over 227 days were matched (4.96 journey pairs per day), representing 88.7% (SenseCam) and 84.7% (travel diary) of the total eligible journeys recorded and 80.0% (SenseCam) and 78.8% (travel diary) of the days of data. Of these pairs, 444 (39.4%; 3.01 per day) were sedentary modes and 638 (60.6%; 1.96 per day) were active modes. These data compare well to the unmatched data (Table 3) for journey frequency and proportions of active and sedentary modes, suggesting the process of selecting matched journeys did not introduce meaningful bias. Journey frequency was slightly increased for both the SenseCam and travel diary (see Table 3) as a result of extracting matched pairs. Journey durations for pairs compared well to those for the unmatched data.

At the individual journey level the mean difference between SenseCam and the travel diary pairs was 2:08 min (representing an over-reporting bias in the travel diary). This compares favourably to the difference in unmatched data (shown in Table 3). A low standard error (0:10 min) gave a narrow 95% confidence interval around this estimate (1:48 to 2:28 min). A relatively large standard deviation (5:46 min) gave wide 95% limits of agreement around this estimate (−9:10 to 13:26 min).

Fig. 1 shows a Bland–Altman plot of difference in journey duration (between SenseCam and travel diary data) against the average duration (of SenseCam and travel diary data) for all matched pairs; the mean difference (bias in minutes) and the 95% limits of agreement are displayed. The Bland–Altman plot illustrates the small, statistically significant fixed bias (over-reporting of journey duration independent of journey length) revealed by the paired t-statistic analysis. It also shows a large random error at the individual journey

Table 3
Journey frequency, percentage of total and mean duration by mode and sedentary or active modes for eligible journeys for SenseCam ($n=1270$) and travel diary ($n=1328$).

	SenseCam (274 days)					Travel diary (278 days)				
	Journey frequency	%	Number of journeys per day	Mean duration of journey (min)	Standard deviation (min)	Journey frequency	%	Number of journeys per day	Mean duration of journey (min)	Standard deviation (min)
All	1270	–	4.64	15:40	15:03	1328	–	4.78	17:46	16:45
Car	338	26.6		14:54	11:40	378	28.5		16:01	11:33
Car (passenger)	95	7.5		17:06	12:37	91	6.9		17:59	14:00
Bus	40	3.1		18:28	15:55	52	3.9		29:41	25:05
Train	13	1.0		20:51	20:51	15	1.1		24:36	21:05
Boat	1	0.1		52:43	-	1	0.1		60:00	–
Walk	451	35.5		12:21	12:55	400	30.1		14:25	13:59
Cycle	332	26.1		19:50	19:28	391	29.4		20:52	20:46
Sedentary	487	38.3		15:52	12:32	537	40.0		18:00	14:49
Active	783	61.7		15:32	16:26	791	60.0		17:36	17:57

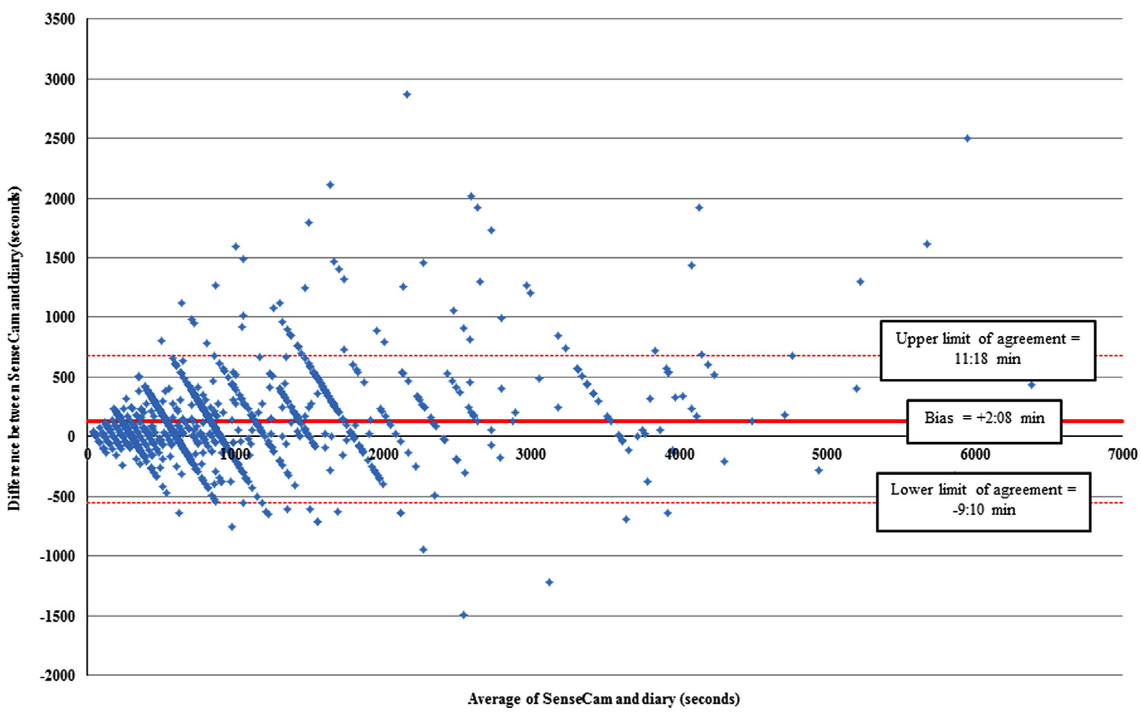

Fig. 1. Journey level Bland–Altman plot for all matched-pairs of self-reported and sensecam recorded journeys. Each point above the $y=0$ line indicates a journey that was over-reported in the diary and each point below the line indicates a journey that was under-reported in comparison to SenseCam recorded journey duration ($n=1127$).

level. The confidence intervals on the difference between means ($\pm$19.6 s) and on the limits of agreement ($\pm$35.0 s) were calculated as described by Bland and Altman (Bland and Altman, 1986, 2007) but not plotted for visual clarity. The apparent tendency for data points to lie on diagonal lines on the plot is a demonstration of the preference of participants to report durations to the nearest five minutes. The between-subject correlation coefficient was $r=0.953$. The within-subject correlation coefficient was $r=0.924$.

For 14 of the 1127 matched journeys, the travel diary reported a different mode to that shown by the SenseCam (1.2%). The most common misreporting was to record a "car passenger" journey as "car driver" ($n=5$). Using the Kappa statistic, agreement on mode was found to be $\kappa=0.98$ (95% CI=0.97–0.99).

3.2.1. *Effect of journey, participant and study characteristics*

We investigated differences in the journey level bias by conducting further paired-samples *t*-tests for variables hypothesised or previously shown to influence misreporting. We examined journey characteristics (mode, whether active or sedentary, duration, purpose, whether reported to nearest five minutes, whether journey was part of a trip-chain), participant characteristics (sex, age, educational status, whether memory jogger used), and study characteristics (location, day of week, day of study). Results are displayed in Table 4.

Walking and cycling journeys were over-reported to a greater extent than sedentary travel modes. Increasing journey duration suggested greater positive bias or over-reporting. Testing the data for heteroscedasticity was inconclusive; log-transformations were tested but made no significant difference and were therefore not used for these analyses. Journeys reported (assumed rounded) to five-minute intervals displayed greater positive bias than those not rounded. Use of the memory jogger reduced the over-reporting bias on journey duration. There were no obvious patterns for journey purpose; whether part of a trip-chain; for participant age, sex, or education status; study location; or for day of study.

3.2.2. Missed journeys

A total of 495 journeys recorded by SenseCam were not reported in the travel diary. SenseCam images revealed that 227 of these journeys (45.9%) were eligible (3–120 min duration) and should have been recorded in the travel diaries. Of the remaining journeys, 263 (53.1%) were of less than 3 min duration and five (1.0%) were greater than 120 min duration. Therefore, the travel diary recorded 1328 (85.4%) and missed 227 (14.6%) journeys from a possible total of 1555 that should have been reported. The most commonly missed eligible journeys by mode were walking (62.5%) followed by car driving (17.7%).

The SenseCam also missed journeys recorded by the travel diary. Across 278 travel diary days SenseCam missed 206 eligible journeys due to: "Camera off" ($n=99$), "Camera not worn" ($n=77$), "Upload fail or corrupted data" ($n=16$), or "Image too dark or obscured" ($n=14$). SenseCam also revealed 25 "phantom journeys" which appeared in the diary but the images revealed no such journey had taken place on that day. This suggests SenseCam should have recorded 1476 eligible journeys during the study period; the data show it recorded 1270 (86.0%) and missed 206 (14.0%).

Study aim 1 – key findings
At a journey level there is a small significant positive bias of 2:08 min (95% CI=1:48–2:28). This shows that on average journey durations are over-reported by 12.1% of reported time. Wide limits-of-agreement (−9:10 to 13:26 min) demonstrates the high random error of individual journey duration reporting.
Walking and cycling, greater journey duration, reporting to 5 min intervals (assumed rounded), and not using the memory jogger were all factors that lead to greater over-reporting bias. Other factors including sex and age appeared to have no obvious influence.
Analysis of missed journeys (identified during journey pair matching) showed both SenseCam and the travel diary recorded over 85% of the known journeys.

3.3. Study aim 2 – assessing the validity and reliability of self-reported daily summary estimates of travel

SenseCam missed no known journeys on 201 days (73.4% of the 274 days any SenseCam data was collected, and 65.5% of the 307 days the device was expected to collect data). For these days SenseCam was assumed to be a criterion measure of daily summary duration. Table 5 shows the average daily summary travel duration and frequency for (A) all SenseCam data (274 days), (B) the criterion SenseCam data (201 days), (C) all travel diary data (278 days) and (D) the days of travel diary data corresponding to the criterion SenseCam days (201 days). Journeys had to meet the 3–120 min duration criteria to be included in summary analysis.

Compared to the crude SenseCam data (A), the 201 criterion days of SenseCam data (B) had more journeys per day (0.36) as expected with incomplete days excluded. Mean journey duration was highly comparable (18 s difference) and mean summary travel time per day was greater by 4:15 min. In comparison to the crude SenseCam data (A), the crude travel diary data (C) over-estimated journeys per day (0.12) and journey duration (2:06 min). The overall effect was a disagreement in travel duration of 12:18 min per day (14.5% of reported time).

Compared to the criterion SenseCam data (B), the travel diary data from the 201 criterion days (D) under-estimated journeys per day (0.63) and over-estimated journey duration (2:37 min). The mean difference between criterion SenseCam data (B) and corresponding travel diary data (D) was 1:41 min (95% CI = −2:00 to 5:24 min) or 2.1% of reported time. A large standard deviation gave a wide 95% limits of agreement around this estimate of −50:29 to 53:41 min (Fig. 2). Negligible confidence intervals on the difference between means and the 95% limits of agreement were not plotted for visual clarity. Similar patterns were revealed when assessing daily active and sedentary travel durations (Table 5). There were greater differences between crude data (A vs. C) than criterion data (B vs. D).

Testing for the sensitivity of including the 1–3 min journeys ($n=263$) detected by SenseCam made a non-significant 30 s reduction to daily summary bias (results not displayed), suggesting these short journeys do not play an important role in total summary travel times. We also tested journey pairs included in the criterion days. These pairs displayed very similar durations and biases (using the paired-*t* statistic) as the journey pairs included in the overall journey level analysis (Study aim 1); this suggests the journeys included in criterion daily summary analysis were representative of the total dataset.

3.3.1. Effect of journey, participant and study characteristics

Results for the differences in the daily summary bias (using sex, age, educational status, whether memory jogger used, study location, day of week and day of study) and summaries for active and sedentary daily travel are displayed in Table 6. Due to large 95% confidence intervals there were few clear differences by these variables.

Study aim 2 – key findings
Analysis of criterion SenseCam data and corresponding travel diary data showed frequency was under-reported by 0.63 journeys per day. Combined with an over-reporting bias of 2:37 min per journey, bias on daily summary travel from these data was a small non-significant 1:41 min per day (95% CI = −2:00 to 5:24 min) or 2.1%. A large standard deviation gave a wide 95% limit of agreement around this estimate of −50:29 to 53:41 min.
Analysis of crude data showed frequency was over-reported by 0.14 journeys per day and duration by 2:06 min per journey; this combined to over-reporting in daily summary travel by 12:18 min per day (14.5%).

Table 4
Comparison of means for travel diary and SenseCam (paired-samples *t*-test) by journey, participant and study characteristics.

	Mean SenseCam time per journey (min)	**Mean diary time per journey** (min)	**Difference between means** (min)	**Difference as % of reported time**	**95% Confidence interval of difference** (min)	**95% Limits of agreement** (min)
Overall (*n*=1127)	15:31	17:39	2:08	12.1	1:48 to 2:28	−9:10 to 13:26
Mode						
Car (*n*=312)	14:42	16:19	1:36	9.8	0:57 to 2:16	−10:00 to 13:12
Car passenger (*n*=84)	17:08	18:18	1:09	6.3	−0:08 to 2:27	−10:33 to 12:51
Bus (*n*=34)	19:31	20:18	0:47	3.9	−0:51 to 2:26	−8:30 to 10:04
Train (*n*=13)	20:51	19:04	−1:47	−9.4	−6:17 to 2:42	−12:47 to 16:21
Boat (*n*=1)	52:43	60:00	7:17	12.1	–	–
Walk (*n*=358)	11:27	13:52	2:24	17.3	1:45 to 3:04	−9:53 to 14:41
Cycle (*n*=325)	19:36	22:28	2:52	12.8	2:21 to 3:24	−6:32 to 12:17
Active or sedentary						
Sedentary (*n*=444)	15:48	17:10	1:22	8.0	0:49 to 1:55	−10:12 to 12:56
Active (*n*=683)	15:19	17:57	2:38	14.7	2:12 to 3:03	8:23–13:39
Purpose						
Commute (*n*=380)	18:40	20:56	2:16	10.8	1:46 to 2:45	−7:12 to 11:44
Work related (*n*=192)	9:34	11:33	2:18	19.9	1:28 to 3:09	−9:14 to 13:50
Shopping (*n*=181)	11:52	13:08	1:15	9.5	0:42 to 1:49	−6:18 to 8:48
Leisure (*n*=283)	17:48	20:24	2:36	12.7	1:43 to 3:28	−12:06 to 17:18
Visit friends (*n*=37)	14:35	15:04	0:29	3.2	−0:30 to 1:29	−5:22 to 6:20
Other (*n*=54)	16:28	18:47	2:19	12.3	0:21 to 4:17	−11:48 to 16:26
Duration						
0–5 min (*n*=240)	4:07	4:16	0:08	3.1	−0:07 to 0:24	−3:53 to 4:09
6–10 min (*n*=295)	8:22	9:12	0:50	9.1	0:29 to 1:11	−5:07 to 6:47
11–15 min (*n*=228)	12:44	14:15	1:31	10.6	0:59 to 2:02	−6:14 to 9:16
16–30 min (*n*=233)	21:08	24:36	3:27	14.0	2:38 to 4:16	−9:00 to 15:54
31–120 min (*n*=131)	47:16	54:46	7:29	13.7	5:37 to 9:21	−13:37 to 28:35
Digit-preference						
Rounded to five min (*n*=831)	16:46	19:20	2:34	13.3	2:08 to 2 59	−9:33 to 14:41
Not rounded (*n*=296)	12:56	12:00	0:56	7.8	0:28 to 1:24	−7:08 to 9:00
Part of chain						
Chained (*n*=398)	15:32	17:58	2:28	13.7	1:43 to 3:08	−11:37 to 16:33
Not chained (*n*=729)	15:30	17:29	1:58	11.2	1:37 to 2:19	−7:28 to 11:24
Sex						
Male (*n*=626)	15:20	17:38	2:17	12.9	1:50 to 2:45	−9:09 to 13:43
Female (*n*=501)	15:44	17:41	1:57	11.0	1:27 to 2:27	−9:09 to 13:03
Age						
18–30 years (*n*=523)	14:53	16:58	2:04	12.2	1:40 to 2:29	−7:22 to 11:30
31–40 years (*n*=240)	14:42	16:23	1:40	10.2	0:58 to 2:21	−8:57 to 12:17
41–50 years (*n*=170)	14:34	16:32	1:58	11.9	1:11 to 2:44	−8:04 to 12:00
51–60 years (*n*=194)	19:01	22:04	3:02	13.7	1:51 to 4:13	−13:26 to 19:30
Education						
Completed high school (*n*=133)	15:45	19:08	3:23	17.7	2:24 to 4:22	−7:51 to 14:37
Current undergraduate student (*n*=226)	13:05	14:16	1:10	8.2	0:42 to 1:39	−5:57 to 8:17
Completed undergraduate degree (*n*=269)	18:14	21:16	3:02	14.3	2:10 to 3:54	−10:49 to 17:29
Completed graduate degree (*n*=499)	15:05	16:50	1:45	10.4	1:16 to 2:15	−9:08 to 12:38
Memory jogger use						
Yes (*n*=233)	15:40	17:00	1:19	7.7	0:48 to 1:50	−6:29 to 9:07
Sometimes (*n*=142)	17:35	20:20	2:44	13.4	1:22 to 4:07	−13:44 to 18:32
No (*n*=724)	14:55	17:12	2:16	13.2	1:52 to 2:41	−8:48 to 13:20
Do not know (*n*=28)	19:06	21:21	2:14	10.5	0:09 to 4:20	−8:21 to 12:49
Location						
Auckland (*n*=151)	15:11	16:24	1:13	7.4	0:19 to 2:06	−9:42 to 12:08
San Diego (*n*=648)	16:28	18:53	2:25	12.8	1:57 to 2:52	−9:17 to 14:07
Oxford (*n*=277)	14:03	16:11	2:07	13.1	1:28 to 2:47	−8:46 to 13:00
Romford (*n*=51)	12:20	13:48	1:28	10.6	0:13 to 2:42	−8:53 to 11:49
Study day						
Day 0 (*n*=142)	15:04	17:24	2:20	13.4	1:33 to 3:07	−6:53 to 11:33
Day 1 (*n*=368)	14:39	17:13	2:33	14.8	1:55 to 3:12	−9:44 to 14:50
Day 2 (*n*=305)	17:26	19:12	1:46	9.2	1:04 to 2:27	−10:13 to 13:45
Day 3 (*n*=285)	14:28	16:23	1:55	11.7	1:18 to 2:32	−8:24 to 12:14
Day 4 (*n*=27)	18:45	20:38	1:53	9.1	0:11 to 3:36	−6:35 to 10:21

Table 4 (*continued*)

	Mean SenseCam time per journey (min)	Mean diary time per journey (min)	Difference between means (min)	Difference as % of reported time	95% Confidence interval of difference (min)	95% Limits of agreement (min)
Week day						
Weekday (n=954)	15:08	17:14	2:06	12.2	1:45 to 2:27	−8:43 to 12:55
Weekend (n=173)	17:37	19:37	2:19	11.8	1:16 to 3:22	−11:24 to 16:02

Table 5
Daily summary travel time analyses for all journeys 3 – 120 min.

	All (crude) SenseCam data	SenseCam criterion data	All (crude) travel diary data	Travel diary data corresponding to criterion SenseCam data days
Days of data (n)	274	201	278	201
Total journeys (n)	1270	1005	1328	878
Journeys per day (n; SD)	4.64 (2.48)	5.00 (2.56)	4.78 (2.55)	4.37 (2.26)
Mean journey duration (min; SD)	15:40 (15:03)	15:22 (15:31)	17:46 (16:45)	17:59 (16:56)
Mean daily summary travel time (min; SD)	72:35 (52:02)	76:51 (52:46)	84:53 (56:10)	78:33 (52:45)
Active travel (walking and cycling)				
Journeys per day (n; SD)	2.86 (1.97)	3.30 (2.09)	2.84 (2.17)	2.78 (2.06)
Mean journey duration (min; SD)	15:32 (16:26)	14:57 (16:48)	17:36 (17:57)	18:16 (18:36)
Mean daily summary travel (min; SD)	44:23 (41:16)	49:22 (42:07)	50:02 (48:06)	50:47 (46:30)
Sedentary travel (motorised travel)				
Journeys per day (n; SD)	1.78 (2.12)	1.70 (2.17)	1.94 (2.31)	1.57 (2.02)
Mean journey duration (min; SD)	15:52 (12:32)	16:09 (12:41)	18:00 (14:49)	17:30 (13:30)
Mean daily summary travel (min; SD)	28:12 (36:36)	27:30 (37:12)	34:51 (44:10)	27:46 (38:42)

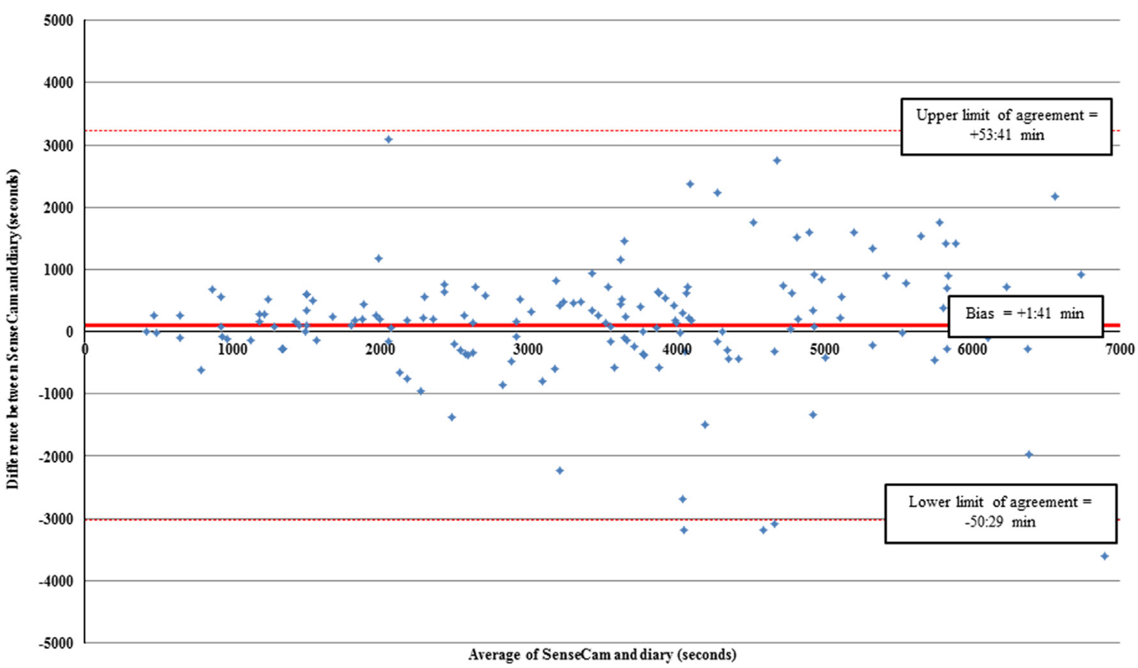

Fig. 2. Daily summary level Bland–Altman plot for days of self-reported and sensecam recorded travel. Each point above the y=0 line indicates a daily summary that was over-reported in the diary and each point below the line indicates a daily summary that was under-reported in comparison to SenseCam recorded travel (n=201).

4. Discussion

We aimed to investigate the validity and reliability of self-reported travel diary data in terms of systematic bias and random error by comparison to SenseCam image data. The key findings showed that at a journey level there was a significant over-reporting bias of 2:08 min (12.1% of reported time). At a daily summary duration level, self-report by travel diaries revealed a positive bias of 2.1% or 14.5% depending on analysis strategy employed. It is important to note that if the travel diary had been used in this population without the SenseCam, we would have taken the crude daily summary of travel duration (Column C; Table 5) as the group estimate. In this way the most appropriate assessment of bias may be through comparison of crude travel diary data (C) to either (A) crude SenseCam data

Table 6
Comparison of daily summary means for travel diary and SenseCam (paired-samples *t*-test) by participant and study characteristics.

	Difference between mean journey summary duration (min)	SD (min)	95% Confidence interval of difference (min)	**Difference between mean journey frequency**	SD (min)	95% Confidence interval of difference
Travel diary – SenseCam (*n*=201)	1:41	26:37	−2:00 to 5:24	−0.63	1.12	−0.79 to −0.48
Sex						
Male (*n*=84)	0:09	27:09	−5:09 to 4:50	−0.69	1.06	−0.88 to −0.50
Female (115)	4:13	25:48	−1:20 to 9:47	−0.55	1.19	−0.81 to −0.30
Age						
18–30 (*n*=92)	3:56	18:02	0:12 to 7:40	−0.62	1.09	−0.85 to −0.39
31–40 (*n*=43)	0:44	20:14	−5:29 to 6:58	−0.65	0.97	−0.95 to −0.35
41–50 (*n*=27)	−10:30	49:38	−30:08 to 9:08	−1.22	1.40	−1.78 to −0.67
51–60 (*n*=39)	5:53	25:59	−2:31 to 14:19	−0.23	0.96	−0.54 to 0.08
Education status						
Completed high school (*n*=83)	1:45	25:04	−3:43 to 7:13	−0.49	1.09	−0.73 to −0.26
Current undergraduate student (*n*=26)	12:38	18:00	5:21 to 19:55	−0.54	1.07	−0.97 to −1.1
Completed undergraduate degree (*n*=52)	3:18	18:51	−1:56 to 8:33	−0.69	1.09	−1.00 to −0.39
Completed graduate degree (*n*=40)	−7:37	38:19	−19:53 to 4:37	−0.90	1.22	−1.23 to −0.51
Study day						
Day 0 (*n*=33)	4:09	17:27	−2:10 to 10:20	−0.58	0.94	−0.91 to −0.24
Day 1 (*n*=58)	3:05	20:26	−2:16 to 8:28	−0.90	1.36	−1.25 to −0.54
Day 2 (*n*=48)	0:53	33:30	−8:50 to 10:36	−0.48	1.01	−0.77 to −0.19
Day 3 (*n*=50)	0:41	32:14	−8:27 to 9:51	−0.44	0.99	−0.72 to −0.16
Day 4 (*n*=12)	−4:23	18:17	−16:00 to 7:13	−0.92	1.00	−1.55 to −0.28
Location						
Auckland (*n*=20)	−8:08	27:44	−21:07 to 4:50	−0.60	1.31	−1.22 to 0.02
Oxford (*n*=45)	8:04	21:02	1:44 to 14:23	−0.69	1.10	−1.02 to −0.36
Romford (*n*=10)	8:14	17:14	−4:05 to 20:34	−0.40	1.43	−1.42 to 0.62
San Diego (*n*=126)	0:27	28:22	−4:32 to 5:27	−0.64	1.07	−0.82 to −0.45
Memory jogger use						
Do not know (*n*=5)	−30:28	47:34	−1:29:33 to 28:36	−2.20	1.92	−4.59 to 0.19
No (*n*=125)	−0:31	27:36	−5:24 to 4:21	−0.80	1.19	−1.01 to −0.59
Partial (*n*=27)	12:40	25:55	2:25 to 22:56	−0.44	0.58	−0.67 to −0.22
Yes (*n*=44)	4:55	16:06	0:02 to 9:49	−0.09	0.68	−0.30 to 0.12

(12:18 min; 14.5%) or (B) criterion SenseCam data (8:02 min; 9.5%). We recommend considering the over-reporting bias on duration to be within this range.

For many epidemiological studies this may be an acceptable range for validity. However, if daily estimates were extrapolated to months or years, rather than three to four days, the compounded error could greatly overestimate travel times. These findings allow researchers to assess whether self-reported journey and daily summary durations are valid for their own study aims. Corrections according to our reported bias could be considered in future work, including differential corrections for active and sedentary travel if appropriate.

For both journey and daily summary duration assessment, very wide limits-of-agreement were apparent. This suggests large random error or uncertainty in individual journey and daily summary measurements and therefore low reliability of self-reported travel behaviour. This suggests the potential for substantial measurement error in individual journey and daily summary measurements, and therefore that such data come with considerable uncertainty.

Agreement on mode was demonstrated for over 98% of the pairs of matched journeys. The 29 phantom journeys in the travel diary were excluded from this analysis and could be considered as misreported modes. Without a comparison measure (such as SenseCam) they would wrongly be included in analysis of self-report data. However, even when included, this still yields agreement for over 95% of journeys. These results suggest self-report of travel modes is valid, and researchers can have confidence in using such an approach (e.g., studies of modal share).

4.1. Comparison to literature

Our previous pilot study is the only comparison of adult self-reported diary, and image assessed, journey durations (Kelly et al., 2011). Our present findings with a much larger sample confirm the small positive and significant bias and wide limits of agreement on journey duration. In a second pilot study we found no such bias, but this was in adolescents and used a daily researcher administered questionnaire of school travel rather than a travel diary (Kelly et al., 2012). This is the first study using images to assess daily summary travel durations.

We previously conducted a systematic review of GPS comparisons to self-reported travel (Kelly et al., 2013a). The studies suggested a range of 1.2–13.8 min for over-reporting of journey durations. However, the eight included studies were limited by uncertainty in the GPS comparative measurements (e.g. data loss, limited journey identification algorithms, etc.) and there was no consensus on the magnitude of over-reporting. One of the included studies used the NTS as their comparative measure and reported much larger positive bias (7:00 min per journey; 46.7%) than in this study (Department for Transport, 2009).

The agreement on mode was exceptionally high and is higher than reported in previous GPS studies (Bricka et al., 2009; Stopher et al., 2007). This high agreement may be a result of only assessing matched trips. However, it is the objective nature of SenseCam image data that allows this matching process which is much harder when processing GPS data (Kerr et al., 2011; Oliver et al., 2010).

4.2. Possible explanations and mechanisms

Rounding of reported times to "5 min" or "10 min" intervals is known as digit preference; rounding usually occurs "up" rather than "down" and has been suggested as a major contributor to the small systematic bias at a journey level (Rietveld, 2001). It is also possible that participants include travel related activities (transitory activities) in their estimates.

The between-subject and within-subject correlation coefficients for journey duration were both very high. This tells us that longer travel diary journey durations were strongly associated with longer SenseCam recorded journey durations (and vice versa). This analysis is also an indication of whether the variance on the differences is clustered within certain participants. Two strong correlation coefficients suggest that it is not. If it was, then frequent travellers with specific reporting behaviours could influence the results if they report with greater or lesser error than infrequent travellers. As such the reporting errors appear to be randomly distributed across the population. Additionally there were no obvious age, education or sex effects.

4.3. Implications

Group level validity for diary-derived self-reported travel modes and durations mean they can be considered suitable for monitoring and surveillance, and for studies requiring assessment of group or population means. Corrections to population data could be considered according to our reported biases.

However, the very wide limits-of-agreement demonstrate that self-report of journey duration or daily travel exposure has a large random error. This implies that self-report may be subject to unacceptably high levels of error for assessing exposure at the level of the individual level, for categorising individuals to an exposure category, or for detecting individual behaviour change. Considering behaviour change, the size of the limits-of-agreement suggests the random error is in excess of plausible changes in active travel behaviour. To be suitable for these purposes multiple measurements on the same individual should be taken. Alternatively, objective measures such as GPS or pedometers could be considered.

These interpretations agree with recommendations made by Dollman et al., that self-report may be more suitable to population monitoring of physical activity and descriptive and needs assessment based research than for intervention and individual based research (Dollman et al., 2009). Matthews et al. stated that when estimating population values (e.g. of active travel) higher levels of random error at the individual level may be acceptable; with a large enough sample these errors do not bias population averages (Matthews et al., 2012). The authors go on to say that large random error is detrimental to the internal validity of association studies and in particular dose response studies and "lead to misclassification of exposure, loss of statistical power and attenuation of effect sizes" (Matthews et al., 2012). It has also been reported that understanding the magnitude of random error (referred to by many authors as precision) may allow for better estimates of required sample size and data collection periods (Stopher et al., 2008). This allows the avoidance of unnecessary participant burden and survey costs.

4.4. Strengths and limitations

We were able to recruit the desired number of volunteers, with low numbers unwilling to consent or dropping out of the data collection protocol. Among the final 84 participants compliance was high in terms of days of data returned. Analysis of missed journeys (identified during journey pair matching) showed both SenseCam and the travel diary recorded over 85% of the known journeys.

There have been calls for technological advances in physical activity assessment which will yield methods with increased validity (Medical Research Council) – this study is a direct response to these calls. Previous studies have used accelerometer or GPS but we believe SenseCam images give a more valid measure of travel mode, frequency, and duration. The criterion validity of SenseCam compared to direct observation, inter-instrument reliability, and inter- and intra-rater reliability on coding has been previously demonstrated (Kelly, 2013, pp. 86–122; Kelly et al., 2011, 2012). However it should be noted that criterion validity was assessed in a semi-experimental setting rather than free-living and was conducted at a journey, rather than daily summary level. Inter-rater reliability for coding has also been shown on free-living journeys in teenagers (Kelly et al., 2012). This is also the largest study using wearable cameras in travel measurement to date.

We attempted to maximise internal validity and minimise potential sources of bias. In an attempt to minimise reporting reactivity we did not explicitly tell participants we were assessing how well they report travel behaviour. We also assessed this potential reactivity by comparing our data to publically available National Travel Survey (NTS) data and analysing results by study day for evidence of habituation to the device and reducing reactivity. We found the travel diary data was similar to existing public datasets; these data are reported in detail elsewhere (Kelly, 2013, pp. 167–170). It is worth noting that participants completing the diary may have already been subject to the reactivity from the travel diary itself. Therefore, introducing the camera might not have any substantial extra effect on awareness of journey behaviour.

We also attempted to maximise external validity. Age, sex, ethnic origin, socio-economic status of experimental samples may all limit generalisability (Medical Research Council). We conducted data collection in different locations to give a range of participants, travel environments, and travel behaviours. However, the demographics of the participants could not be considered representative of the wider populations from which they were recruited; the sample was relatively young and highly educated; a common problem in such travel research (Dalton et al., 2013). Whether these issues were a function of the university setting where recruitment was active or is representative of the sort of people willing to wear a camera is unclear. We suggest it was both. Recruitment also resulted in a high number of cyclists. All participants in this study were able-bodied and therefore we do not have information on feasibility of use with disabled individuals. Neither did we collect any data on ethnicity of participants.

Foster et al. (2011) have previously shown that the recruiting technique employed can influence the study sample. In Romford, where we were able to recruit a group not associated with a University setting (and with a different educational status), the compliance was lower. However, the camera compliance was substantially higher than the diary compliance, suggesting this is a limitation relevant to survey methods, rather than specific to camera methods. The NTS 2010 Technical Report found similar differences between response rates in outer London and the rest of the country; the overall NTS response rate in 2010 was 60% but this was lower in Inner London (48%) and Outer London (55%), and higher in the rest of the country (61%) (Rofique et al., 2011).

There is also a question as to whether people who volunteer to wear SenseCam are representative (and have representative reporting behaviour) of the usual survey populations whose reporting we are trying to investigate. It should be noted that there was a low proportion of people refusing to take part once they were informed about the study, suggesting bias from this reason may not be large. However, the generalisability is limited to the population we approached rather than the NTS population. Perhaps the similarity in diary reporting between our population and NTS populations (mentioned above) is a better indication of the generalisability of our study population to previous NTS populations. However, there were low numbers of public transport journeys due to lower recruitment followed by the lower compliance in Romford (UK) where high bus and train use was anticipated. It is unclear why SenseCam missed a higher proportion of bus journeys than any other mode and whether this biased the results.

There are certain limitations to SenseCam as a device for measurement. These include participant burden, forgetting to wear the device, obscured or low light images, possible reactivity, and substantial manual data coding time. This suggests that currently the device is suited to small-medium size studies, for example validating or calibrating measures that can be used on larger samples as in this study.

This study only considers certain aspects of validity and reliability. For example we did not investigate or discuss such aspects as test–retest or inter-instrument reliability, or face validity of the diary. We used an "NTS-like" diary, but this did not follow the exact protocol which only focuses on walking on day 7 of 7 and excludes travel on routes motor vehicles cannot use, thereby underestimating walking and cycling. Therefore our study represents the validity and reliability of a typical diary rather than the strict NTS method. Finally, this study did not investigate possible associations between journey variables. For example, longer journeys and rounded journeys both indicated greater bias; however, we did not test if longer journeys were more likely to be rounded.

4.5. Future study

We make the following recommendations for future research,

- Repeat the experiment with a more varied sample to investigate the feasibility of using in wider population groups, and the bias on reporting in these groups. This might include low- and middle-income countries, though the cost of the device and resources available may prohibit this. For example, active travel is a particularly important intervention in "Ciclovia" in Bogota, Columbia, and valid measurement could be particularly instructive (Sarmiento et al., 2010; Torres et al., 2013). The device was recently trialled in Kingston, Jamaica.
- Repeat the study with other self-report measures of travel and active travel commonly used in public health research, for example habitual measures such as the International Physical Activity Questionnaire (Craig et al., 2003) or the EPIC Physical Activity Questionnaire (Wareham et al., 2002).
- Conduct a similar study comparing SenseCam to other devices such as accelerometers or GPS to validate and calibrate their algorithms for detecting and measuring journeys.

5. Conclusions

Valid and reliable measurement is fundamental to epidemiological study of physical activity and health. Self-reported travel diary data is a relatively valid measure of journey level and daily summary travel duration at the group level. However, it is an unreliable measure at the level of the individual journey, day, or participant. These findings should direct the study designs where self-reported travel diary data can be used with confidence.

Funding

This work was supported by Microsoft Research through their Ph.D. Scholarship Programme and the British Heart Foundation.

Acknowledgements

We would like to thank the following: Professor Alan Batterham (Teesside University) for statistical support; and Dr. Steve Hodges and Dr. Emma Woodberry (Microsoft Research, Cambridge) for device support.

References

Armstrong, N., Welsman, J.R., 2006. The physical activity patterns of European youth with reference to methods of assessment. Sports Med. 36, 1067–1086.
Bland, J.M., Altman, D.G., 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1, 307–310.
Bland, J.M., Altman, D.G., 1995a. Calculating correlation coefficients with repeated observations: Part 1 – correlation within subjects. Br. Med. J. 310, 446.
Bland, J.M., Altman, D.G., 1995b. Calculating correlation coefficients with repeated observations: Part 2 – Correlation between subjects. Br. Med. J. 310, 633.
Bland, J.M., Altman, D.G., 2007. Agreement between methods of measurement with multiple observations per individual. J. Biopharm. Stat. 17, 571–582.

Bricka, S., Zmud, J., Wolf, J., Freedman, J., 2009. Household travel surveys with GPS an experiment. Transportation Research Record 2105, 51–56, http://dx.doi.org/10.3141/2105-07.
Bull, F.C., Gauvin, L., Bauman, A., Shilton, T., Kohl 3rd, H.W., Salmon, A., 2010. The Toronto Charter for physical activity: a global call for action. J. Phys. Act. Health 7, 421–422.
Carmines, E.G., Zeller, R.A., 1979. Reliability and Validity Assessment. SAGE Publications, California, USA.
Craig, C.L., Marshall, A.L., Sjostrom, M., Bauman, A.E., Booth, M.L., Ainsworth, B.E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J.F., Oja, P., 2003. International physical activity questionnaire: 12-country reliability and validity. Med. Sci. Sports Exerc. 35, 1381–1395.
Dalton, A.M., Jones, A.P., Panter, J.R., Ogilvie, D., 2013. Neighbourhood, route and workplace-related environmental characteristics predict adults' mode of travel to work. PLoS One 8, e67575.
Department for Transport, 2009. National Travel Survey GPS Feasibility Study Final Report, London.
Department for Transport, 2010. National Travel Survey 2009: Statistical Release, London.
Department of Health, 2004. At least five a week: evidence on the impact of physical activity and its relationship to health. A report from the Chief Medical Officer. In: Prevention, P.A.H.I.a. (Ed.), London.
Doherty, A.R., Hodges, S.E., King, A.C., Smeaton, A.F., Berry, E., Moulin, C.J.A., Lindley, S., Kelly, P., Foster, C., 2013a. Wearable cameras in health: the state of the art and future possibilities. Am. J. Prev. Med. 44, 320–323.
Doherty, A.R., Kelly, P., Kerr, J., Marshall, S., Oliver, M., Badland, H., Hamilton, A., Foster, C., 2013b. Using wearable cameras to categorise type and context of accelerometer-identified episodes of physical activity. Int. J. Behav. Nutr. Phys. Act. 10, 22.
Doherty, A.R., Moulin, C.J., Smeaton, A.F., 2011. Automatically assisting human memory: a SenseCam browser. Memory 19, 785–795.
Dollman, J., Okely, A.D., Hardy, L., Timperio, A., Salmon, J., Hills, A.P., 2009. A hitchhiker's guide to assessing young people's physical activity: deciding what method to use. J. Sci. Med. Sport/Sports Med. Aust. 12, 518–525.
Florez Pregonero, A., Gomez, L.F., Parra, D.C., Cohen, D.D., Arango Paternina, C.M., Lobelo, F., 2012. Time spent traveling in motor vehicles and its association with overweight and abdominal obesity in Colombian adults who do not own a car. Prev. Med. 54, 402–404.
Foster, C.E., Brennan, G., Matthews, A., McAdam, C., Fitzsimons, C., Mutrie, N., 2011. Recruiting participants to walking intervention studies: a systematic review. Int. J. Behav. Nutr. Phys. Act. 8, 137.
Gemming, L., Doherty, A., Kelly, P., Utter, J., Ni Mhurchu, C., 2013. Feasibility of a SenseCam-assisted 24-h recall to reduce under-reporting of energy intake. Eur. J. Clin. Nutr. 67, 1095–1099.
Hamer, M., Chida, Y., 2008. Walking and primary prevention: a meta-analysis of prospective cohort studies. Br. J. Sports Med. 42, 238–243.
Kelly, P., 2013. Assessing the Utility of Wearable Cameras in the Measurement of Walking and Cycling. University of Oxford.
Kelly, P., Doherty, A., Berry, E., Hodges, S., Batterham, A.M., Foster, C., 2011. Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study. Int. J. Behav. Nutr. Phys. Act. 8 (1), 44, http://dx.doi.org/10.1186/1479-5868-8-44.
Kelly, P., Doherty, A.R., Hamilton, A., Matthews, A., Batterham, A.M., Nelson, M., Foster, C., Cowburn, G., 2012. Evaluating the feasibility of measuring travel to school using a wearable camera. Am. J. Prev. Med. 43, 546–550.
Kelly, P., Krenn, P., Titze, S., Stopher, P., Foster, C., 2013a. Quantifying the difference between self-reported and global positioning systems-measured journey durations: a systematic review. Transp. Rev. 33, 443–459.
Kelly, P., Marshall, S.J., Badland, H., Kerr, J., Oliver, M., Doherty, A.R., Foster, C., 2013b. An ethical framework for automated, wearable cameras in health behavior research. Am. J. Prev. Med. 44, 314–319.
Kerr, J., Duncan, S., Schipperjin, J., 2011. Using global positioning systems in health research: a practical approach to data collection and processing. Am. J. Prev. Med. 41, 532–540.
Kerr, J., Marshall, S., Godbole, S., Chen, J., Legge, A., Doherty, A., Kelly, P., Oliver, M., Badland, H., Foster, C., 2013. Using the SenseCam to improve classifications of sedentary behavior in free-living settings. Am. J. Prev. Med. 44, 290–296.
Krizek, K.J., Handy, S.L., Forsyth, A., 2009. Explaining changes in walking and bicycling behavior: challenges for transportation research. Environ. Plan. B: Plan. Des. 36, 725–740.
Lubans, D.R., Boreham, C.A., Kelly, P., Foster, C.E., 2011. The relationship between active travel to school and health-related fitness in children and adolescents: a systematic review. Int. J. Behav. Nutr. Phys. Act. 8, 5.
Matthews, C.E., Hagstromer, M., Pober, D.M., Bowles, H.R., 2012. Best practices for using physical activity monitors in population-based research. Med. Sci. Sports Exerc. 44, S68–76.
Medical Research Council. Diet and physical activity measurement toolkit (accessed 07.07.12).
Ogilvie, D., Foster, C.E., Rothnie, H., Cavill, N., Hamilton, V., Fitzsimons, C.F., Mutrie, N., 2007. Interventions to promote walking: systematic review. Br. Med. J. 334, 1204.
Oja, P., Titze, S., Bauman, A., de Geus, B., Krenn, P., Reger-Nash, B., Kohlberger, T., 2011. Health benefits of cycling: a systematic review. Scand. J. Med. Sci. Sports 21, 496–509.
Oliver, M., Badland, H., Mavoa, S., Duncan, M.J., Duncan, S., 2010. Combining GPS, GIS, and accelerometry: methodological issues in the assessment of location and intensity of travel behaviors. J. Phys. Act. Health 7, 102–108.
Rietveld, P., 2001. Rounding of Arrival and Departure Times in Travel Surveys: an Interpretation in Terms of Scheduled Activities. Tinbergen Institute, Amsterdam, The Netherland.
Rofique, J., Humphrey, A., Pickering, K., Tipping, S., 2011. National Travel Survey 2010 Technical Report. Department of Transport, London.
Sallis, J.F., Owen, N., Fotheringham, M.J., 2000. Behavioral epidemiology: a systematic framework to classify phases of research on health promotion and disease prevention. Ann. Behav. Med. 22, 294–298.
Sarmiento, O., Torres, A., Jacoby, E., Pratt, M., Schmid, T.L., Stierling, G., 2010. The Ciclovia-Recreativa: a mass-recreational program with public health potential. J. Phys. Act. Health 7 (Suppl 2), S163–180.
Stopher, P., FitzGerald, C., Xu, M., 2007. Assessing the accuracy of the sydney household travel survey with GPS. Transportation 34, 723–741.
Stopher, P., Kockelman, K., Greaves, S., Clifford, E., 2008. Reducing burden and sample sizes in multiday household travel surveys. Transp. Res. Rec.: J. Transp. Res. Board 2064, 12–18.
Stopher, P.R., Greaves, S.P., 2007. Household travel surveys: where are we going? Transp. Res. A – Policy 41, 367–381.
Torres, A., Sarmiento, O.L., Stauber, C., Zarama, R., 2013. The Ciclovia and Cicloruta programs: promising interventions to promote physical activity and social capital in Bogota, Colombia. Am. J. Pub. Health 103, e23–30.
Wareham, N.J., Jakes, R.W., Rennie, K.L., Mitchell, J., Hennings, S., Day, N.E., 2002. Validity and repeatability of the EPIC-Norfolk Physical Activity Questionnaire. Int. J. Epidemiol. 31, 168–174.
Weisberg, H.F., 2009. The Total Survey Error Approach: A Guide to the New Science of Survey Research. University of Chicago Press.
Welk, G., 2002. Physical Activity Assessments for Health-Related Research. Human Kinetics, Champaign, IL.
Woodcock, J., Franco, O.H., Orsini, N., Roberts, I., 2011. Non-vigorous physical activity and all-cause mortality: systematic review and meta-analysis of cohort studies. Int. J. Epidemiol. 40, 121–138.
Woodcock, J., Givoni, M., Morgan, A.S., 2013. Health impact modelling of active travel visions for england and wales using anIntegrated Transport and Health Impact Modelling Tool (ITHIM). PLoS One 8, e51462.
World Health Organisation, 2007. A guide for population based approaches to increasing levels of physical activity: implementation of the WHO global strategy on diet. Phys. Act. Health, Geneva, http://www.who.int/dietphysicalactivity/PA-promotionguide-2007.pdf, ISBN 978 92 4 159517 9.
Yang, L., Sahlqvist, S., McMinn, A., Griffin, S.J., Ogilvie, D., 2010. Interventions to promote cycling: systematic review. Br. Med. J. 341, c5293.