University of Oxford

Doctoral Thesis

Assessing the utility of wearable cameras in the measurement of walking and cycling

---

Paul Kelly

St Cross College

A thesis submitted for the degree of Doctor of Philosophy

in the

Department of Public Health

Medical Sciences Division

Trinity 2013

Assessing the utility of wearable cameras in the measurement of walking and cycling

Paul Kelly, St Cross College

Submission for Doctor of Philosophy in Public Health

Trinity Term 2013

Abstract

Walking and cycling are considered very important behaviours in public health. This thesis is a study of the measurement of walking, cycling and overall travel behaviour.

In the first part I present a systematic review of studies comparing Global Positioning System (GPS) measured travel to self-report. I found 12 results from eight eligible studies. All studies showed self-reported journey times were greater than GPS measured times. The differences ranged from over-reporting of +2.2 to +13.5 minutes per journey.

In the second part of this thesis, I describe the development of a feasible, valid and reliable method for measuring travel with automated wearable cameras (SenseCam) through field testing and two pilot studies. I compared my new method to direct observation (considered a criterion measure) and found very good agreement and reliability (inter-rater, intra-rater and inter-measure). I also present an ethical framework for the measurement of health-related behaviours using automated wearable cameras.

In the third part of the thesis, I report findings from an experiment designed to assess a well known UK travel diary (The National Travel Survey). Across four locations (Oxford, UK; Romford, UK; San Diego, USA; and Auckland, New Zealand) I collected 3-4 days of SenseCam and travel diary data from n=84 participants (convenience sample). Compliance with the collection protocol was high and inspection of the crude results suggests relative agreement between measures. Analysis of matched pairs of measurements (n=1,127 journeys) suggests a significant positive bias on self-reported durations (2:08 minutes; 95% CI = 1:48 to 2:28; 95% limits-of-agreement = -9:10 to 13:26).

These results suggest self-reported journey exposure is valid at a population level, though corrections according to my reported bias could be considered. The large limits of agreement on duration estimates suggest self-report may be unsuitable for assessment of individual travel behaviour.

Contents

List of Tables, Figures and Text boxes

Acknowledgements

I would like to thank my supervisor Dr Charlie Foster for his help and guidance through the course of my DPhil. Quite simply this would not have happened without his support.

I have also been supervised by Dr Aiden Doherty who has been a great source of advice, feedback and support in many ways. Again I owe him huge thanks.

Steve and Emma took a chance on me in the early days and I am very grateful for their encouragement.

From my research group, I would like to thank Gill, Anne, Alex and Anja. I would like to say a special thanks to Nick and Justin for putting up with all my questions. Further afield, I would also like to thank Alan B, Alan S, Melody, Hannah, Jacqueline, Simon, Sylvia, Patricia, Marie, Elaine, Pekka and Amanda.

My parents have been incredibly supportive of my prolonged stay at University and I thank them deeply for the opportunity. I must also mention my sisters; Rachel, who has explained many difficult concepts to me; and Helen who has always been excited for the latest updates from my research efforts.

I am lucky to have some friends who have helped, encouraged, consoled and celebrated with me; JK, Ben, James, Iana, Alexandra W, Alexandra M, Andy, Colin, Jeremy and Toby.

Declaration of authorship

I have read the Oxford University regulations and declare this thesis is my own work. Any material presented from other sources is clearly indicated.

The idea to use SenseCam to investigate the error on self-reported travel behaviour was my own. I was working for my supervisor Dr Charlie Foster at the time the idea was conceived and he has been involved through every stage of developing and enacting the ideas and writing. Dr Aiden Doherty has co-supervised my thesis, and has also been involved in many aspects of this work.

I received expert assistance and advice from many different individuals at different stages of production of this thesis and the work presented within. Where appropriate, I state their role and involvement in this work.

Funding

I am very grateful to Microsoft Research who funded my studies through their PhD Scholarship Programme.

Word count

Main text: 46,500 words

Contents, Abstracts, Acknowledgements, Glossary, Overview, etc.: 4,500 words

References: 5,700 words

Tables and Figures: 7,000 words

Abbreviations

ADNFS	Allied Dunbar National Fitness Survey
AUT	Auckland University of Technology
DH	Department of Health
DfT	Department for Transport
GPS	Global Positioning System(s)
HIC	High income country
HR	Heart rate
HSE	Health Survey for England
ICC	Intra-class Correlation Coefficient
IRB	Institutional Review Board
LMICs	Low- and middle- income countries
MVPA	Moderate to vigorous physical activity
NICE	National Institute for Health and Clinical Excellence
NTS	National Travel Survey
QALY	Quality Adjusted Life Year
UCSD	University of California, San Diego
WHO	World Health Organisation

Durations

In this thesis I often report time durations. I use various units as needed and have set some rules for consistency. As an example, 236 seconds could be reported in one of the following ways:

- 236 seconds
- 3:56 minutes
- 3.93 mins

Glossary

Active travel	Travel that can also be considered physical activity. Most commonly walking and cycling, but also skateboarding, roller-blading and other similar behaviours
Automated image capture	Recording of images with a wearable camera that happens automatically without any user action
Broken journey or trip	A journey or trip that is paused without the purpose changing
Chained journey or trip	A journey or trip that is made of two or more independent stages
Digit preference	Propensity for survey and diary respondents to report durations to the nearest 5 or 10 minutes. Commonly called rounding
Digital life logging	Automated digital capture of everyday life activities through first-person point-of-view images
Global Positioning System (GPS)	Used to record location. Data can be converted into travel behaviour data
Institutional Review Board	Sometimes called Ethics Committee. Assesses and approves research applications for adherence to ethical codes of conduct
Journey or trip	A one-way course of travel having a single main purpose. May be comprised of individual components called stages. E.g. a walk-bus-walk to work would be defined as a single journey of three stages
Sedentary travel	Travel by motorised modes, such as car, bus, train, boat or tram, and not considered physical activity
Transitory activities	Activities before, after or during journeys, associated with the journey but not travel (e.g., waiting for public transport, or transitioning between modes).
Travel diary	Also record or log. Usually recall for a certain period of time
Travel survey	Research instrument into travel behaviour. Usually contains a questionnaire element for demographics and a questionnaire or diary element for travel activities
Travel questionnaire	Can be self- or researcher- administered. Usually collects information on usual or habitual behaviour
Journey or trip stage	The constituent parts of a journey or trip. A one-way course of travel having a single main purpose and mode. E.g. a walk-bus-walk to work would be defined as three stages

Overview

This thesis describes an investigation of SenseCam in the measurement of travel, with a focus on walking and cycling. SenseCam is an automated wearable camera. It takes a photograph every 10-20 seconds. By the end of a day it will have taken 2000-3000 photographs of the daily life of the wearer. I have tried to use these photographs to understand the travel behaviour of the wearer.

Specifically I have assessed whether these images can inform about the error on self-reported travel behaviour. This is an important question as much of the existing evidence for the public health benefits of walking and cycling is based on self-reported measurement.

In 2009 I was working as a researcher in the British Heart Foundation Health Promotion Research Group at Oxford University. I was working on a project investigating national approaches to promoting physical activity in children, funded by WHO Europe. I was also involved in various active travel projects for my then line-manager Dr Charlie Foster. I had formed an interest in measurement of physical activity the previous year during my MSc thesis. I conducted a secondary analysis of HSE data, investigating possible associations between physical activity and diet. I was intrigued by the idea that we don't just know the amount of exercise (or eating) someone is doing.

In July 2009 Charlie sent me a Youtube link of some computer scientists wearing SenseCam around Cambridge. You see them cycling, walking, visiting shops, drinking coffee, etc. We instantly thought all problems with physical activity assessment were solved for ever more. Case closed. We contacted Microsoft Research who had developed SenseCam and persuaded them to lend us two devices. I set about collecting the evidence that would conclusively end all debate about physical activity assessment. I would do the same for diet and sedentary behaviour a few months later I assumed.

Of course I was wrong and naive in equal measure. I have spent the last four years learning about measurement error, measures of agreement, feasibility, pilot testing, validity, reliability, coding protocols, accuracy, precision, uncertainty, sample sizes, GPS measurement, systematic reviews, ethical frameworks, recruitment, consent, missing data and bias. I ended up looking at one or two dimensions of one domain of physical activity behaviour. I conclude that I may have shown something about self-reported travel duration, certain to some assumptions and limitations. I have certainly learned how complex the issues surrounding measurement of physical activity can be.

However, I remain positive that active travel is a very important direction for public health in the coming years. I also believe that improvements in our ability to measure it will only strengthen our efforts to increase it. I hope to continue trying to make a contribution to this area.

PK.

Part 1 – The literature

CHAPTER 1 - LITERATURE REVIEW

Summary

My first chapter outlines the importance of active travel, specifically walking and cycling, to public health. I discuss the role of valid and reliable measurement when considering walking and cycling evidence, and the strengths and weaknesses of current methods.

Self-report is widely used to measure travel behaviour. I review the National Travel Survey (NTS), one of the UK's primary data sources on travel behaviour and the way the data is used in Public health. As a self-report method the NTS is subject to the same challenges of validity and reliability as any measure that relies on participant responses. I discuss what is currently known about the error on the NTS and other self-report measures.

I introduce SenseCam, a passive automated digital camera, and describe its possible role in this field. This thesis is ultimately an investigation into the potential utility of this device in active travel measurement. Finally, I state my thesis objectives and proposed work plan.

Physical activity and Public health

Physical activity in children, adolescents, adults and the elderly contributes to short and long term physical, social, emotional and psychological health and well-being [1-5]. It is associated with reduced morbidity and mortality, and the specific health benefits of regular physical activity include reduced risk of cardiovascular disease, coronary heart disease and stroke, type II diabetes, hypertension, colon cancer, breast cancer, obesity, dementia and depression [1-5]. The prevalence and impact of physical inactivity is independently emerging as one of the largest public health problems [4, 5]. The Global Burden of Disease Study 2010 ranked physical inactivity as the tenth leading preventable risk factor for global mortality [6].

Physical activities take place in diverse domains[1] in the life-course including work, domestic, leisure, and the focus for this thesis, transport [3]. Transport related physical activity usually refers to walking and cycling and is termed active travel[2]. It is worth noting that walking and cycling can also be leisure activities, conducted for sport or recreation, rather than transport purposes [7].

Walking and cycling are important forms of physical activity to the field of public health [1, 2]. They have the potential to confer substantial health benefits as well as environmental and congestion benefits when they replace vehicle travel [8, 9]. Walking and cycling provide an opportunity to accumulate health enhancing physical activity which can contribute to meeting physical activity recommendations [1-3, 10, 11]. In contrast, travel by motor vehicle is sedentary and carbon emitting [12, 13]. In the UK, data from the 2002 General Household Survey showed that walking and cycling were the first and fourth most common physical activities respectively [14]; however, Ogilvie et al., reported that data from the National Travel Survey (NTS) showed a 15% reduction in walking and cycling trips between 1996 and 2001 [10]. More recent NTS data from the Department for Transport has shown the average time spent walking or cycling decreased from 12.9 minutes to 11 minutes between 1995/7 and 2007 [7].

Travel behaviour features within many conceptual models and frameworks for health. For example it appears in reference to transport in the WHO model for the Social Determinants of Health [15] and within individual activity and physical activity environment on the Foresight obesity map [16].

Walking

Many studies have shown the health benefits of regular walking [17-21]. Two recent meta-analyses demonstrated regular walking is significantly associated with reduced risk for all-cause mortality [22, 23]. In terms of public health it is considered an important form of activity because of an

---

[1] As well as domain classification, physical activity can be classified in other ways such as by behaviour (e.g. walking, housework, gardening, football), by intensity (e.g. light, moderate or vigorous (possibly using MET values) depending on the physiological response), by frequency (e.g. daily or monthly bouts), and by volume or duration (e.g. energy expenditure or total time).

[2] It could also include activities such as skate boarding, roller blading, kayaking, running, or any physical activity for which the purpose includes travel or transport

unwillingness or inability of a large proportion of the population to participate in more vigorous activities [24]. It has been described as the safest, most convenient form of physical activity as it is low-impact, low cost, and readily accessible, requiring no special skills or equipment [8, 25, 26]. Furthermore, walking can be easily assimilated into daily life and continued into old age [8, 27]. Systematic reviews of exercise interventions report walking to have the greatest potential for increasing activity levels in sedentary individuals [28, 29]. In their influential walking paper, Jerry Morris and Adrian Hardman described walking as 'the perfect form of physical activity' [26] and as a behaviour it now features in many national and international physical activity and disease prevention strategies [1, 3, 7, 10].

In terms of trends, analysis of Health Survey for England (HSE) data suggest walking levels in the UK increased between 1999 and 2004 (from 50 to 60 min per week for men and from 38 to 50 min per week in women) [30]. It should be noted this was assessed by number of days on which a minimum level was attained, rather than an actual record of walking. This might explain the contradictory finding to the previously cited NTS data which reported a decrease in walking from 1995/7 to 2007 [7]. In a recent analysis of the 2008 Health Survey for England data, Belanger et al., showed walking (for all purposes) was the most important contributor to the total amount of moderate to vigorous physical activity (MVPA) for both men (26-42% of total MVPA across all age ranges) and women (37-45%) [31].

One of my first thesis tasks was to estimate the potential health impact of increased walking at a population level. I used data from the Allied Dunbar National Fitness Survey (ADNFS) in which participants completed a range of fitness and biometric tests [32]. By combining cardiovascular data from 1,741 individuals (aged 18-64) that walked one mile on a treadmill at 3mph with English population statistics I estimated that 1.5 million men (95% C.I. 0.9-2.2 million) (from 13.4 million corresponding to 11.6% (95% C.I. 7.0-16.2%)) and 3.9 million women (95% C.I. 3.0-4.8 million) (from 13.6 million corresponding to 28.6% (95% C.I. 22.0-35.1%)) in England aged 25-64 years were so unfit they could receive health benefits from regularly walking at a relatively slow speed of 3mph. This was assessed by the proportion of participants that exceeded 70% heart rate (HR) max, taken as the threshold for substantial fitness gains. These data are displayed in Table 1.1, Figure 1.1 and Figure 1.2. In total, a projected 5.4 million individuals (95% C.I. 3.9-6.9 million) aged 25-64 (from 27.0 million corresponding to 20.1% (95% C.I. 14.6-25.7%)) could receive health benefits from walking at 3mph [33]. This is a very conservative estimate as self-selected walking speeds are often higher than 3 mph [34], and it only considers those who receive substantial fitness and health gains. This estimate was published in the Journal of Sports Sciences [33] and the full paper is presented in Appendix 1.

This estimate has some fairly significant limitations. The ADNFS was collected in 1992 and the population statistics were collected in 2009. The trends in physical activity patterns discussed earlier in this chapter suggest population fitness levels may have changed during this 17 year gap. The sample for the fitness testing may also have been subject to bias and not be representative. Additionally the estimate does not specify the dose or frequency of walking required for health gains. However, it serves to show the potential population impact of walking, especially in light of the conservative assumptions (pace and % HR max). It also suggests the potential population impact may be greater in women.

Table 1.1 Estimated numbers of individuals in England who would attain vigorous intensity activity by walking 1 mile at 3mph (reproduced from Kelly et al., 2011)

	Age	**Number in English population (x1000)**[1]	**Percentage exceeding 70%HR max walking @ 3mph (4.8kmh)**[2] **(with 95% C.I.)**	**Number getting training effect (x1000) (with 95% C.I.)**	**Percentage of population getting training effect (with 95% C.I.)**
Men	25-35	3,352.7	2 (0-4.4)	67.1 (0-147.5)	
	35-44	3,875.1	8 (4.2-11.8)	310.0 (162.8-457.3)	
	45-54	3,230.6	9 (5.0-13.0)	290.8 (161.5-420.0)	
	55-64	2,928.6	30 (21.1-38.9)	878.6 (617.9-1,139.2)	
		13,387.0		**1,546.5** (942.2-2,164.0)	**11.6** (7.0-16.2)
Women	25-35	3,355.5	12 (7.6-16.4)	402.7 (255.0-550.3)	
	35-44	3,917.5	19 (13.7-24.3)	754.6 (536.7-952.0)	
	45-54	3,285.7	38 (31.3-45.7)	1,248.6 (995.6-1,501.6)	
	55-64	3,028.8	49 (39.7-58.3)	1,484.1 (1,202.4-1,765.8)	
		13,587.5		**3,890.0** (2,989.7-4,769.7)	**28.6** (22.0-35.1)

1. Population estimates from http://www.statistics.gov.uk.
2. Direct measure of heart rate for adults taken from Morris and Hardman Walking to Health paper, p324, Br J Sports Med, May 1997. They referenced Allied Dunbar National Fitness Survey data.

Figure 1.1 Estimated numbers of each age group for men who are over or under 70% heart rate max when walking for 1 mile at 3 mph (reproduced from Kelly et al., 2011)

Image subject to copyright

Original available at

http://www.tandfonline.com/doi/abs/10.1080/02640414.2011.609897#.UoOV5nCWZo4

Figure 1.2 Estimated numbers of each age group for women who are over or under 70% heart rate max when walking for 1 mile at 3 mph (reproduced from Kelly et al., 2011)

Image subject to copyright

Original available at

http://www.tandfonline.com/doi/abs/10.1080/02640414.2011.609897#.UoOV5nCWZo4

Cycling

Cycling is usually considered a more vigorous activity than walking [35, 36]. As well as a means of travel, it is often conducted as a sport, for exercise and training, and for pleasure [37]. A systematic review by Oja et al., reported cycling to be associated with reduced morbidity and reduced cardiovascular risk though showed equivocal results from three studies for all-cause mortality [38]. A longitudinal study, with follow up at six years reported better cardiovascular risk factor profiles in those children who cycled to school [39]. A Department for Transport (DfT) report measuring baseline cycling in intervention cities and towns reported that of adult respondents aged 16 and over, 28% had cycled in the previous twelve months; further, one in five adults (19%) could be described as frequent cyclists, in that they said they cycled at least once a week [40].

De Hartog et al., modelled the risks and benefits of a modal shift from car to cycling journeys in the Netherlands, taking into account health benefits, traffic accidents and exposure to air pollution; the benefits were found to be 11 times the risk [41]. Yang et al.'s review suggested community wide promotion and improving infrastructure both have the potential to increase cycling levels, though by modest amounts [35]. A 2011 report from the London School of Economics estimated that cycling contributed £3 billion to the UK economy through health benefits, reduced sickness-absence days, employment, manufacturing, and reduced congestion [42]. It should be noted that the report was commissioned by UK Cycling, so may be optimistic in certain assumptions, but it demonstrates the wider importance of cycling beyond traditional public health considerations.

Uptake of cycling may not always be equitable. An analysis of the London Cycle Hire Scheme found that females and those living in deprived areas were less likely to register [43]. However, among those that did, usage was higher in individuals from deprived areas. The authors suggest this indicates the potential for cycling interventions in deprived groups [43]. A study by Goodman et al., demonstrated an increase in non-school active travel (walking and cycling) predicted more time in MVPA in children. There was no evidence of substitution behaviour, when one physical activity replaces another and overall levels do not change, which is a common criticism of behaviour specific interventions [44].

Sedentary travel and health

Sedentary travel is transport that does not include physical activity. Generally speaking these are motorised forms of transport such as car, bus or travel by train. A recent study from Columbia showed that time spent travelling in motor vehicles was positively associated with overweight and obesity, after adjustment for socioeconomic status and other physical activity [45]. In a similar study, the same association was found for car ownership in males (though no significant result was found in females) [46]. It has been suggested that short motorised trips offer particularly good possibilities for active travel interventions [13].

Environmental benefits

There is environmental and public health synergy between reducing sedentary, carbon emitting activities such as car travel and replacing with walking and cycling. This modal shift will reduce pollutants and emissions, and can help to reduce traffic levels [8, 12, 26, 47]. Public transport with, for example, train or bus is more complicated. These modal share journeys still have emission costs, but also offer opportunities for increased active travel to and from bus stops and stations; the design of the transport network is an important factor in this regard [48, 49].

NICE Guidance on walking and cycling

In 2012 The National Institute for Health and Clinical Excellence[3] (NICE) produced Public Health Guidance on walking and cycling [7]. As the guidance was commissioned by the Department of Health (DH), this reflects the increasing national policy relevance of walking and cycling in the UK. The aim was to set out how people can be encouraged to increase their levels of walking and cycling for travel and recreation purposes [7].

The guidance covered; policy and planning; local programmes; schools workplaces; and the NHS. There were also research recommendations regarding population wide interventions, factors that influence walking and cycling, specific approaches, sustainable change and differences within

[3] NICE changed to the National Institute for Health and Care Excellence in April 2013 and are classified as a Non Departmental Public Body (NDPB) as a result of the 2012 Health and Social Care Act

population groups. NICE have produced an estimated cost per QALY (Quality Adjusted Life Year) for walking of approximately £2700[4] from an evaluation of the Get Walking, Keep Walking programme [7, 50].

[4] When considering drug treatments, costs of £20,000-30,000 are considered the threshold for cost-effectiveness (Measuring effectiveness and cost effectiveness: the QALY; available at: http://www.nice.org.uk/newsroom/features/measuringeffectivenessandcosteffectivenesstheqaly.jsp)

Why do we measure physical activity and active travel?

Measurement is fundamental to epidemiological study of physical activity and health [51]. The evidence discussed so far in this chapter is derived from research based on some measurement of walking and cycling. Measurement of walking and cycling allows us; to monitor prevalence and trends with populations and sub-groups; to understand the exposure to the behaviour and associations or relationships with certain health outcomes; to investigate the potential determinants and correlates of the behaviour; to test if interventions and policies have succeeded in increasing or maintaining activity levels; to identify appropriate populations or sub-groups for intervention; and to model and calculate the economic or environmental benefits [52-54]. A framework for this is discussed below.

Beyond public health, travel planners use travel demand models that are calibrated and validated using input data obtained from measuring travel behaviour [37, 55, 56] and policy makers and use these travel data to direct traffic safety and injury prevention policies [57]. Therefore, much research in both the public health and transport sectors is aimed at understanding who is making journeys, when and why; the distance, duration and speed (or intensity if walking and cycling); and what routes and modes are used and why [13, 55, 58-60].

Studies of travel behaviour may be cross-sectional or longitudinal, and vary from population, region or city-wide investigation [60-62] to small scale evaluation of walking interventions [63]. The results from these studies often inform research, practice, policy, travel planning, and funding decisions [60, 64-66]. Measures of travel behaviour with appropriate validity and reliability are therefore often required. Poor validity and reliability limit the conclusions that can be drawn from any type of research [67].

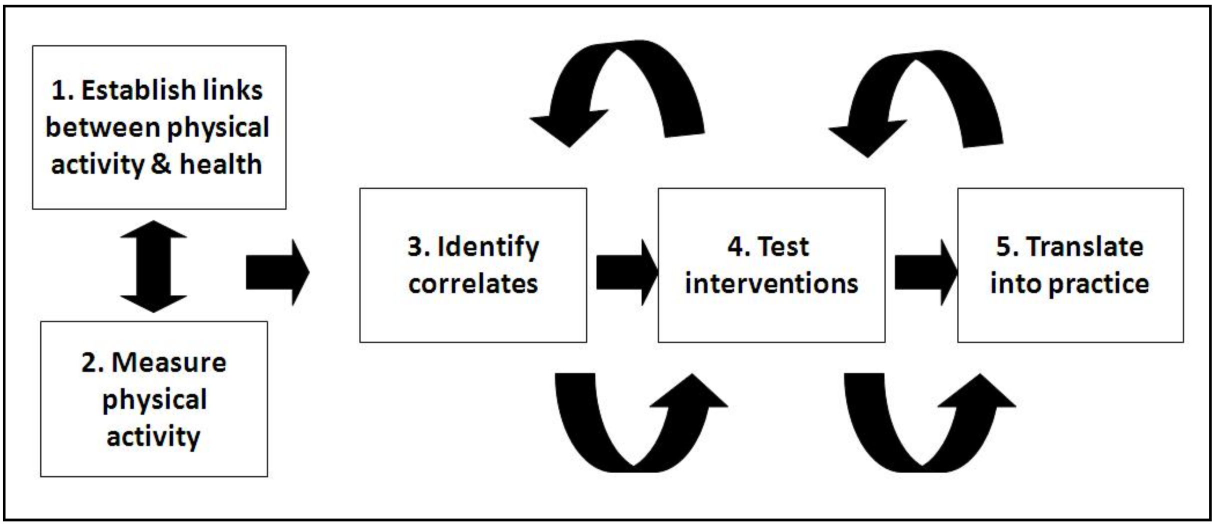
What is the theoretical framework?

The rationale for improving behaviour measurement is based on the theoretical framework for epidemiology as proposed by Sallis and Owen (1999) [68] and Sallis, Owen and Fotheringham (2000) [52]. They describe five main phases of 'behavioural epidemiology' research in relation to physical activity;

1. Establish the links between physical activity behaviour and health: The associations between behaviours and health outcomes give the foundation for subsequent research.
2. Develop measures for validly and reliably measuring physical activity behaviour: High quality measures are required.
3. Identify factors that influence the behaviour: The correlates and potentially modifiable risk factors that can be targeted by interventions or policies.
4. Develop and evaluate interventions to change the behaviour: Interventions with an evidence base.
5. Translate research into policy and practice.

This framework is represented in Figure 1.3. The place of measurement (Phase 2) and its relationship to the other phases is displayed. It is clear that the other phases are reliant on measurement.

Figure 1.3 Owen and Sallis' Behavioural Epidemiology Framework



Other behavioural frameworks exist, for example McLeroy et al.'s Ecological Model for Health Promotion [69]. Krizek et al., report sectors beyond public health have their own concepts of travel behaviour; for example in transport, models are often based on choice and utility maximisation [62]. However, I feel Owen and Sallis' best describes the role of measurement in public health and this is why I have chosen to use it to frame this thesis with an active travel focus.

Why is measurement of active travel difficult?

Like many physical activity behaviours, active travel is difficult to measure [10, 62, 67]. This is because it covers a spectrum of durations and intensities, it cannot be quantised into easily measurable packets, and there is no precise biological marker of completion or the health outcomes it influences. The instruments used to measure walking and cycling must be feasible, reliable, valid, non-reactive and able to assess usual and habitual or specific activities, depending on the research question or study design. Further they may also be required to measure vehicle travel or public transport use in studies of overall travel behaviour. That these instruments are low cost and confer low participant burden is also desirable [70]. Finding measures that sufficiently fulfil these criteria is a public health challenge, and it has been for some time.

How do we measure travel behaviour?

Krizek et al., propose three types of walking and cycling measurement; self-report, observation and instrumentation [62]. I believe that this proposal holds when considering both walking and cycling, and overall travel behaviour. *Direct observation* of participants is considered the criterion measure of journey behaviour [71, 72]. However, this is too resource and time consuming to be available or practical in anything but the smallest studies [71, 72].

A common technique for travel behaviour remains self-report by, for example, *Travel Diary* or *Questionnaire* [20, 62]. These self-report data can be collected in a number of ways including researcher visits, postal surveys, web surveys and telephone interviews. Despite the emergence of new technologies, self-report remains in high use for reasons of feasibility, cost, comparability and data handling [73]. Self-report is considered a subjective measure of travel behaviour as it is based on respondent opinion, attitudes, memory and perception [71].

In addition to self-report, there are various tools and technologies available for travel researchers [71]. For example, *accelerometers* are used to measure activity intensities and patterns [74]; *pedometers* are used to record steps counts [74]; and *global positioning systems (GPS)* are used to investigate journey routes and speeds, which can be a proxy for mode [75]. These technologies allow us to assess the dynamic properties of travel behaviour by collecting detailed spatial, temporal and characteristic conditions throughout the journey experience [72]. They are considered objective measures as they directly assess a physical parameter or property e.g. steps per day [71].

Technologies can also be used in combination, for example to improve mode prediction [76-80]. This is discussed further in the next section. However, a recent paper from New Zealand was critical of the level of integration between systems for measuring transport related physical activity [13] and automated techniques for aligning GPS data and self-report data, for example, are said to be still in their infancy [81].

Strengths and weaknesses of travel measurement techniques

Self-report

Self-report of travel behaviour is one of the most common and important ways of obtaining data on travel behaviour [20, 65, 72, 82]. Unfortunately, the potential for human error means self-report can lack the required validity and reliability. Due to problems with adherence, interpretation of question, memory and judgement, the data can be subject to missed trips, incomplete entries and misreported journey characteristics (e.g. duration, distance or purpose) [56, 58, 64, 83-87]. Self-report is a measure of the participant's perception of what they think they did, or habitually do, rather than an objective measure, and can be subject to social desirability bias [62, 88]. The Yang review of cycling interventions stated the reliance of studies on self-reported measures was a major limitation as they are of unknown validity and reliability [35].

Measurement error obscures true relationships between behaviour and health, and the effects of interventions may go undetected if the change in behaviour is lost in the error in the measure. The assumed error in self-report is often countered by employing large sample sizes [87]. However, this increases the scale of any data collection effort, and will not improve measurement at the level of the individual.

A particular criticism of self-report travel is forgetting trips altogether and this is said to be particularly true for short trips [88]. Trips undertaken using active modes are more likely to be missed (possibly as they are the shorter trips), and this is thought to be especially true for longer recall periods [66]. Self-report of motorised travel has also been shown to be limited [72, 89]. Trips may not be reported due to length of survey, respondent forgetting or considering the trip unimportant and selective omission [61, 90]. In travel surveys, "short trips, walk trips, and shopping trips" are more likely to be omitted, as respondents are "time-penalised" for being truthful and complete [88]. Trip reporting fatigue in multi-day travel data collection has also been reported; the Dutch National Mobility Panel found recorded trip rate decreased by 1% per day; in walking trips this was 3.5% [91].

There may be other reasons for the error or bias in self-report. For example, diaries may be completed at the last minute or at the end of the study period rather than daily or as instructed [92]. Beyond memory, respondents can have a tendency to round start time, end time and overall duration (known as digit preference) to the nearest 5, 10 or 15 minutes [93-95]. Rietveld has discussed rounding times, saying respondents are drawn to "anchor points" when recalling durations [95].

However, self-report is considered to have good validity for assessing travel purpose and mode [72]. It also generates raw data that is easier to manage than, for example GPS or accelerometer, and self-report is currently the most cost efficient method [96]. Beyond mode, frequency and duration, travel diaries can be used to collect detailed information about journey purpose, destination, ticket price and accompanying passengers [89]; variables often beyond other measures. Self-report measures can also add depth and context. For example, some self-report measures are designed to measure perceptions about active travel and active travel environments [97, 98].

Global Positioning Systems (GPS)

As a result of the limitations in self-report, there have been calls for objective measures of travel behaviour that measure the desired criteria [13]. First suggested for physical activity assessment in 1997 [76], *Global Positioning Systems (GPS)* devices (wearable and in-vehicle) are increasingly used in large scale travel surveys to supplement self-report measures [61, 80]. For some time there has been discussion about ultimately replacing self-report with GPS measures [55, 61, 86, 99, 100]. With advances in technology, GPS are also increasingly used to investigate journey routes, destinations, modes and speeds [13, 59, 75, 86, 101]. City and state-wide GPS surveys are now common in the USA and Australia [21, 58, 86, 102, 103]. A GPS pilot was recently added to the NTS in the UK [99]. GPS is regarded as having a lower participant burden than active techniques such as diary or survey [88]. However, like most technologies GPS still require some active participant intervention in wearing or operating the device depending on the specific protocol and unit. Though not entirely error-free, it is described as a "more objective" and "accurate" measure of travel behaviour than self-report [101, 102]. However, in many cases, the primary measurement technique remains self-report for reasons of feasibility, cost and data handling [20, 73].

GPS can be subject to cold starts (delayed signal acquisition) and signal loss around tall buildings, tunnels, trees, etc. [79, 104]. A review of GPS use to locate physical activity reported variable data loss of 2.5% to 92% across 17 eligible studies [105]. A 2011 travel study from New Zealand reported 89% of trips were lost to GPS for a range of reasons [13]. GPS present the challenge of transforming an enormous data set of time-stamped, geo-coded data points into a useful data base for analysis [79, 81, 106]; it is said they generate a dataset that is highly detailed, but highly complex to interpret [107]. They also rely on algorithms to detect trips in the data, and impute mode and purpose [81, 86]. Their particular strength is in objectively obtaining journey route, previously only available to subjective participant recall [55, 96]. It has been argued that they should primarily be used to assess where physical activity (e.g. active travel) takes place and to understand the role of the environment as a determinant of this activity in combination with *GIS (Geographical Information Systems)* [79, 96, 104]. In 2012, authors in the field of GPS were still asking "whether the traditional survey-reported collection method may be preferred with regards to some types of travel" [86].

A 2012 study from New York City, USA reported a "promising 82.6% success rate" for mode detection from GPS data. The authors said that management of the enormous data sets generated is still a challenge, especially in an urban setting with "urban canyons" and complicated transport networks [108]. Some authors have suggested that GPS equipped mobile phones may be the future of such research due to their ubiquity and decreasing unit cost. These technologies employ Assisted GPS

(AGPS) and wireless transfer to provide additional information and can cover indoor areas. However, they do have issues with privacy, consent and working across multiple providers and networks [109].

Motion and physiological sensors

Pedometers are good at measuring outcomes such as steps per day, and accelerometers can assess total physical activity as well as patterns and intensities [110]. However, they have difficulty detecting individual journeys making them unsuitable for travel behaviour studies when used alone. They cannot measure driving, bus or train use, and have difficulty detecting cycling [111]. Used in combination they can provide useful data. For example Cooper et al., used GPS and accelerometer to assess the contribution to total MVPA of the journey to school in 11-12 year children [112] and Rodriguez et al., recently demonstrated a technique for identifying walking trips in adolescent females by merging accelerometer and GPS data [113]. Similarly, heart rate monitors (e.g. Acti-Heart) can be combined with other technologies to assess the physiological response to the activity.

GPS, pedometers, accelerometers and heart rate monitors are considered objective measures as they directly measure a physical property [67]. However, the data from these measurements still have to be analysed and processed for travel behaviour to be determined. These data require interpretation by researchers, or researcher written algorithms, and therefore also have elements of subjectivity. I would say they directly measure parameters which are considered indicators of travel behaviour, but they are themselves indirect measures of travel behaviour.

It is worth noting, these measures discussed are all measures of individual behaviour. There are other ways to assess travel behaviour I would classify as within the observational category proposed by Krizek. These include traffic surveys where pedestrian, cyclist and vehicle numbers can be counted by researchers or counting devices set up at strategic travel routes or nodes [62]. Additionally infrastructure monitoring can assess usage of cycle paths or walkways, and tickets sales can be used to monitor use of public transport. These techniques are outside the scope of this thesis.

Summary of travel measures

To summarise this section I have created Table 1.3 to show my interpretation of the capabilities of the various measures of travel behaviour discussed. I present this as my opinion; each cell is a large body of work on its own. I have deliberately left this unreferenced as the range of effectiveness of each measure at each dimension will vary widely between study design, testing conditions and device used. This also assumes no combinations of devices. However, I believe this is a helpful reminder of the strengths and weaknesses of each measure and their suitability, or not, to different research questions and outcomes.

Table 1.3 My summary of the capabilities of common travel measurement techniques

	Measure					
Dimension of travel behaviour	**Self-report**	**GPS**	**Accelerometer**	**Pedometer**	**Heart rate**	**Direct observation**
Mode	Subject to participant memory	Subject to detection algorithm	Subject to behavioural detection	Only walking	No	Yes
Frequency	Subject to participant memory	Subject to detection algorithm	Subject to behavioural detection	No	No	Yes
Duration	Subject to participant memory	Subject to detection algorithm	Subject to behavioural detection	Step counts per day	No	Yes
Intensity	Subject to participant memory	Speed information subject to detection algorithm	Estimated energy expenditure subject to behavioural detection	No	Heart rate response	Some researcher estimation
Context or purpose (work, leisure)	Subject to participant memory	Some inference from land use	No	No	No	Yes
Travel Environment	Subject to participant memory	If combined with GIS	No	No	No	Yes

Analysis of self-report in this thesis

Having considered the widespread and continued use of self-report techniques in travel and active travel measurement, as well as the uncertainty about the inherent error, I have decided to focus on investigating self-report in this thesis. I acknowledge technologies such as GPS, accelerometers and combinations of both will increasingly be used as the technology and data processing improve. Research into these methods is also warranted, but will not be included in the scope of my investigations. I have decided to focus particularly on the National Travel Survey (NTS) and will discuss this in the following section.

The National Travel Survey (NTS)

For this thesis I have chosen to focus on one particular self-report measure of travel behaviour; the National Travel Survey (NTS) commissioned by the Department for Transport (DfT). The NTS provides up-to-date and regular information about personal travel within the UK and monitors trends in travel behaviour. Details are given in Text box 1.1.

Text box 1.1 Details of the National Travel Survey (reproduced from National Travel Survey 2010 Technical Report)

The NTS is one of the Department for Transport's main sources of data on personal travel patterns in Great Britain. The Ministry of Transport commissioned the first NTS in 1965/1966 and since July 1988, the NTS has been a continuous survey (fieldwork conducted on a monthly basis) with an annual set sample size of 5,040 addresses. In 2002 the annual set sample size increased to 15,048 addresses. Since January 2002, the Department for Transport has commissioned the National Centre for Social Research (NatCen), an independent social research institute, as the contractor for the NTS.

Individuals within each participating household are asked to complete a seven day travel record. They are also interviewed and detailed information on the key characteristics of each participating household and any vehicle to which they have access is collected.

The NTS produces data at a number of levels; household, individual, vehicle, long distance journey, day, trip and stage. Data from the NTS is used extensively by the Department for Transport to monitor changes in travel patterns and to inform the development of policy. The findings and data are also used by a variety of other organisations including: other Government departments; university academics and students; transport consultants; local authorities and voluntary sector organisations representing a wide range of interests including motorists, cyclists, the elderly, rural communities and children.

The NTS measures both sedentary and active travel modes. While this thesis is motivated by measurement of walking and cycling, the previously mentioned health associations with sedentary travel further the rationale for investigating the NTS.

Uses of the NTS in public health

There are many uses of NTS data in the field of public health including prevalence estimates, trend surveillance and health modelling. For example, Woodcock et al., modelled the health and environmental impacts of different walking and cycling scenarios in the UK using baseline walking and cycling data from the National Travel Survey [12].

The UK Public Health Observatories have used NTS data on walking and cycling to monitor overall changes in physical activity [114]; as have the Office for National Statistics in their 2005 report; Focus on Health [115]. The previously mentioned apparent decline in walking and cycling was based on NTS data [10]. The recent HEAT for Cycling and HEAT for Walking[5] projects were designed to allow practitioners to model the economic benefits of cycling and walking interventions [37]; they have also been used with input data from the NTS [12].

Other studies using NTS data include an investigation of the likelihood of walking predicted by bus pass provision in older people [116] and analysis of road injury risk compared to time spent travelling; in this study the authors found that for young males cycling is safer than driving [117].

Brennan et al., produced a health and economic modelling report for the previously mentioned NICE Walking and cycling guidance [118]. In this report they present *The ScHARR Walking and Cycling Model* which uses NTS data to estimate the relationships between journey distance, duration and frequency.

Validation and investigation of the NTS

While deciding to focus on the NTS I travelled to NatCen (London) on a number of occasions to meet the team involved in running the NTS. They informed me that the main work validating the NTS was

[5] HEAT = Health Economic Assessment Tool

cognitive validation of respondents understanding of questions and definitions [65, 119]; this work suggested two types of responders; readers and skimmers, with varying time spent completing the diary affecting error on reported data. Beyond this, the NTS undergoes substantial weighting and correction to account for sample bias [120]. However, less work had been conducted trying to validate, investigate the error, or calibrate the self-reported data using objective measures.

There was an attempt to use GPS in conjunction with the NTS, with a 2009 feasibility study and 2011 pilot study [99, 100]. However, these studies had difficulty when attempting to match the GPS data to the diary data. Due to the algorithms and processing used, the average GPS journey duration was calculated at 51 minutes, while the average reported duration was 21 minutes. There were also 30% fewer journeys identified in the GPS data than were reported in the diary data [99]. The authors concluded that the GPS devices (Amtel) and processing system used did not offer an acceptable alternative to the self-report method. In other words, the difference between GPS and diary data was not primarily thought to be a result of error on the self-reported diary data.

I came across the following reference in an NTS technical publication; *"Analysis of the accuracy of NTS respondents' estimates of trip distance and duration"; Cronberg and Bonsall (2006); Technical annex 1 of Research Project UG599 (From Review of the Potential Role of 'New Technologies' in the National Travel Survey), Department for Transport, London* [121]. This is the only publication I have found that was a deliberate attempt to quantify the error on self-reported journey data on the NTS. However, at the time of writing, the report is not available on any website or online resource. I was able to obtain a copy by emailing the author (Professor Peter Bonsall; Leeds University) directly.

The analysis of 110,563 daily records compared reported trip distances and durations to three different route formulations using Microsoft MapPoint software (version 11.0); quickest-route-by-car; shortest-route-by-car; and crow-fly distances using recorded post codes. Cronberg and Bonsall found evidence for positive differences indicative of over-reporting. However the estimate of how much journeys were over-reported varied from zero to 6 minutes per journey depending on the method used. Overall, only 11% of reported durations fell within 10% of the MapPoint estimate for that trip, and 23% falling within 20% of it. The accuracy was variable by journey duration, mode, and purpose; closest for car trips and worst for walking trips [121]. The authors also reported a tendency for respondents to round duration estimates.

However, the authors acknowledge these MapPoint estimates are a proxy for "objective" measurements of true journey time; for example they do not take into account the effect that variations in traffic levels may impact on journey times or the routes people may choose and therefore they are unlikely to provide good estimates. The authors concluded that "the true distances and durations are, of course, unknown" [121].

Use of the NTS in this thesis

Having reviewed the importance of, and issues with, measurement of travel behaviour I have identified the NTS as a widely used measure that has not been objectively validated. In response, this thesis is an investigation of the error on self-reported journey behaviour in the NTS. Specifically I will attempt to investigate the error on exposure to travel, active and sedentary travel, and walking and cycling. The following sections will outline the benefits of such work, the literature on previous attempts to validate or calibrate self-reported travel measures, and a potential new technology to test.

Investigating the size and source of error on self-reported travel behaviour

As self-report remains in common use, there have been various attempts to validate, or otherwise investigate the error or bias on these measures. Studies have sought to identify which population groups or travel modes provide better or worse data and find the factors that may influence or predict error.

Comparing GPS and self-report

Though GPS and self-report are often used to complement each other, a series of studies that have used GPS and self-report methods to measure travel behaviour and have found differences in the obtained data [61, 89, 122, 123].

In the US, GPS has been used to supplement over ten state and regional travel surveys and to identify the rates and characteristics of mis-reported trips and their respondents. The authors sought to develop ‘‘correction factors’’ for adjustment of trip estimates for travel models [86, 102, 124]. A study using in-vehicle GPS showed increasing age may be a factor in increased error in self-report [89]. A similar study found that self-reported driving was particularly unreliable in those aged 60-80 years [125]. Walking and cycling have been found to be the most poorly reported modes [61] and it appears incidental walking may be reported less well than purposeful walking [126, 127]. In terms of purpose, home based and non-work journeys are more likely to be missed than work journeys [122].

As previously mentioned, journey under-reporting (i.e. reporting fewer journeys than measured by GPS) is a common finding. Of regional and state-wide studies in the US the range of under-reporting in self-report was found to be from 10% in Kansas City to 81% in Laredo Texas [86]. “Trip under-reporting” was found to be influenced by survey length and method, respondent burden and comprehension, respondent demographics and attitudes, and travel characteristics [86, 124].

Duration

One particular aspect of travel behaviour that is often investigated in travel research is journey duration [60, 61, 128]. In public health duration indicates exposure to active travel as well as identifying which motorised trips might be suitable for health interventions [13]. In transport research, duration can indicate traffic congestion and emissions and the data are critical to transport modelling [60, 64, 65]. In comparison of GPS and self-report studies, differences in measured and reported durations have been found. A 2011 study by Housten et al., reported "Mean per cent of day" engaged in travel or waiting during travel; in this study a self-report group reported 6.9% (95% CI = 6.1 to 7.8) of day while a different but comparable group wearing GPS was found to spend 8.3% (95% CI = 8.0 to 8.6); further, participants did not report nearly half (49%) of the locations and trips identified in GPS-enhanced data on their activity diaries, resulting in about 3 h/day in unreported locations and 0.6 h/day in unreported trips [129]. While most studies agree there is error on self-reported travel behaviour, to date I have not been able to identify any consensus on the magnitude of this error.

The use of GPS and comparison to self-report has raised a fundamental question about what a trip or journey actually is; when does it start, when does it end, and when can it be considered "paused" (referred to as dwell time) [86]? Movements within large, multi-building areas such as university campuses, hospitals or shopping centres will be detected by GPS, but may not be perceived as journeys by the participant; this may explain some of the apparent under-reporting of trips when comparing GPS to self-report [86]. This highlights the difficulty of designing questions for travel behaviour research that respondents understand and that will capture the right information [62].

As stated previously, GPS has its own limitations and any validation or error investigations using GPS will themselves be limited. It is a good measure of location, and can be a good measure of mode, but signal loss, signal acquisition and data interpretation can lead to variable trip detection and duration estimates. A better measure of travel behaviour, closer to direct observation, would allow for improved validation or error estimation.

Wearable cameras may offer a solution - SenseCam

"Valid methods of measuring diet and physical activity remain a challenge and invalid measures may yield erroneous or inaccurate data which has serious implications. The hope is that technological advances in both fields will yield methods with increased validity" (MRC Website - Diet and physical activity measurement toolkit)

SenseCam is a technology that may provide objective measurement and therefore be able to assess the error or bias on self-reported journey behaviour. It is a lightweight digital camera worn around the neck (see Figure 1.4) that passively captures images approximately every 10-30 seconds throughout the day [83, 130]. On board heat, light and motion sensors are used to detect changes in environment and "intelligently" record more images at times of high activity. The camera was developed in the field of *digital life-logging* [83].

Figure 1.4 SenseCam is a lightweight digital camera that can be worn around the neck (reproduced from Kelly et al, 2011)

Image subject to copyright

Original available at

http://www.ijbnpa.org/content/8/1/44

What is digital life-logging?

"Life-logging" refers to the automated digital capture of everyday life activities through first-person point-of-view images. First conceived by Vannevar Bush in the 1940's [131], it has traditionally been a pursuit of those in the computing and engineering domains, where much effort was placed in miniaturising device size and increasing battery capture time [132]. In 2003, the Sensors and Devices Group at Microsoft Research Cambridge developed SenseCam. It has been used in a variety of applications including artistic capture of life experiences, a therapeutic aid for those with Aphasia (speech-related brain disorder), market research analysis, social sharing of everyday images and a memory aid for Alzheimer's patients [133]. The strength of SenseCam is in its ease of use, a long battery life (up to 16 hours of continuous use) and storage capacity (capable of holding over one week's worth of images, ca. 32,000 life-log images).

Images have previously been used in social sciences for a number of purposes; to elicit or provoke discussion and other data; for feedback and documentation of the research process; and as a mode of interpretation and representation [134]. This thesis is an investigation of their potential utility in measurement of health behaviours, specifically travel and active travel. I will assess whether they can be used to investigate the suspected errors or bias on self-report.

Have wearable cameras been used in Public health before?

Those interested in measurement of health behaviours (e.g. physical activity, active travel, sedentary behaviour or nutrition) often wish to record periods of at least a day. Therefore the use of passive digital image capture (e.g. with SenseCam) for health behaviour research has only recently become possible due to technological advances in device miniaturisation, memory storage and battery life. Since 2010, the first feasibility and methodological studies in physical activity and nutrition employing these techniques have been published [135, 136]. In March 2013 The American Journal of Preventive Medicine published a series of articles on the emerging theme of wearable camera research [137-142]. I was lead author for one of these and co-author on two others.

SenseCam images in travel measurement

If participants wear the SenseCam device during their travel activities, I will be able to collect a continuous visual log of their journey. From the visual inspection of the images I will be able to directly observe the mode of travel;

- In car from vehicle features with steering wheel indicating if driver or not
- On bus or train from in vehicle or carriage features
- On bike from handle bars
- Walking from progression through environment

This technique is closer to objective assessment of mode than GPS or accelerometer which relies on inference from speed or movement and activity counts. While this assumes that the participant is actually wearing the device, rather than a friend or family member, this assumption applies to the aforementioned devices too.

Furthermore, as each image is time-stamped, by taking the first and last image of the journey, the total journey duration can be calculated. Microsoft state the camera records an image at least every 30 seconds [83]. This means that the SenseCam journey duration will be within 60 seconds of the true value if I can identify the correct start and end images. If SenseCam is able to provide more objective assessment of journey mode and duration than previously available from other technologies, I may be able to design an experiment to find the best estimate yet of the error on self-reported travel data.

Benefits of validating or correcting self-reported travel data

The data from the NTS is subject to the usual limitations of self-report. By assessing possible error on the self-reported NTS I can investigate such questions as if there are systematic biases or large individual variations in error.

Calculating the error on self-reported journey duration will potentially;

- Allow for correction factors to be applied to existing data that reports travel levels
- Enhance transport related models by allowing correction of self-reported data
- Increase the validity and confidence of health associations or carbon emissions calculated using NTS-type walking, cycling and driving data
- Indicate the sensitivity of similar self-report when used to measure behaviour change

It has also been reported that understanding the magnitude of random error (referred to by many authors as precision) allows for better estimates of required sample size and data collection periods [143]. This allows the avoidance of unnecessary participant burden and survey costs.

Thesis objectives

In this section I will outline the overarching aim and primary objectives for my thesis. Specific methodologies and research questions will be covered in the appropriate sections. This section is designed to allow the reader to understand the research arc of my thesis.

In this chapter, I reviewed relevant travel and measurement literature. I identified the importance of research that improves the way we are able to measure travel behaviour (and the important dimensions within travel behaviour). I discussed the uses of the NTS in public health, and the issues with self-report. I also discussed SenseCam, its origins in the field of life-logging and computer science, and its potential applications in the field of travel measurement.

This leads to the thesis aim:

"…to assess the efficacy of utilising SenseCam to improve our understanding of the error on self-reported travel duration using a modified version of the National Travel Survey"

To achieve this aim (and based on the content discussed in this chapter), I formulated three thesis objectives:

1. Systematically review the existing literature comparing GPS measured and self-reported journey durations
2. Develop a new method for travel measurement with SenseCam, with a view to valid and reliable journey assessment
3. Test my developed SenseCam method in a suitably powered sample, and investigate the error on self-reported journey durations (and compare to results from GPS review)

Objective 1: Systematically review the existing literature comparing GPS measured and self-reported journey times

My review of the literature in Chapter 1 revealed GPS and self-report are the two primary methods of investigating travel behaviour. As the technology improves, GPS is increasingly considered an objective measure.

Some studies present comparison of simultaneously collected self-report and GPS travel data; in some cases this will provide information on the difference between GPS measured and self-reported travel duration. If GPS can be considered the criterion measure, then this difference is indicative of the error in self-report.

Before developing my own method with SenseCam it will be appropriate to appraise the current "best evidence" from other methods for assessing the error on self-reported journey duration. I decided to do this by systematically reviewing the evidence for GPS assessed error of self-reported journey duration.

Objective 2: Develop a new method for travel measurement with SenseCam, with a view to valid and reliable journey assessment

Before SenseCam can be used to assess the error on self-reported travel behaviour I need to develop a feasible, valid and reliable method for generating travel data. This will require field testing, pilot testing, comparison to a criterion measure and appropriate assessment of burden, ethics and practical considerations.

I will describe the process by which SenseCam can be used to generate meaningful travel data and how I evaluated its performance to do so. This will involve data collection procedures, data management, coding protocols, and analysis techniques.

Objective 3: Test my method in a suitably powered sample, and investigate the error on self-reported journey durations (and compare to results from GPS review)

Once my method has been developed, I will conduct an experiment to assess the error on self-reported travel duration. This experiment will have a suitably powered sample size, such that estimates of the error or bias have acceptable and relevant confidence. Results from my pilot study indicate I will need approximately 1000-1500 measurements (journeys)[6] for suitable confidence on assessment of error; this will be achieved by 70-90 participants wearing SenseCam and completing a travel diary for a three day period. I will compare this result to that of the systematic review. The aim is to demonstrate a "better" assessment of the error on self-reported journey duration.

Chapter conclusion

This chapter reviewed the literature on walking and cycling as important behaviours to public health. I discussed the importance of being able to measure walking and cycling behaviour, with a focus on the National Travel Survey which generates important UK data on travel and active travel. I then introduced SenseCam (a wearable digital camera) and suggested it may have the potential to improve how we are able to measure travel and active travel. Finally I outlined my three thesis objectives.

[6] The required number depends on desired confidence. Formal sample size calculations will be presented in Chapter 7.

CHAPTER 2 – SYSTEMATIC REVIEW OF GPS AND SELF-REPORT TRAVEL BEHAVIOUR

Summary

This chapter describes a systematic review of the literature comparing self-reported and GPS measured journey durations. This review was a response to the literature I discussed in Chapter 1, which covered the importance of measuring travel behaviour, the suspected error on self-report and the potential utility of GPS as a way of investigating this error. This review is designed to answer my first thesis objective.

In this review I found that despite the prevalence of GPS studies, there were only eight that presented a useable comparison between GPS and self-report. All these studies found that self-report over-estimated travel duration in comparison to GPS (with a range of +2.2 to +13.5 minutes per journey).

This systematic review suggests that when using self-reported journey behaviour, the journey durations should be treated as an over-estimation. The study methods and findings were heterogeneous, and there is still ambiguity about the magnitude of the bias on self-report. I take this as a rationale for developing a SenseCam method which has the potential to be a more objective measure of journey duration than GPS.

Manuscript from this work

The work presented in this chapter is based on a manuscript which is currently under review at Transport Reviews (Quantifying the difference between self-reported and GPS measured journey durations: a systematic review – see Appendix 2). My co-authors were Patricia Krenn, Sylvia Titze, Peter Stopher[7] and Charlie Foster. I am the first author for this work. Certain sections and passages in this chapter were used in the submitted version of the manuscript.

Background

As discussed in the first chapter, measurement of travel behaviour is often by self-report, by GPS or some combination of the two; however, the quality of the data collected is subject to certain limitations. In response my first thesis objective (previous chapter) I conducted a systematic review of studies that compare GPS measured and self-reported journey duration. This would allow me to appraise the current evidence for the bias or error on self-reported journey duration. I found no evidence of an existing systematic review of this question during my review of the literature.

The objective of my review was to answer the following research question:

What does the available evidence reveal about GPS measured error or bias on self-reported journey durations?

I would do this in the following way:

1. Systematically reviewing the available evidence from studies measuring journey duration using both GPS and some self-report method (diary, questionnaire or interview)
2. Identifying studies that report the size and direction of bias on self-reported duration

[7] Prof Titze and Dr Krenn are based at the University of Graz. I know them as we are members of the HEPA Europe network and they have visited our group to test the feasibility of using SenseCam in their own GPS work. I invited them to contribute to this review as Dr Krenn (supervised by Prof Titze) conducted a review of GPS and data loss for her own doctoral studies, and this informed the design of my review. I read a lot of Prof Stopher's work during my literature review; he has published previously on GPS and self-report comparisons. When I travelled to Melbourne for the ISBNPA 2011 annual meeting I visited his unit at Sydney University and he advised me on this review. One of his PhD students has subsequently visited our group to collect SenseCam data for their own work.

Methods

The following protocol is presented here as it is appears in the submitted manuscript. I had previously been involved in a systematic review into active travel in children [20], specifically in the quality assessment and data extraction stages. This was the first time I led a systematic review from start to finish; I developed and refined my method with advice from Dr Charlie Foster, Professor Sylvia Titze and Professor Peter Stopher. I was advised on search terms (see Appendix 2) and databases by Nia Roberts who works at the Bodlean Healthcare Library in Oxford University.

Inclusion criteria

Studies were considered for inclusion if they measured and reported travel behaviour, specifically journey duration (or frequency and total time) by GPS and self-reported methods.

Papers were included if:

- They were cross sectional, longitudinal or intervention studies
- Journey duration was assessed by GPS and self-reported methods
- Journey duration was reported for both measures

Papers were excluded if they did not meet all inclusion criteria.

Search strategy

Searches were conducted in the electronic databases of publications listed below. Nia Roberts helped me complete these searches. The dates refer to the available databases at the time of searching. I acknowledge GPS only became available for research use in the late 1990s, and I did not anticipate any included studies from dates earlier than this.

Health databases:

- Embase (OvidSP) (1974 – present)
- Medline (OvidSP) (1948-present)
- Web of Knowledge (1945 – present)

Travel databases:

- GeoREF (OvidSP) (1966 – present)
- Transport Database (OvidSP) [includes International Transport Research Documentation (ITRD) and the Transportation Research Information Services (TRIS) database] (1988 – present)

Searches were originally conducted on 11th October 2011 and were updated on 1st December 2013 prior to submission of the paper[8]. The reference list of each included study was searched for further relevant studies. The search strategy was conducted using only English search terms but language was not an inclusion-exclusion criteria.

Dr Charlie Foster and I screened all titles to exclude irrelevant studies. Disagreements were resolved by discussion. Abstracts were retrieved for remaining titles and were assessed for relevance by me, Dr Patricia Krenn and Prof Sylvia Titze. Disagreements at this stage were settled by Dr Charlie Foster. Full texts were retrieved for remaining studies and assessed according to the inclusion-exclusion criteria.

Data extraction

I used a data extraction form (agreed with all authors) to extract information from included studies. Extracted data included location, year, study population, sampling, GPS device, self-report technique, analysis approach, and data treatment and outcome. Any uncertainties were resolved by discussion with Dr Charlie Foster.

Study quality

Each of the included studies was appraised against seven quality criteria (by me and Dr Patricia Krenn): year of study, sample size, representativeness of sample, number of measured trips, discussion and quality of data treatment, number of travel modes investigated and peer review of report. These categories were modified from a previously published quality criteria matrix used by Dr

[8] The updated searches revealed 230 more titles, but no full texts suitable for inclusion. The results presented in this chapter refer to the most up to date searches

Krenn [105]. Three categories were scored by applying binary metrics (0/1) and four by applying tertiary metrics (0/1/2), with any between-rater discrepancies being resolved by discussion (see Table 2.1). The quality of each paper was described by a score summarising the metrics to provide an overall impression of the quality of the available evidence. The possible scores ranged from 0-10.

Table 2.1 Criteria used to assess quality of studies

Criteria	**Possible score**		
	0	1	2
Year of study*	Pre 2005	2005-present	-
Population size	<50	>50, but <=100	>100
Number of measured trips	<100	>100, but <=200	>200
Data Quality**	Low	Medium	High
Population representative?	No	Yes	-
Modes measured	1	2 or more	-
Peer review	No	Yes	-

*Year of study meant as indicator of device quality. Based on assumption that devices and software have improved since 2005.
**Data quality is a subjective judgement regarding the proportion of collected data (trips) that were eligible and/or retained for analysis

Statistical significance

Where available standard deviations on mean durations were used to calculate 95% confidence intervals on the differences between means as described in Kirkwood, et al (2005) [144].

Results

Study selection

1,232 titles were obtained by applying the search strategy and five more titles were identified from online and "grey" sources. After screening for relevance 240 titles were retained and abstracts retrieved, with agreement between researchers on 97.3% of titles (Cohen's Kappa 0.895). From these abstracts, 56 full texts were retained (with five further full texts from reference lists of these full texts). This left 61 full texts that were assessed according to the inclusion-exclusion criteria and 14 full texts were retained for data extraction. Figure 2.1 shows the study flow chart. The majority of excluded full texts did not report self-reported duration or present a comparison of GPS and self-reported journey duration.

Figure 2.1 Study flow chart -selection and screening process for included studies

Image subject to copyright

Original available at

http://www.tandfonline.com/doi/abs/10.1080/02640414.2011.609897#.UoOV5nCWZo4

Description of studies

The 14 included studies were from the USA (n=5) [88, 103, 145-147], Australia (n=3) [61, 148, 149], the UK (n=2) [99, 100], Canada (n=1) [89], Sweden (n=1) [150], Switzerland (n=1) [151] and South Africa (n=1) [152]. In terms of sample size they ranged from a study of five households [148] to an 897 participant study [99]. Data collection was mostly conducted over a single day, although some studies had a collection period of up to one week. A range of GPS devices were used and self-reported travel was by recall diary, face to face interview or CATI (Computer assisted telephone interview). The studies assessed a variety of modes and combinations of modes: driving [88, 89, 103, 145, 148, 150]; walking: [147]; walking and driving: [61]; all modes: [99, 100, 146, 149, 151, 152].

I identified two distinct techniques for comparing and presenting the differences between GPS and self-reported journey duration; (1) **Matched-trips** analysis where only the trips that could be identified as the same from both methods were retained for analysis (e.g. from 100 self-report and 125 GPS trips, 89 could be identified as the same trip from both methods, and were used for analysis, with the others discarded); and (2) **Unmatched-trips** analysis where GPS and self-report data sets of the same study sample were analysed independently before comparison (e.g. 100 self-report trips compared to 125 GPS trips collected over the same period). These are presented in Figure 2.1. Only two studies presented both types of analysis [100, 148].

Figure 2.2 Representation of analysis types

Image subject to copyright

Original available at

http://www.tandfonline.com/doi/abs/10.1080/02640414.2011.609897#.UoOV5nCWZo4

Of the included studies, six compared GPS data that was not comparable to the self-reported data; the self-report was from a different sample population or data collection period to the GPS data. These studies were included in the quality assessment but excluded prior to analysis as the differences between measures will include temporal and sample differences in travel behaviour, and will not isolate the difference of interest.

Quality scores

The quality scores for the studies ranged from 2/10 to 9/10 (see Table 2.2). There was 94.9% agreement on quality criteria between raters (myself and Dr Krenn).

Table 2.2 Quality criteria of 14 studies retained for data extraction; ordered by date

	Criteria							**Total Score**
Study	**Year of study (0/1)**	**Popula-tion size (0/1/2)**	**Measure-ments (0/1/2)**	**Data Treat-ment quality (0/1/2)**	**Representa-tive sample (0/1)**	**Travel modes (0/1)**	**Peer review (0/1)**	
Murakami (1999) [88]	0	2	1	2	1	0	1	**7**
Bachu (2001) [145]	0	0	0	1	1	0	1	**3**
Stopher (2002) [148]	0	0	0	2	0	0	1	**2**
Wolf (2003) [103]	0	2	2	2	1	0	1	**8**
Schoenfelder (2003) [150]	0	0	2	2	0	0	1	**5**
Stopher (2007) [61]	1	2	1	2	1	1	1	**9**
Schuessler (2008) [151]	1	2	2	1	1	1	1	**9**
Krygsman (2009) [152]	1	1	2	2	1	1	1	**9**
Bricka (2009) [146]	1	1	1	0	1	1	1	**7**
Stopher (2010) [149]	1	2	1	2	1	1	1	**9**
Blanchard (2010) [89]	1	1	1	1	0	0	1	**5**
Cho (2010) [147]	1	0	0	2	0	0	1	**4**
NTS Pilot (2009) [100]	1	2	2	2	1	1	0	**9**
NTS (2011) [99]	1	2	2	2	1	1	0	**9**

The difference between self-reported and GPS measured trip duration

Table 2.3 shows the 12 results from the eight included studies (two studies had both analysis types presented [61, 100] and one study presented independent results from three locations [103]). All studies showed self-reported journey times were greater in duration than GPS measured times. The difference between self-report and GPS time per journey ranged from +2.2 to +13.5 minutes and from +9.2 to +75.4% of GPS measured time. Standard deviations were reported for five studies and used to calculate 95% confidence intervals on the difference between means; for four of these results the difference between mean durations was significant at the 95% level. One study did not report standard deviations, but reported that the difference between means was significant at the 95% level based on their own analysis. Table 2.4 shows the characteristics of the included studies.

Table 2.3 Included studies (n=8) by date that compare GPS and self-reported journey duration, absolute difference (minutes) and percentage difference (GPS comparator)

Study Result	**Analysis**	**Result – average journey duration (min)**		**Difference per trip**		
		GPS	**Self-report**	**Absolute (min)**	**95% CI of difference**	**Percentage**
Bachu (2001) [145]	Unmatched	10.1	12.3	+2.2	2.2	+21.8
Stopher (2002) [148]	Unmatched Matched	17.9 27.8	31.4 31.4	+13.5 +3.6	- -	+75.4 +12.9
Wolf (2003) [103]*	Matched[a] Matched[b] Matched[c]	12.3 11.1 11.4	15.6 14.9 14.7	+3.3 +3.8 +3.3	1.3** 1.5** 1.4**	+26.8 +34.2 +28.9
Stopher (2007) [61]	Matched	13.0	14.2	+1.2	***	+9.2
NTS Pilot (2009) [100]	Unmatched Matched	15.0 15.0	22.0 18.0	+7.0 +3.0	- -	+46.7 +20.0
Stopher (2010) [149]	Matched	17.1	18.9	+1.8	-	+10.5
Blanchard (2010) [89]	Matched	35.4	41.9	+6.5	-	+18.4
Cho (2010) [147]	Unmatched	12.5	20.3	+7.8	4.5**	+62.4

*Wolf 2003 presented results separately for 3 locations; (a) San Diego (b) Alameda and (c) Sacramento

**Results significant at 95% level based on our analysis

***Reported to be significant at 95% level

Table 2.4 Characteristics of included (n=8) studies

Study	Location	Study population as reported by authors	Age range	Collection period	GPS unit (s) as reported by authors	Self-report measure	Travel dimension assessed
Bachu (2001) [145]	Louisiana, USA	10 households	27-62 years	1 day	Garmin GPS III+	Time use diary	Driving
Stopher (2002) [148]	Sydney, Australia	5 households	n/a	1 day	In-vehicle GeoLogger	Time use diary	Driving
Wolf (2003) [103]	San Diego, Sacramento and Almeda, USA	Data from 254 households (517 surveyed)	n/a	1 day	GeoLogger	Travel diary and CATI (Computer Assisted Telephone Interview)	Driving
Stopher (2007) [61]	Sydney, Australia	70 households recruited; data from 45 households	15+ years	1 day	In-vehicle and wearable devices	Sydney Household Travel Survey (face to face interview)	Walking and driving
NTS Pilot (2009) [100]	United Kingdom	90 eligible households; 52% agreed; 134 eligible adults; 121 agreed; 106 completed protocol	16+ years	7 days	Amtel BTT08	UK National Travel Survey (recall diary)	All modes
Stopher (2010) [149]	Victoria, Australia	58 complete households and 18 partially complete. Take up rate 12-56%	14+ years	1 day	Amtel GPS device	VISTA07 Travel diary	All modes
Blanchard (2010) [89]	Ontario, Canada	n = 61 (57% women)	67-92 years (Mean = 80.4)	1 week	CarChip E/X® (Davis Instruments) and an Otto Driving Mate (Otto; Persen Technologies Inc.)	24h daily activity diary	Driving
Cho (2011) [147]	North Carolina, USA	n = 40 adult volunteers (21 women)	Adults	4 days	Garmin Foretrex 201	2x2day trip diary	Walking

Considering the eight results from the six studies that reported matched-trip analysis all showed that self-reported journey times were greater in duration than GPS measured times. The difference between self-report and GPS times per journey ranged from +3.0 to +6.5 minutes and from +9.2 to +34.2% of GPS measured time. Considering the four results from included studies that reported unmatched-trip analysis, again they all showed that self-reported journey times were greater in duration than GPS measured times. The difference between self-report and GPS times per journey ranged from +2.2 to +13.5 minutes and from +9.2 to +75.4% of GPS measured time.

Table 2.5 displays results from the six studies that were excluded after data extraction as the data were not comparable. That is, the study sample or collection period for the GPS data did not match that for the self-report data. The difference between self-report and GPS times per journey ranged from -30.0 to +34.3 minutes and from -58.8% to +140.6% of GPS measured time.

Table 2.5 Studies excluded (n=6) as the analysis was between incomparable datasets

Study	Result – average journey duration (mins)		Difference per trip		
	GPS	Self-report	Absolute (mins)	95% CI of difference	Percentage
Murakami (1999) [88][1]	14.2	13.9	-0.3	-	-2.1
Schoenfelder (2003) [150][2]	24.4	58.7	+34.3	-	+140.6
Schuessler (2008) [151][3]	26.2	14.8	-11.4	-	-43.5
Krygsman (2009) [152] [4]	42.7	24.4	-18.3	-	-42.9
Bricka (2009) [146][5]	13.0	18.0	+5.0	0.1	+38.5
NTS (2011) [99][6]	51.0	21.0	-30.0	-	-58.8

[1]Compares 6 days of GPS data to 1 day of recall data
[2]Compares data from GPS to data from a separate National Travel Survey. Data shown for fulltime workers; also available for retirees
[3]Compares data from GPS to data from a separate National Travel Survey. Self-report duration averaged from 3 sites presented.
[4]Compares self-report data that included weekend travel to GPS data for week days only
[5]Compares data from 2 matched groups
[6]Compares data from larger sample using diary (n = 1726) to sub-sample who also carried GPS (n = 897)

Aggregated difference between self-reported and GPS measured journey duration

Table 2.6 shows the aggregated, pooled estimate of the difference across the 12 results from the eight studies included for analysis, weighted by number of trips measured. It also shows the aggregated difference separately for the four unmatched-trip analysis results and eight matched-trip analysis results. The difference between self-report and GPS is greater when unmatched-trip data are analysed (6.8 min; 46.2%), compared to matched-trip data (3.2 min; 20.3%). There was no apparent association between magnitude of over-reporting and quality score of study.

Table 2.6 Aggregated absolute and percentage difference between GPS and self-reported journey duration, for all results, for unmatched-trip analysis results and for matched-trip analysis results

Analysis type	**Average journey duration (min)**		**Difference per trip**	
	GPS	**Self-report**	**Absolute (mins)**	**Percentage**
All (n=12 results; 6,607 GPS journeys; 6,177 self-report journeys)	15.3	19.7	+4.4	+28.6
Unmatched-trips (n=4 results; 2410 GPS journeys; 1,980 self-report journeys)	14.8	21.6	+6.8	+46.2
Matched-trips (n=8 studies; 4,197 journey pairs)	15.6	18.8	+3.2	+20.3

Discussion

To my knowledge this is the first systematic review to report studies that assess the difference between self-reported and GPS measured journey duration. The results show that average self-reported journey durations are consistently over-reported in comparison to GPS measured durations. With all 12 results included, the summary difference on duration was +4.4 mins (+28.6%) per journey (with over 6,000 journeys from each measure).

The existing literature suggests a number of possible reasons why self-reported journey durations may be an over-estimation when compared to GPS. For example, respondents often under-report journey frequency and group multiple short journeys into one long self-reported journey [103, 124, 153]. This raises questions about participants' perceptions of what a single journey is, or how to report their journey behaviour appropriately. My own pilot work (to be presented in Chapters 3-5) with SenseCam has also suggested that participants may include travel related journey activities (for example, loading a car or looking for a space to rack a bike) even when instructed not to [154]. A tendency for respondents to round travel times (digit preference) has also been identified as a factor in self-reported error [93, 95].

Another possible reason for self-report error is perception; if travel is generally considered to be onerous, it is more likely to be perceived as taking longer than it actually does, and could be over-reported [61]. It is likely that some or all of these factors contribute to give the longer journey times recorded by self-report methods.

These explanations are based on the assumption that GPS is a valid and reliable measure of journey duration but we know GPS have some limitations [59, 81]. Studies by Clark and McKimm [155] and Li [156] using direct observation have shown GPS is a valid measure of journey duration in experimental settings. However, GPS is not currently a perfect objective measure in free living situations, and there are sources of error on the data collected. Decisions made by researchers regarding definitions of a journey, inclusion and exclusion criteria, and missing GPS data will affect the obtained results. Similarly, the signal acquisition or connection-lag times (cold starts) for different devices and differential data drop out due to buildings and tree cover will influence the underestimation of GPS derived journey duration [64, 105]. It is therefore possible that for some journeys, the difference in journey duration is from GPS error rather than error in the self-report measure. Participants can forget to wear GPS devices at times and therefore completely miss trips that are recorded by self-report. This would increase the apparent over-reporting in unmatched analysis.

To my knowledge (and that of my co-authors on the submitted manuscript), the systematic comparison of two different analysis techniques (matched-trip and unmatched-trip) presented here has not been conducted or reported previously. The unmatched-trip analyses show a greater range of results and aggregated over-report than the matched-trip (by a factor of two), showing the important difference between these analysis techniques. This finding is not surprising; matched trip analysis compares like journey to like journey, and discards trips not appearing in one or other method; unmatched trip analysis compares average journey behaviour from one measure to average journey behaviour from the other and includes error inducing trips that would be excluded by the other analysis method.

The excluded studies presented in Table 2.5 displayed a much greater range of values for the difference between self-report and GPS, from large negative (under-report) to large positive (over-report). It is not surprising these non-comparable data sets had greater differences than the comparable ones displayed in Table 2.3. Although these studies generally scored highly for quality, this calls into question the usefulness of non-comparable analysis for individual journey duration comparison.

Strengths and limitations

The review presented in this chapter is clearly subject to some important limitations; primarily there is not a wide publication history on this topic which has an impact on the quantity and quality of studies reviewed, and strength of the conclusions reached from interpreting the results. The main strength comes from this review being the first such attempt to appraise the available evidence, and that the results identified are consistent in their findings.

The included studies displayed heterogeneity in design and methodology, and this will limit the usefulness of the aggregated results. Importantly, Table 2.4 shows different self-report methods were used, and this may introduce variable self-report error. My subsequent pilot work with SenseCam recorded journey times has suggested error or bias varies by type of self-report [154, 157]. The scope of the included studies is a general limitation. These studies come from seven different countries (all but one being a developed, Western setting) and six of the studies only investigate one mode of travel. The "accuracy" of GPS can vary according to mode assessed [158] and this varied between studies. Ultimately, this review was able to include only eight studies for analysis and the representativeness of the findings is not clear.

Only four of the studies presented their data in a way that allowed for the statistical significance of the results to be calculated. Again, this limits the usefulness of the aggregated error result. Without limits of agreement analysis (reported in no study) we are unable to assess the extent of random error as opposed to systematic bias and this is a major limitation to this review.

In terms of the quality of the studies, two scored from 0-3, three from 4-6 and nine from 7-9. Overall the 14 studies can be considered to be of moderate to high quality, when assessed against the criteria chosen for this review and this strengthens the conclusions of this review. In a systematic review it is important to appraise the quality of the studies [159, 160] and the quality criteria were developed from those used in a previous GPS review [105]. However, due to the low number of available studies I was reluctant to exclude on the basis of quality, particularly as the magnitude of over-reporting was not associated with study quality. Further, the criteria I used are not previously validated, and boundary limits (e.g. on sample size) were chosen pragmatically, rather than with a predetermined rationale.

As with any review, there is a possibility of publication bias. Large differences may not be reported if they jeopardise current or future work, or may be kept in-house for privacy reasons. Furthermore, I was unable to include a number of validation studies that only reported correlations, even though this is an inappropriate test of error or agreement between measures [161].

Later in my thesis (when analysing my main data set) I realised it was a limitation to only investigate journey duration. This is just one potential source of bias when investigating travel behaviour [162]. Another important and related component of travel behaviour is journey frequency or trip rate which is said to be underreported [103, 124, 153]. It has been suggested that over-reporting of duration is a function of under-reporting journey frequency (trip-chaining or reporting multiple short trips into fewer long trips). However, over-reporting was also evident in matched analysis studies suggesting that self-reported trip-chaining cannot account for all duration over-reporting.

Conclusions from this review

This review demonstrates that there is a disagreement between self-reported and GPS measured journey time; it is consistent in direction, but variable in magnitude which means that the two measures should not be used interchangeably. It also means that based on the assumption GPS is the more objective measure, studies using self-reported journey duration over multiple journeys and days are likely to be substantially over-estimating the travel time or exposure to active and sedentary travel. Whether this is in transport modelling, or to calculate the association with a health outcome, these results will have been attenuated by the error on self-report.

The results from this review also suggest that any study relying on self-report to assess journey duration should go beyond coefficient of validation calculations, and seek to quantify the error on their measure with a more objective measure (such as GPS or SenseCam). Continuous, on-going studies (such as the NTS) considering a move from self-report to GPS based measurement should carefully consider the implications of the difference between measures.

This review also demonstrates that, at present, studies reporting these data are sparse. Studies measuring travel duration with self-report and an objective measure such as GPS should report their data in such a way that a more comprehensive review can take place in the future.

Regarding matched and unmatched-trip analysis, the choice of analysis method clearly varies between studies and researchers. The chosen method seems likely to affect the error, or over-reporting, detected. Unmatched analysis shows the total discrepancy between what people tell us in self-report measures and what the GPS measures over the period of data collection. However, the differences can be due to the amount of data collected, rather than errors of participant report. The matched tells us how much people misestimate only what they remember and report doing and is usually only possible on a smaller proportion of the data due to the application of filters and data inclusion criteria.

Future study from this review

As GPS technology improves, and more studies are published, I would recommend further analysis of the bias on self-report data. This should investigate by study quality, study method, different types of self-report measure: questionnaire, diary-log, and interview, and by different mode of travel, particularly sedentary versus active modes for health intervention work.

Furthermore, I believe the large range of bias detected in this review supports testing other methods for assessing the bias on self-reported travel behaviour. As presented in the previous chapter, I think SenseCam has considerable potential to allow a better assessment of the error on self-report than GPS. The next few chapters will present my attempts to develop and test a suitable method.

Chapter conclusion

This chapter was an attempt to summarise and appraise the available evidence on a question that is central to my thesis; namely, how does self-reported journey data compare to objectively (GPS) measured journey behaviour. This was in response to my first thesis objective. I found self-reported journey duration was over-estimated (by variable amounts) in all eligible studies.

The rest of this thesis will describe how I attempted to contribute new evidence to this topic through developing a new method to measure travel behaviour. This responds to my second and third thesis objectives.

Part 2 – Developing a SenseCam method

CHAPTER 3 – A FRAMEWORK FOR DEVELOPING A SENSECAM METHOD

Summary

The Chapters in Part 2 aim to address my second thesis objective; namely developing a method to use SenseCam in travel measurement. This chapter assesses the appropriate methodological and theoretical steps to establish SenseCam as a feasible, valid and reliable measure so it may be used to investigate the error on self-report. In this chapter, I propose my own validity and reliability framework developed from the existing health behaviour measurement literature.

Introduction

I will start by exploring the literature and terminology associated with health behaviour measurement. I will then propose my own framework from the aspects selected as appropriate to my second thesis objective.

Feasibility

Feasibility describes whether a tool or measure can be used for its designed purpose [163]. A device may produce valuable data, but if it weighs 35 kg it will be too heavy to be worn during free-living physical activity and it would not be feasible for a physical activity measurement study (see early life-logging devices which resemble a desktop computer worn in a harness [132]). Feasibility can be assessed by conducting field testing and small scale "proof of concept" studies. These should assess participant feedback (e.g. satisfaction and burden) as well as data issues. For example, can meaningful information be produced in reasonable time frames?

Validity

Validity refers to the extent to which a test result is representative of the true scientific value. In this scientific context I use "true" to mean an exact representation of what happened, free from all possible sources of error or bias. I found the following educational literature definitions of validity useful:

- the degree to which tools accomplish the purpose for which they are being used defined by Worthen et al., 1993 [164]

- the extent to which certain inferences can be made from test scores or other measurements defined by Mehrens and Lehman, 1987 [165]

Validity is said to be a distinct construct from reliability (discussed below) and one that is more difficult to measure [166]. Validity can refer to Test validity and Experimental validity, themselves consisting of a number of components. The following section will outline the common sub-types and definitions used when discussing behavioural measurement with human participants. Within the literature some terms are used interchangeably and different categorisations are used. I will present some commonly accepted definitions and categorisation, taken primarily from two sources directly relevant to my Thesis; the MRC Diet and Physical Activity Measurement Tool Kit [163] and the Social Research Methods Knowledge Base [167].

Test validity

Test validity is the degree to which a tool measures what it claims to measure. Part of the difficulty in validating physical activity (as well as other health behaviour) measurements is because the 'truth' is very hard (if not impossible) to know with absolute certainty. This is due in part to a lack of exact biomarkers or easily quantifiable "packets" of the behaviour. Poor overall validity is often a result of systematic errors in the measurement tool or technique [163].

The primary types of test validity as defined by the MRC Toolkit are **Content**, **Construct** and **Criterion**. It is this model of validity that I will use for my thesis as it has been defined from the point of view of physical activity and nutrition measurement. However, it is worth noting that other models exist; for example, Patterson et al argue that criterion and content validity are really estimation strategies for construct validity [168].

Content and Face validity

Content validity (also called logical validity) was defined (in a psychological testing context) by Anastasi and Urbina as a *"systematic examination of the test content to determine whether it covers a representative sample of the behaviour domain to be measured"* [169]. The MRC Toolkit defines it as how well the "*domain intended for measurement*" is covered by or tied to the test or measure. It is important to consider if any aspects of the behaviour being measured are over or under represented, for example if we consider physical activity, does the measure capture exercise well, but occupational activity poorly.

Face validity is a distinct concept, but similar to Content validity. It is taken to mean a measure looks like it will do what it is designed to do, rather than it has been demonstrated or proved to do so.

Construct validity

Construct validity regards whether, and to what degree, a tool measures what it purports to measure. For my thesis, this means to what degree SenseCam measures travel behaviour(s). The following variants of construct validity can be found in the literature:

- **Convergent validity** is whether a tool is correlated to another tool that it should theoretically be related to. This is examined by having two or more different instruments measure the same construct. A high degree of agreement or concordance between instruments indicates good convergent validity [170]

- **Discriminant validity** (also Divergent validity) conversely this examines if two measurements that are supposed to be unrelated, are in fact, unrelated

- **Nomological validity** is whether the concept or measures being tested behave as they should with other related concepts or relevant variables

Criterion validity

Criterion validity concerns whether a tool correlates to a criterion tool already held as valid [163] (making criterion validity analogous to convergent validity with a gold standard or criterion measure). There are two common types:

- **Concurrent criterion validity** is comparison of one measure, to a measure previously shown to be a gold standard or criterion measure, when the data are collected at the same time. When this "gold standard measure" is a truly objective measure (e.g. double labelled water for energy expenditure) we are said to be assessing absolute validity

- **Predictive validity** is the extent to which a result from one measure predicts a result from a criterion measure when the data are collected at a different times

Other terms

Other validity terms exist in the literature but are not discussed here. For example, Comparative validity can be used to describe concurrent or convergent typologies [171]. Further, Representation validity or Translation validity, concerns the extent to which an abstract theoretical construct can be turned into a specific practical test and are not discussed in this chapter.

Experimental validity

Experimental validity refers to the wider study design and procedures. Experimental validity is comprised of two sub-groups:

Internal validity

Internal validity is the extent to which conclusions drawn from the experimental data are free from confounding issues which cause bias such as reactivity and missing data; methodological quality. It is most often used in the context of causal relationships.

External validity

External validity is the extent to which conclusions drawn from the data are generalizable to the larger population (e.g. do not exhibit sample bias)[9]. Age, sex, ethnic origin, socio-economic status may all limit generalizability (MRC website). For example with SenseCam, while the camera will perform the same whoever wears it (i.e. it will take photos automatically), the way it is worn in terms of compliance and wear-time may vary by individual or population group.

Assessing validity

The different types of validity discussed above can be assessed in different ways, and there is debate about the most appropriate [172]. The choice may be influenced by whether the data are continuous, ordinal or categorical. Generally, validity will be assessed by a correlation coefficient, a regression analysis, a Bland-Altman analysis, or a combination of these techniques [161, 163, 172].

Reliability

Reliability[10] is a related but different concept to validity; it is the extent to which a tool gives measurements that are consistent, stable, and repeatable [163]. While high reliability does not imply high validity, a lack of reliability does place limits on a tests overall validity. Two accepted definitions of reliability[11] are;

- the degree of consistency between two measures of the same thing defined by Mehrens and Lehman [165];

[9] Ecological validity is similar to external validity, but refers to the extent to which experimental conditions accurately reflect real world conditions

[10] I am only considering absolute reliability. Relative reliability, which refers to degree to which the position of individuals in a sample is consistent over repeated measurements, is not relevant to my research objective (Baumgartner TA: Norm-referenced measurement: Reliability. In *Measurement concepts in physical education and exercise science.* Edited by Safrit MJ, Wood TM: Human Kinetics Books; 1989)

[11] Again, these definitions are from the educational literature

- how stable, dependable, trustworthy, and consistent a test is in measuring the same thing each time defined by Worthen et al., [164]

Reliability is affected by random error; the literature is equivocal as to whether systematic error (bias) influences reliability. For this thesis I will assume that stable systematic bias does not influence reliability, but inconsistent or proportional bias could. The MRC Toolkit notes high reproducibility is possible due to a consistent over- or under- reporting of diet (or activity); the method can repeatedly report incorrect data in a reliable manner [173]. There are several general classes of reliability[12], and procedures to assess or estimate them.

Test-retest reliability

Also called repeatability or stability, this assesses the degree to which test scores are consistent from one test administration to the next. Measurements are gathered from a single experimenter or researcher who uses the same methods or instruments and the same testing conditions. This estimate incorporates intra-rater reliability, or the stability of the person taking or interpreting the measurement.

Test-retest reliability is commonly assessed by the Coefficient of variation (standard deviation of repeated test as a percentage of the mean). The Intra-class Correlation Coefficient (ICC), Pearson correlation and Spearman correlation can be used for continuous data (depending on whether the data is normally distributed). Ordinal data can be assessed by Weighted Kappa and binary data with the Kappa Statistic.

An important point is that if the behaviour varies widely between tests (i.e. has poor stability) the test-retest on that behaviour will show poor results. This can certainly be the case in when test-retest is assessed by measuring physical activity behaviour from day to day. I will address this with respect to establishing test-retest reliability for SenseCam in Chapter 5.

[12] There are other forms of reliability not presented here. For example, Internal consistency reliability is the correlation of several items in the same test that are designed to measure the same thing.

Inter-rater reliability

Also called concordance, this assesses the degree to which test scores are consistent when measurements are taken by different people using the same methods. This is usually assessed by percentage agreement for categorical variables and Intra-class correlation coefficient (ICC) for continuous variables.

Inter-instrument reliability

This assesses the degree to which test scores are consistent when measurements of the same thing are taken by different devices. In other words, would two SenseCams looking at the same journeys give consistent measurements compared to each other? This is usually assessed by percentage agreement for categorical variables and ICC for continuous variables.

A note on accuracy and precision

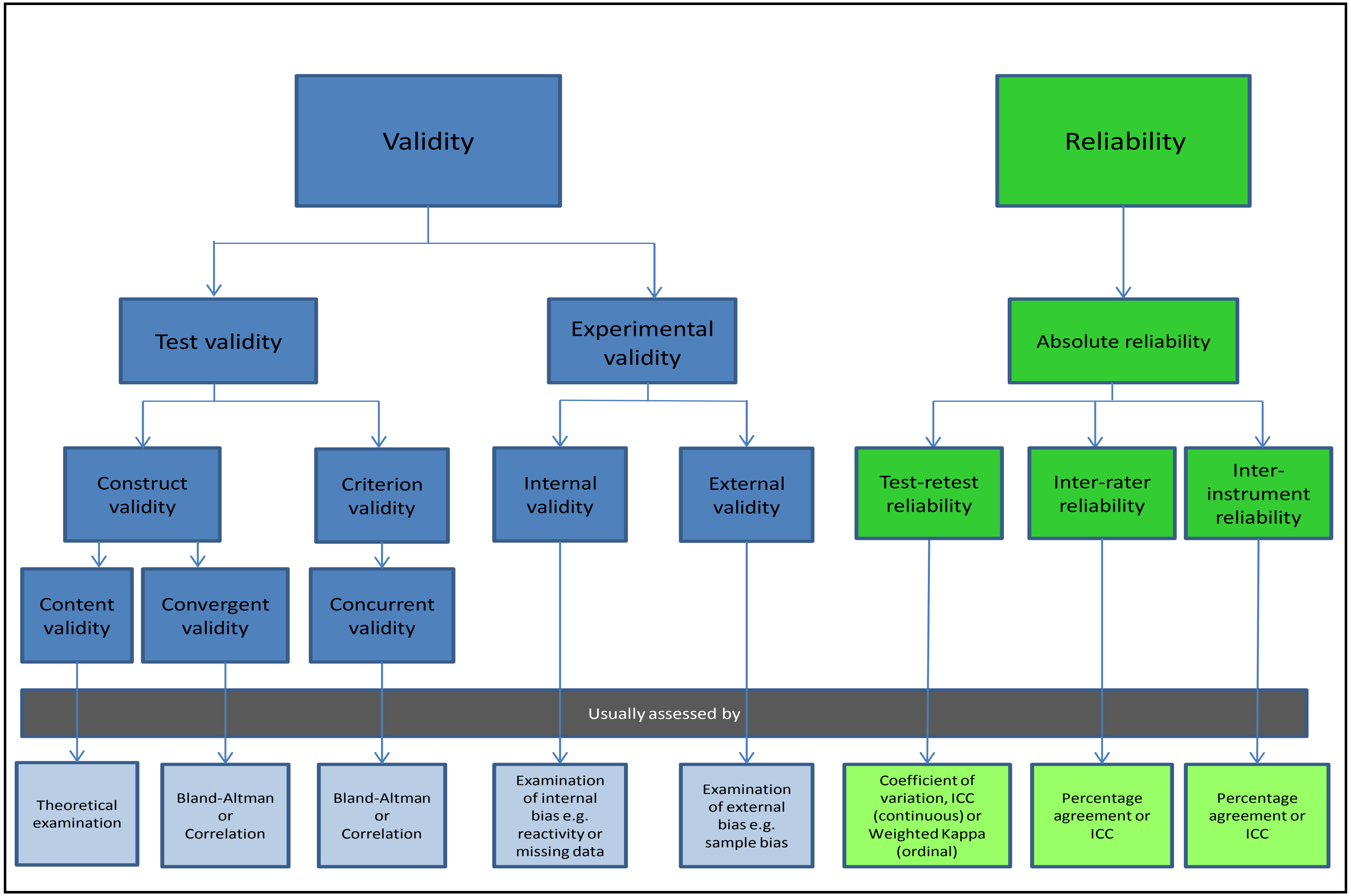
The terms accuracy and precision are widely used within the measurement literature. I have found that their use is variable; often they are used interchangeably with validity and reliability; other times they are used to represent specific aspects of validity or reliability. Hanneman reports they are often used when the authors mean bias and reliability [174] and Mantha et al., recommend the use of agreement and reliability instead [175].

Streiner and Norman discuss this issue and suggest it is both "inaccurate and imprecise" to use the terms accuracy and precision in place of validity and reliability [176]. They argue both validity and reliability are more complex and incorporate more ideas than accuracy and precision which are more specific terms. However, they suggest accuracy is interchangeable with validity when there is a "known true value". For reasons of consistency and clarity I have done my best to avoid using accuracy and precision throughout this thesis wherever possible and use agreement, bias, error and reliability.

Validity and reliability for SenseCam measurement of travel

Figure 3.1 shows the components of validity and reliability I have identified as relevant to my second thesis objective; developing a method to measure travel behaviour with SenseCam. It also shows the common tests used to assess these components.

Figure 3.1 Schematic representation of the different types of validity and reliability I have identified for my research objective



Framework for establishing the feasibility, validity and reliability of SenseCam in travel behaviour measurement

Based on the material discussed in this chapter I have developed a methodological framework[13] for establishing the feasibility, validity and reliability of SenseCam (see Table 3.1). This is based on the elements I deemed relevant and suitable for the assessment of a device used to measure travel behaviour. In the following chapters I describe how I assessed SenseCam against each part of the framework.

I acknowledge there are certain limitations to this framework. For example, it does not include predictive validity, a common technique in health research, nor internal consistency reliability, nomological validity or discriminant validity [163, 167]. While I have not selected these for establishing my SenseCam method they are well known aspects in measurement research. Nor does this framework account for important issues such as device wear-time criteria (important if we are investigating normal daily behaviour for example).

[13] I take methodological framework to mean a strategy for combining research methods and analytic techniques to answer a research question. In this case, can SenseCam measure travel behaviour well enough to investigate the error on self-report?

Table 3.1 Methodological framework for establishing feasibility, validity and reliability of SenseCam

Stage	**Process**	**Outcome assessed**
Feasibility		
Proof of concept	1. Field testing and Pilot testing of device in controlled and free-living settings	Does SenseCam generate meaningful travel data?
Test Validity		
Construct content and Face validity	2. Examination of transport and active travel literature 3. Consultation with transport and health experts	Does SenseCam measure the relevant travel behaviours, and the purported dimensions within them?
Construct convergent	4. Assessment of the correlation between SenseCam data and Self-reported data from Pilot 1 and 2	Does SenseCam data agree and correlate to that from an established measure (self-report)?
Criterion concurrent	5. Comparison to direct observation for mode and duration	Do SenseCam measurements of journey mode and duration agree with those from direct observation?
Experimental Validity		
Internal validity	6. Examination of bias such as reactivity and missing data	Are conclusions drawn from experimental data free from bias?
External validity	7. Examination of sample bias (age, sex, ethnic origin, socio-economic status)	Are the conclusions generalisable to the larger population?
Reliability		
Inter-rater	8. Assessment of correlation (ICC) between two independent researchers coding the same data for mode and duration	Can we develop a coding protocol so that different researchers interpret the same journeys and durations from the images?
Inter-instrument	9. Assessment of correlation (ICC) between two independent devices recording the same durations	Do different devices record consistent data?
Test-retest	10. Comparison to direct observation for repeated journeys	Does SenseCam record the same modes and durations over consecutive tests?

Chapter conclusion

This chapter attempted to appraise and summarize the literature on measurement feasibility, validity and reliability. I have selected the aspects relevant to establishing SenseCam as a measure of journey behaviour suitable to investigate the error on exposure to active and sedentary travel behaviours.

The framework presented provides a basis and rationale for my field testing and pilot testing. It also allows me to present the results in a logical way that takes the reader through the process of establishing the SenseCam method.

CHAPTER 4 - TESTING SENSECAM

Summary

This chapter will describe how I developed a method for using automated wearable cameras to measure exposure to travel behaviour, in terms of journey frequency, mode and duration. This is in relation to Thesis Objective 2, described previously. I will outline the process of field testing, pilot testing and criterion testing and developing the necessary procedures and protocols to use SenseCam as a measure of travel behaviour in a volunteer population.

Methods

Pre testing (prior to start of DPhil studies)

In early summer 2009 our research group obtained two loan SenseCam devices from the Sensors and Devices Research Group at Microsoft Research (Cambridge, UK). They had asked us to assess how SenseCam might be used in the field of Public health. With two other researchers (Dr Charlie Foster and Dr Justin Richards) we trialled the devices at the Iffley Road Sports Complex in Oxford, UK. We observed each other conducting a variety of activities selected from a list of 20 most common physical activities in the UK (taken from the Active People Survey conducted by Sport England). The activities included stationary cycling, treadmill walking and running (with and without loads), rowing ergometers, a range of gym and free weight activities and a range of speeds on an outdoor running track. We also tested sporting activities such as tennis and squash but did not test any contact sports such as soccer or rugby, or any water sports such as swimming. During the activities the wearer was asked about comfort and how noticeable the device was.

Self-testing (prior to start of DPhil studies)

Myself and Dr Charlie Foster wore SenseCam for various trips and journeys and half and whole days between October 2009 and November 2009. We aimed to assess the feasibility of SenseCam for identifying our own journeys for a range of common travel behaviours. These included walking, cycling, driving, car passenger, bus, train, London Underground and also when we flew to Dublin for a meeting with the Clarity Research Centre at Dublin City University (a computing group also using SenseCam). My assessment metrics were as follows; comfort; ease of wear during travel and otherwise; identification of any problems during use; and demonstration that travel behaviour (mode and duration) could be identified from the images. It was at the completion of this stage I first started to conceptualise my thesis aims and objectives.

Volunteer testing (prior to start of DPhil studies)

The next stage was to ask volunteers in my research group to wear SenseCam, so I could attempt to identify journeys of third parties (i.e. journeys I had not completed myself). I also used this stage to develop and refine instructions and protocols for wearing the device. SenseCam came with basic wear instructions developed by the researchers who invented the device and I adapted these for a travel research setting. This work took place from December 2009 to March 2010. The primary method for protocol and instruction development was in depth interviews after the device had been worn.

Pilot Study 1 – Oxford SenseCam study

The pilot study was designed as a standard measure of agreement investigation, as described by Bland and Altman [161]. In September 2010 I travelled to the National Centre for Social Research who administer the National Travel Survey (NTS) and obtained permission to base my self-report diary on their measure. Through consultation with physical activity researchers (Professor Ilse De Bourdeaudhuij, Ghent University and Professor Alan Batterham, Teeside University) with experience in similar studies using accelerometers I developed the following design;

- 20 volunteers
- 1 full day of travel wearing camera
- Completion of self-report diary for the same period
- Structured interviews about participant burden and experience

I conducted this pilot between October and December 2010. The pilot was designed to answer the following questions, as directed by the methodological framework;

- Can SenseCam be used to detect and record journey behaviour in free living volunteers?
- Can I develop a coding protocol to determine journey start and end time?
- How does the device perform compared to the self-report survey?
- What are the experiences and issues for SenseCam wearers, and how do these inform user protocol and instruction design?
- Can the data be used to estimate error on self-reported travel behaviour, and if so what are the indications from 20 participants?

The findings from these questions will be presented in the following sections, with reference to the framework of developing the feasibility, validity and reliability of the SenseCam methodology. The study has been written up in full and is published in the International Journal of Behavioural Nutrition and Physical Activity (IJBNPA) [154]. The paper can be found in Appendix 3.

This study received ethics approval from the Social Sciences and Humanities Inter-divisional Research Ethics Committee (IDREC) in accordance with the procedures laid down by the University of Oxford for ethical approval of all research involving human participants (Ref No.: SSD/CUREC1A/10-054 - see Appendix 4).

Pilot Study 2 - The School Food Journey study

The School Food Journey project was a pilot study into the use of SenseCam as a qualitative measure of environment and food purchasing behaviour in 13-15 year old school children on their journey to and from school. The Principal Investigator was Gill Cowburn (Senior Researcher in the same research group as me). It took place between March-May 2011. This presented an opportunity to investigate SenseCam measurement of travel behaviour over a five day period. I was able to incorporate a researcher led travel questionnaire (investigating journey mode and duration) into the existing research design and conduct a similar experiment to Pilot 1.

This second pilot had the following aims;

- To test the feasibility of using SenseCam over multiple consecutive days
- To test the feasibility of using SenseCam in a teenage population and school setting
- To test the inter-rater reliability/agreement of the coding protocol developed in Pilot I
- To investigate if the error on self-reported journey duration from an adult population, as shown in Pilot I was apparent in a teenage population

As with Pilot I, the findings from these questions will be presented in the following sections, with reference to the framework of developing the SenseCam methodology. This study has also been written up in full and published in the American Journal of Preventive Medicine [157]. The paper can be found in Appendix 5.

This study also received ethics approval from the Social Sciences and Humanities Inter-divisional Research Ethics Committee (IDREC) in accordance with the procedures laid down by the University of Oxford for ethical approval of all research involving human participants (Ref No.: SSD/CUREC1A/10-092 - see Appendix 6).

Criterion validity and reliability experiment – SenseCam compared to Direct observation

To establish criterion concurrent validity, I designed an experiment to compare SenseCam data to an accepted criterion (gold standard) measure. For travel behaviour this is direct-researcher observation [155, 156]. This experiment also allowed me to assess inter-instrument reliability, intra-rater reliability and inter-rater reliability.

In a controlled setting I asked a second researcher (Dr Aiden Doherty; AD) to wear a SenseCam and conduct a series of 20 journeys. Additionally I asked a third researcher (Dr Justin Richards; JR) to accompany AD for some journeys (matching his behaviour exactly). For these journeys taken together, I asked them both to wear two SenseCams to assess inter-instrument agreement. I asked them to conduct designated journeys, accompanied them myself, and timed each journey by wrist-worn stop-watch. I used a short protocol for direct observation to identify the start and completion of journeys (see Table 4.1). I then retrieved the recorded image data and coded them according to my coding protocol. I compared the observed durations and modes to the SenseCam (worn by AD) results to assess criterion validity. The data from JR's SenseCam allowed me to repeat the criterion validity measure and assess inter-instrument agreement.

Having coded the data myself, I asked AD code the same set of image data. I compared his coding to the criterion observed data as a further test of criterion validity, and I compared his coding to my coding as a measure of inter-rater agreement.

The travel behaviour was simulated free-living; I informed AD and JR where I required them to go and by what mode. For this experiment I avoided simulating "breaks" (stationary periods of three or more minutes within a journey; see Coding protocol in Chapter 5) for reasons of simplicity, but included some chained journeys (multiple modes between destinations).

Table 4.1 Direct observation protocol

	Journey start	**Journey end**
General principle for all journeys	Start of travel or motion	End of travel or motion
Walking	Transition from inside to outside or start of motion after period of stationary	Transition from outside to inside or start of stationary after period of motion
Cycling	Start pedalling	Stop pedalling and dismount from bike
Driving	Start motion in car from stationary period	Complete motion in car and journey finishes or park car and finish journey
Car passenger	Vehicle starts moving from stationary period	Vehicle stops moving and journeys finishes

Data collection took place on 16th April 2013. Data analysis took place on 17th April 2013. AD and I conducted journeys 1-7 and were accompanied by JR for journeys 8-20 (see Figure 4.1). I timed all journeys using a Casio wrist-worn stopwatch (see Figure 4.2) according to the protocol and recorded the times in a notebook.

Figure 4.1 Drs Doherty and Richards after completing journey 8

Figure 4.2 Drs Richards and Doherty against the clock prior to the start of journey 20

Figure 4.3 Dr Doherty demonstrating how to wear two SenseCams at different heights by adjusting the lanyard length

Outcomes

This section will describe the outcomes from pre-testing, self-testing, Pilot study 1, Pilot study 2 and the Criterion experiment. I will report the outcomes with reference to the methodological framework described in Chapter 3.

Feasibility

Feasibility		
Proof of concept	1. Field testing and Pilot testing of device in controlled and free-living settings	Does SenseCam generate meaningful travel data?

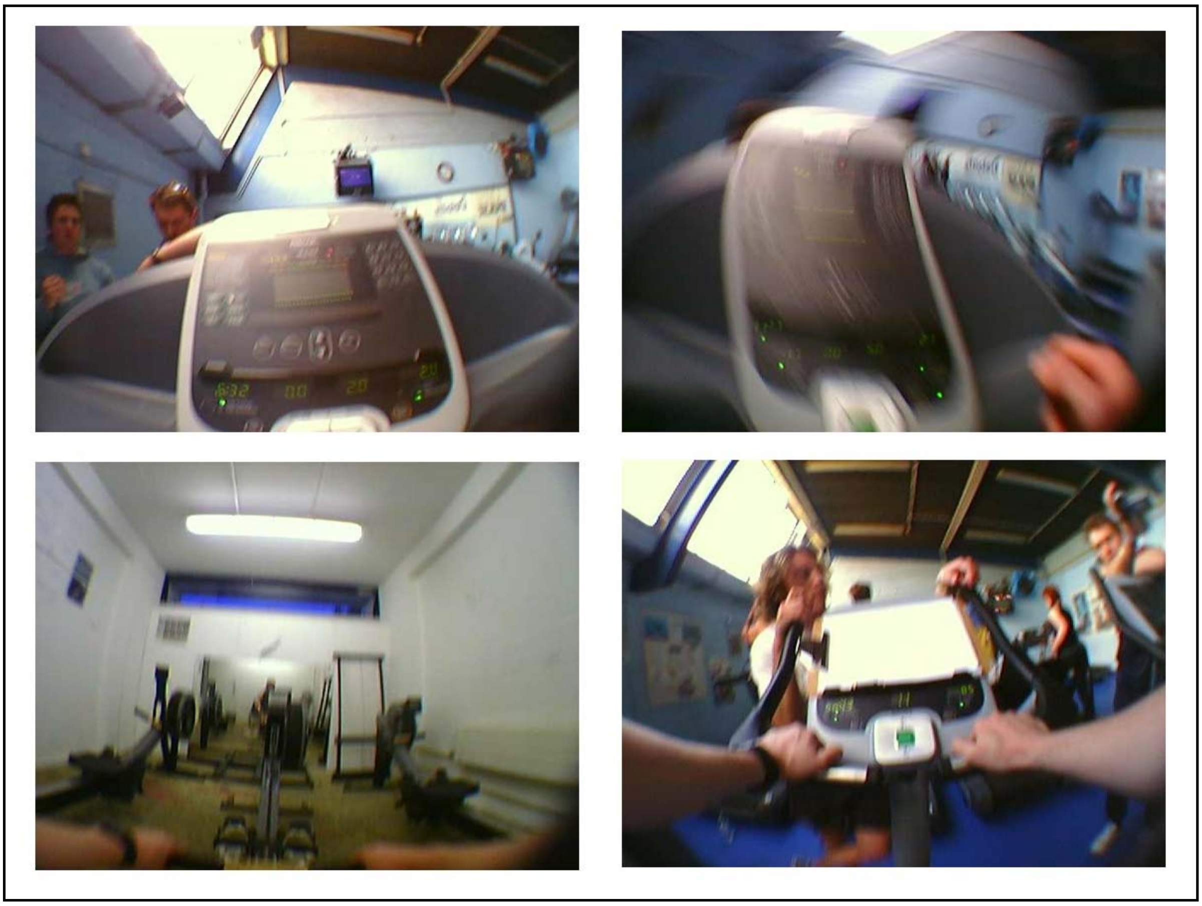
Does SenseCam generate meaningful travel data? This is concerned with the practical element of whether, when worn, does SenseCam provide images that reveal the travel behaviour of the wearer? The conceptual issue of whether these images are a true representation of the travel behaviour will be discussed in the validity sections.

Pre testing

It was immediately apparent that most vigorous activities were uncomfortable for the wearer and when the images were viewed, these activities were not easily identifiable (for example the running shown in Figure 4.4). However, stationary cycling and most walking speeds were comfortable for the wearer, and the images revealed these activities well.

The outcomes of the pre-testing led me to conclude that measurement of walking and cycling showed promise, but that other common physical activities may not be as suitable. I reported these findings to Microsoft Research in Cambridge in Summer 2009 and they encouraged myself and Dr Charlie Foster to apply for a Microsoft Research PhD Scholarship to pursue this work. This application was submitted in September 2009 and approved in January 2010 (with funding to begin in October 2010).

Figure 4.4 Field testing of SenseCam. Clockwise from top left; walking at 3mph, running at 5mph, cycling on a stationary bike at 90 rpm and using a rowing ergometer



Self-testing

Having identified the potential of measuring walking and cycling with SenseCam, I assessed the possibility of measuring other common travel behaviours. In a controlled setting, SenseCam proved able to record the travel behaviour of two researchers (myself and Dr Charlie Foster). Additional travel activities included driving, being a car passenger, taking over-ground and London Underground trains, bus travel and being an airline passenger. There were two notable findings; in a "racing bike" type position the camera "dangled" too much for comfort, and when traveling through airports the camera had to be removed during security clearance and passport control as no photography was allowed.

Upon viewing the images, I determined that my own travel behaviours could be successfully identified and journey -start and -end time determined, from the time stamp the device records for each image.

Volunteer testing

The self-testing process was repeated with volunteers from our research group. Five volunteers wore SenseCam for a day, and again the device successfully captured images that allowed me to identify the volunteers' journeys. This showed that SenseCam images could reveal travel behaviour to a third party viewer who had not completed the journey. I also used this stage to develop and test the protocols and instruction sheets, and interviewed each volunteer after they wore SenseCam for feedback and comments on what should be included in participant demonstrations and information sheets.

Pilot Study 1 – Oxford SenseCam Study

The pilot demonstrated that SenseCam can successfully be used to record travel behaviour in a sample of volunteers (n=20) over a single day of measurement. Mode, duration and number of trips were clearly identified. Figure 4.5 shows a sample of journey modes recorded during the study.

Figure 4.5 A sample of images collected during the first SenseCam pilot (reproduced from Kelly et al., 2011)

These images demonstrate the direct observation of journey mode possible from the first-person point-of-view. Clockwise from top left, walking, cycling, driving, using the London Underground and riding a local bus can be clearly seen and distinguished from each other

Pilot Study 2 - The School Food Journey Study

This second pilot demonstrated that SenseCam could successfully be used to record travel behaviour over a five day period in a small sample of 13-15 years old school children. Figure 4.6 shows a sample of travel images collected in this study.

Figure 4.6 A sample of digital images collected in Pilot study 2 (reproduced from Kelly et al, 2012)

As with Figure 4.5, these images demonstrate the direct observation of journey mode possible from the first-person point-of-view. Clockwise from top left, walking, cycling, car passenger and using public transport (bus). Note the obscured faces of the third parties. This is an ethical issue associated with SenseCam which will be discussed in Chapter 6.

Conclusion 1 - Feasibility

The results from field testing, and both pilot studies demonstrated SenseCam was a feasible measure of collecting travel data. The work also illustrated various practical issues regarding wearing an automated camera, and these are discussed in relation to compliance and external validity later in this chapter. Additionally, in the process of this work I identified some ethical issues unique to SenseCam; these are discussed in Chapter 6. That the work was granted two independent institutional ethics approvals for Pilot 1 and Pilot 2 was also important for establishing SenseCam as a feasible method.

Test Validity

Construct content validity

Construct content and Face validity	2. Examination of transport and active travel literature 3. Consultation with transport and health experts	Does SenseCam measure the relevant travel behaviours, and the purported dimensions within them?

Does SenseCam appear to measure (Face validity) and then measure (Construct content validity) the relevant travel behaviours, and the purported dimensions within them? This stage of validation is a theoretical exercise [163].

Examination of transport and active travel literature

Examination of transport and active travel literature started with the National Travel Survey. As discussed previously, the NTS is the foremost on-going travel study in the UK. Table 4.1 shows the aspects of travel behaviour assessed by the NTS. For comparison, it also shows data for two other established travel surveys, The Sydney Household Travel Survey [177] and the California Statewide Travel Survey [178].

Table 4.1 Aspects of travel behaviour assessed by three common travel surveys

National Travel Survey, UK	**Sydney Household Travel Survey, Australia**	**California State-wide travel survey**
7 days	1 day	1-2 days
Journey purpose	Journey purpose	Activity at location
Time of departure	Time of journey	Time of departure
Time of arrival	-	Time of arrival
Start location	-	Origin
End location	-	Location
Mode	Mode	Mode
Distance	Distance	-
Travel time	Duration	-
People travelled with	-	Vehicle occupancy
Make of vehicle	Vehicle type	-
Driver or passenger	-	-
Trip costs	Trip costs	Trip costs

It is apparent that SenseCam images should be able to objectively assess the following aspects of journey behaviour from Table 4.1:

- Mode (and whether car driver or passenger)
- Duration and time of arrival and departure
- Frequency

This suggests SenseCam is suited to assess exposure to travel, active travel and walking and cycling. It is also apparent SenseCam will not be suitable for objectively assessing[14]:

- Location of origin or destination (and therefore distance and speed or intensity[15])
- Trip costs
- Accompanying passengers
- Vehicle make

In terms of purpose, SenseCam will be able to assess this to some extent based on the images that precede and follow the journey (if recorded). For example images in a home before and images in an office after might indicate a commute to work. Likewise images in a shop might reveal the purpose of the journey there.

The aspects SenseCam is not suited to capture may be considered important to understanding the determinants of travel behaviour. Questions about determinants of travel behaviour are beyond the objectives of this thesis, but existing tools exist to assess these characteristics and SenseCam may have a future role as a memory or recall aid for interview purposes. This indicates that the validity of SenseCam in travel research is dependent on the research question.

Consultation with transport and health experts

In the early stages of my thesis I had a series of meetings with NatCen (National Centre for Social Research) who administer the NTS. I also met with Sustrans Research Director Andy Cope. It was agreed that SenseCam could measure the purported behaviours and dimensions of travel proposed.

I have also been able to present my findings at a number of expert meetings and research conferences. These include Health Enhancing Physical Activity annual meeting 2011 (Olumouc), Association for

[14] SenseCam will not be a criterion measure of these aspects, but could be used as a recall aid to aid better self-report of these aspects in future work

[15] Intensity is another important aspect of walking and cycling behaviour which can be estimated from distance and duration data in the NTS. While SenseCam may give better duration assessment, it cannot assess distance

Study of Obesity Conference 2011 (London), Transport Research Group, University College London 2011 (London), International Conference on Diet and Activity Methods 2012 (Rome) and International Society for Behavioural Nutrition and Physical Activity annual meeting (Melbourne 2011 and Minneapolis 2010, where I also won the Early Career Best Oral Presentation). In all cases there was discussion with the relevant experts and researchers and there was agreement that SenseCam appeared to measure what was claimed. Most concerns focussed on the practical issues around wearing or data management, or ethical concerns around privacy. These are discussed in Chapters 5 and 6.

In terms of over- or under-representing any aspects of travel behaviour, there is no theoretical reason to think that this would be the case[16]. All common journey types should be recorded to the same extent. The device should perform the same for car travel, public transport, walking and cycling. In practice, it may be that people feel more uncomfortable in public than in their own car and may be more likely to remove the device, but the different tests and pilots did not reveal any indications of this. However, the small samples and journey numbers do not allow firm conclusions to be drawn and when I complete my large study, I will investigate if there is any bias on those journeys missed or recorded.

Conclusion 2 and 3 – Construct content and Face validity

The examination of the relevant transport literature and the consultation with experts indicates that SenseCam appears to measure the components of travel behaviour that are claimed, namely frequency, mode and duration. There is no theoretical reason at this stage to suspect that any aspects would be over- or under-represented by measuring travel with this method.

[16] This is probably only true for common journey types. There may be populations that regularly travel by horse, ski, kayak or other extreme journey types that SenseCam is not able to record, but these will be a very small or non-existent fraction of the data a national travel survey is looking to collect

Construct convergent validity

Construct convergent	4. Assessment of the correlation between SenseCam data and Self-reported data from Pilot 1 and 2	Does SenseCam data agree and correlate to that from an established measure (self-report)?

Does SenseCam data correlate to that from an established (previously validated) measure, namely self-report?

Is data from Tool A (SenseCam) correlated to data from another tool that it should theoretically be related to, in this case Tool B (Self-report).

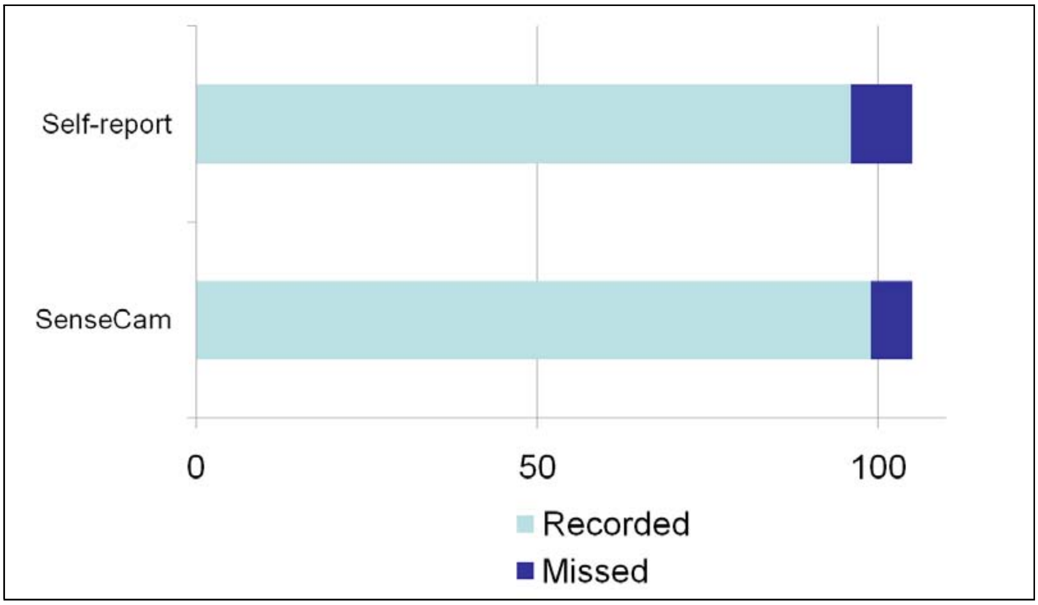
Creating travel data from SenseCam images

In the previous section I describe how I demonstrated it was feasible to wear SenseCam during a range of travel behaviours and the device would collect images of these activities. In order to establish SenseCam as a valid and reliable method of travel assessment I had to develop a systematic method of creating travel data (determining mode and duration at a journey level and frequency at a summary level) from the images collected. The Methods and Coding protocol I developed and used to do this are presented in the next chapter in such a way that they could be followed by another researcher (or research team).

Pilot Study 1 – Oxford SenseCam Study

Combining SenseCam and the travel diary, 105 separate journeys were recorded. This assumes there were no journeys missed by both methods. SenseCam recorded 99 journeys (94%) and the travel diary recorded 94 (90%). These data are shown in Figure 4.7. SenseCam and self-report identified the same journey on 88 occasions (84% agreement).

Figure 4.7 Journeys recorded by Self-report and SenseCam in Pilot study 1



Over the 88 journeys that both measures recorded, the average reported journey length was 1,064 seconds (17.7 mins) and the average SenseCam recorded journey length was 910 seconds (15.2 mins). Overall, at the group mean level, the self-reported journey durations were 154 seconds (or 16%) longer per journey (95% CI = 89 to 218 s; 95% limits of agreement = 154 ± 598 s (-444 to 752 s)). Ordinary least-squares linear regression was used to derive the correlation (validity) coefficient and a strong correlation between the two methods was apparent ($r=0.92$, $p<0.001$).

These results indicate that each measure misses a small proportion of journeys the other records; further, there is a small positive bias on reported duration. However, in the context of health behaviour measurement the percentages and correlation coefficients are high [71] and suggest that SenseCam demonstrate convergent validity for journey duration, frequency and mode when compared to self-report.

Pilot Study 2 – The School Food Journey Study

In Pilot 2, 162 journeys were detected from both measures; 150 were reported in the travel survey (12 were not reported but were shown by SenseCam); 147 journeys were recorded by SenseCam (15 lost were due to camera not worn (n=8) or the lens was obscured for the start or end of the journey (n=7)).

There were data from both measures for 135 journeys (83.3%). Table 4.2 shows the summary of journey durations and modes.

Table 4.2 Journey mode, frequency, self-reported duration and SenseCam recorded duration for journey stages (n=135) for both measures (reproduced from Kelly et al, 2011)

Travel mode	**Frequency**	**Average self-reported duration**		**Average SenseCam recorded duration**	
		Seconds	**Minutes and seconds**	**Seconds**	**Minutes and seconds**
Walk	79	886	14:46	843	14:03
Cycle	6	800	13:20	514	08:34
Car	27	484	08:04	495	08:15
Bus	23	1250	20:50	1098	18:18
Total	135	838	13:58	828	13:48

From the 135 journey stages with data from both measures, 31.4 hours of travel were reported (average stage = 838 seconds; 13min 58sec); SenseCam recorded 31.1 hours (average stage = 828 seconds; 13min 48sec) from approximately 12,000 images of travel. At the group mean level, self-reported journey stage durations were 10 seconds longer per journey (95% CI = –33 to 53 s; 95% limits of agreement = ± 501 s (–491 to 511 s)). Both the within-subjects and the between-subjects correlation between methods was strong (r=0.89 (95% CI = 0.84 to 0.93) and 0.92 (95% CI = 0.79 to 0.97), respectively).

Similar to Pilot 1, these results indicate that each measure misses a small proportion of journeys the other measure records and they do not agree exactly on the duration of these journeys. However, once again the percentage agreements and correlation coefficients are at a level that suggests SenseCam displays convergent validity for journey duration, frequency and mode when compared to self-report.

Conclusion 4 – Construct convergent validity

The results from both pilot studies demonstrate that there is substantial agreement and strong correlations between SenseCam data and self-report data for journey frequency and duration. There is almost perfect agreement for trip mode. This is a strong indication that SenseCam is measuring the same constructs as self-report, in a similar way. Therefore we can conclude that SenseCam has been shown to have high convergent validity with self-report as a method for measuring travel behaviour (mode, frequency and duration).

Criterion validity

Self-report of travel as a generalised method is well established. Some self-report tools have been previously validated or compared to an objective measure (for example, using GPS as discussed in Chapter 2). However, it should be noted self-report methods are known to have limited validity, due in part to human memory and compliance; this was discussed in detail in the opening chapter. Therefore to establish criterion validity, I designed an experiment to test SenseCam against a method considered a gold standard in travel measurement; direct observation by a researcher [155, 156].

Criterion concurrent	5. Comparison to direct observation for mode and duration	Do SenseCam measurements of journey mode and duration agree with those from direct observation?

Does SenseCam data correlate to that from a criterion measure, namely direct observation?

Experiment 1: Criterion testing of SenseCam

The observed and stopwatch recorded durations for the criterion experiment are shown in Table 4.3.

Table 4.3 Observed and recorded durations for the n=20 simulated journeys in the Criterion experiment

Journey number	Journey mode	Time (minutes) Direct observation with stop-watch	Time (minutes) Camera A (PK coding)	Time (minutes) Camera A (PK coding; re-test)	Time (minutes) Camera B (PK coding)	Time (minutes) Camera C (PK coding)	Time (minutes) Camera A (AD coding)
1	Walk	04:43	04:45	04:45	-	-	04:45
2	Car	06:51	06:49	06:49	-	-	06:39
3	Walk	04:37	04:50	04:50	-	-	04:30
4	Walk	02:20	02:24	02:24	-	-	02:24
5	Walk	02:55	02:53	02:53	-	-	02:53
6	Walk	07:14	07:28	07:18	-	-	07:18
7	Walk	10:00	09:59	09:59	-	-	09:59
8	Cycle	07:21	07:20	07:20	07:36	07:42	07:20
9	Walk	08:30	08:20	08:20	08:24	08:26	08:20
10	Walk	13:27	13:29	13:29	13:31	13:17	13:29
11	Cycle	11:24	11:34	11:24	11:33	11:17	11:23
12	Walk	07:25	07:23	07:23	07:37	07:28	07:34
13	Walk	05:43	05:48	06:00	05:55	05:54	05:48
14	Walk	03:07	03:01	03:10	03:08	03:04	03:01
15	Walk	12:35	12:38	12:38	12:57	12:45	12:38
16	Car	10:31	10:26	10:26	10:16	10:34*	10:36
17	Car	11:47	11:46	11:46	12:05	11:55*	11:46
18	Car	05:23	05:17	05:17	05:54	05:38*	05:17
19	Car	06:30	06:43	06:43	06:15	06:27*	06:32
20	Walk	10:32	10:29	11:09	10:30	10:44	10:29

*These journeys were as a car passenger as Camera C was worn by Dr Richards

Mode

For mode, there was perfect agreement between cameras and coders from all cameras for every journey undertaken. As such, there was no need for any statistical analysis on this categorical variable.

Duration

Criterion Validity 1

How well does SenseCam and my coding of data (AD; Camera A) compare to "gold standard" direct observation? Assessed by the paired samples t-test and ordinary linear regression;

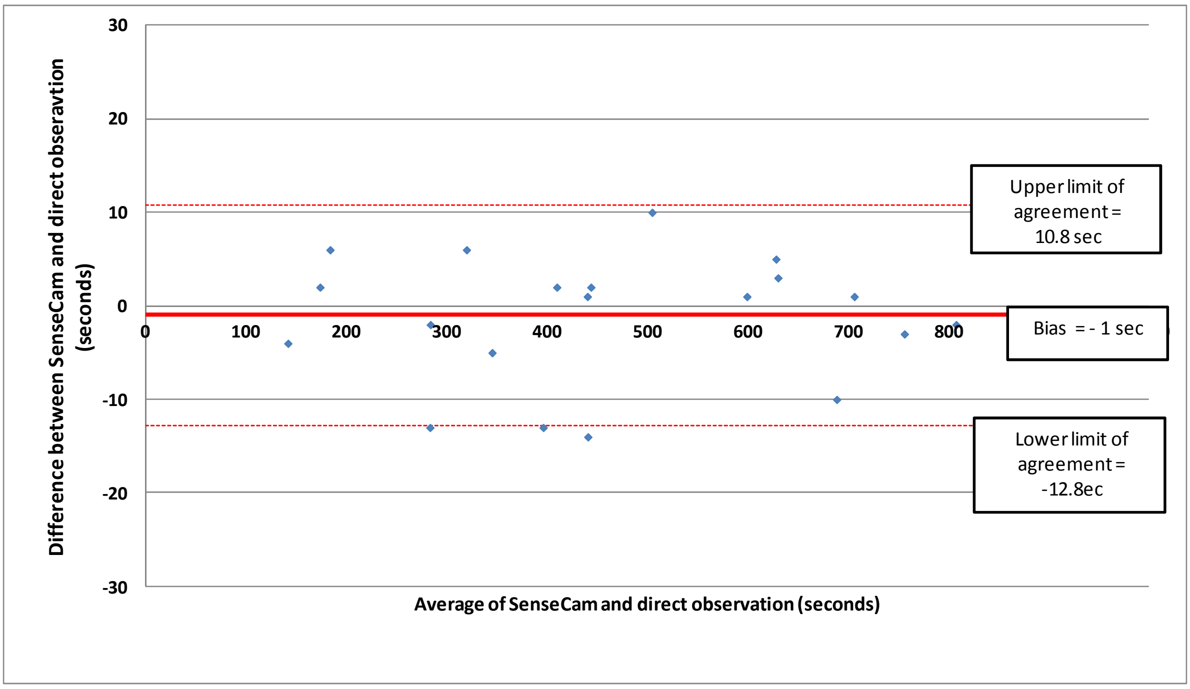
Mean difference = 0:01 minutes (95% CI = -0:01 to 0:04)

Standard deviation = 0:06 minutes (95% limits of agreement = -0:13 to 0:11)

Correlation between methods, r = 0.999

The very small mean difference and very high correlation coefficient suggests very good agreement between SenseCam and direct observation for these 20 controlled journeys. The small standard deviation and narrow 95% limits of agreement (see Figure 4.8) also shows high test-retest reliability considering consecutive tests.

Figure 4.8 Bland-Altman plot for n=20 experiment journeys showing average of SenseCam (as coded by PK) and direct observation durations against the difference (both in seconds)



Criterion Validity 2 and 3

How well does SenseCam and my coding of data (2 = JR; Camera C and 3 = JR; Camera D) compare to "gold standard" direct observation? Assessed by ordinary linear regression and the paired samples t-test;

Mean difference (Camera C) = 0:06 minutes (95% CI = -0:01 to 0:14)

Standard deviation (Camera C) = 0:13 minutes (95% limits of agreement = -0:19 to 0:31)

Correlation between methods, r (Camera C) = 0.997

Mean difference (Camera D) = 0:04 minutes (95% CI = -0:01 to 0:09)

Standard deviation (Camera D) = 0:09 minutes (95% limits of agreement = -0:14 to 0:22)

Correlation between methods, r (Camera D) = 0.999

These data were from a second and third camera on a second wearer (JR) for 13 controlled journeys. The mean differences and correlation coefficients are slightly worse than from Camera A but still suggest very good agreement between SenseCam and direct observation. The small standard deviations and narrow 95% limits of agreement also show high test-retest reliability considering consecutive tests. Bland-Altman plots were created but are not shown as they closely resemble Figure 4.8.

Criterion validity 4 (new coder; AD)

How well does SenseCam and AD's coding of data (AD; Camera A) compare to "gold standard" direct observation? Assessed by ordinary linear regression and the paired samples t-test;

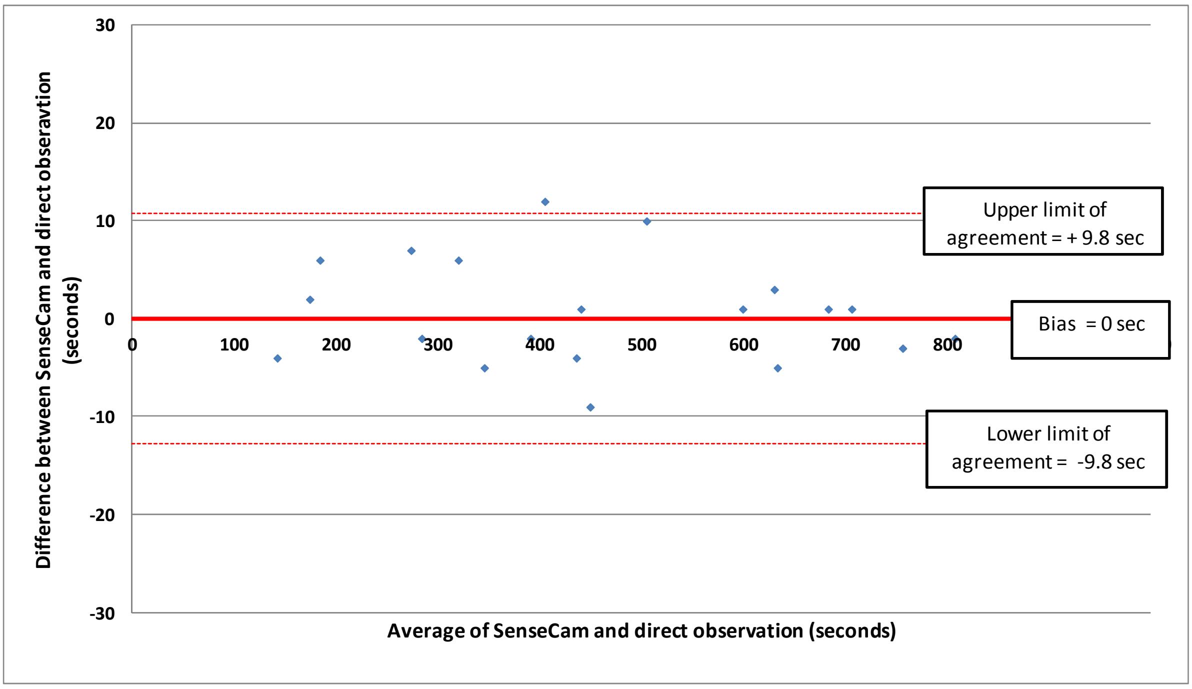
Mean difference = 0:00 minutes (95% CI = -0:01 to 0:05)

Standard deviation = 0:05 minutes

Correlation between methods, r = 1.000

The zero mean difference and the correlation coefficient suggest near perfect agreement between SenseCam data (as coded by AD) and direct observation for these 20 controlled journeys. The small standard deviations and narrow 95% limits of agreement (see Figure 4.9) also show high test-retest reliability considering consecutive tests. This coding was conducted by AD (using my protocol) and suggests that criterion validity for this method is not reduced by introducing a new coder.

Figure 4.9 Bland-Altman plot for n=20 experiment journeys showing average of SenseCam (as coded by AD) and direct observation durations against the difference (both in seconds)



Conclusion 5 – Concurrent criterion validity

The results from the criterion experiment demonstrate that in a controlled journey setting my SenseCam method is a valid near-criterion measure of journey mode and duration, as defined by direct observation. Each test contains multiple journeys, and the 95% limits of agreements are narrow, indicating good test-retest reliability. This will be discussed in the next sections[17]. That SenseCam is not agreeing to the nearest second is likely a function of the sampling rate of the camera (taking photos every 10-30 seconds depending on the activity of the wearer) but this does not introduce any substantial bias. This will be discussed in Chapter 7 – Study design, as it informed me about the required sample size for confidence on my findings when designing my main experiment.

[17] As the journeys were not the same journey conducted multiple times, this is not true test re-test reliability, and it is not the only aspect of reliability which is important. These will be assessed in the Reliability section of this chapter

Experimental validity

Experimental validity refers to the factors or variables that may influence the results, and the generalisability of the conclusions reached to wider populations. In this case it refers to any biases that influence the travel data obtained while using SenseCam. The following section will discuss factors influencing experimental validity, identified through the field testing and pilot studies, and how these may be minimised for the main study and other future work using SenseCam.

Internal validity

Internal validity	6. Examination of bias such reactivity and missing data	Are conclusions drawn from experimental dada free from bias?

The question I will ask in this section is what biases are inherent in the SenseCam method of measuring travel behaviour? This is not a systematic examination, which may only be possible when more studies using the device are completed. However, there are certain key issues that are worth considering.

Reactivity (Hawthorne effect) is when participants change their behaviour as a result of being observed or taking part in the study (in this case wearing SenseCam) [179]. There are two behaviours in question here; their travel behaviour and their reporting behaviour. In regard to the first, there is little that can be done to avoid this. However, all studies should compare their behavioural data to existing data sets to "sense-check" the data collected, and I will do this for my main study. In other words, does the self-reported data look like expected travel behaviour? I also ask participants if they think they changed their behaviour while wearing SenseCam (as part of the end of collection satisfaction questionnaire). Of course this is limited in many ways, but gives some indication of whether the participant was aware of the device or consciously changed behaviour. No-one in Pilot 1 or 2 reported changing their travel behaviour, although some said they ate less and snacked less often, suggesting some related health behaviour reactivity.

It is advantageous that SenseCam has been shown in Pilot 2 to be feasible for multiple days of data collection; this allows time for potential reactivity to lessen (acclimatisation) and I will compare reporting behaviour across the days of data collection. This is similar to accelerometer studies, although I am not planning to discard first and last days of data, which is a limitation.

In my protocols and instructions I also take care not to reveal to participants I am investigating "accuracy" or "error" on their reporting. If asked I say that the images are to "reveal further information about their journeys". This could be considered very mildly subversive, but is not pernicious to the participant, and does not breach accepted ethical standards [139].

"Lost data" while using SenseCam is another source of bias and a limitation to internal validity. Data can be lost when the camera is not worn, is removed, fails due to lack of charge, device malfunction or improper operation. Data is also lost when the images do not reveal sufficient information to determine journeys. For example, when the lens is obscured by clothing (which could be affected by the weather or season) or when light levels are too low at night.

To minimise lost data and protect internal validity, I have developed and refined participant protocols by interviewing participants at data collection. This was particularly informative during the first testing with volunteers. The aim was to make it as easy as possible (straightforward and minimal burden) for maximal data collection. The following issues were found to be important; ethical assurances and pre-empting concerns of participants with instructions for operating the privacy button; informing that data will be scrambled and participants will have the opportunity to review and delete images; emphasising the absence of audio or sound recording; providing information (and a prepared statement) to describe the study to interested third parties; appropriate wear time guidelines; operating and charging the device (with email or SMS reminders); and positioning of the device on the body (in relation to activity, body shape and clothing). Many of these issues relate to ethical research, and I developed a framework based on these findings and examination of the existing visual research literature. This is discussed in the next chapter.

Conclusion 6 –Internal validity

SenseCam based measurement is not free from bias, but I have identified some key areas which can be addressed to minimise the threats to internal validity. These include issues such as minimizing data loss and maximising participant compliance through clear demonstration and instruction to participants, as well as providing easy to follow materials. Data loss and compliance, as well as evidence of reactivity (discussed in Chapter 7) should always be assessed when using SenseCam to measure travel behaviour as indications of internal validity.

External validity

External validity	7. Examination of sample bias (age, sex, ethnic origin, socio-economic status)	Are the conclusions generalisable to the larger population?

External validity is concerned with how generalisable are the conclusions from the research and does the method have sample or selection bias? The key question is whether SenseCam research is selecting certain groups or types of individuals, or if certain groups provide more or less complete data than others (for example, by being more capable with the technology or working in environments not suitable for photography). If this is the case conclusions based on their behaviour may not be generalisable to wider populations. In the field testing and pilot studies there was no obvious evidence there were any particular groups where wearing SenseCam for travel assessment would not be feasible (or valid and reliable) and the pilots included adults and teenagers. However, the samples were very small, had few older adults (65 years and above) and were based on volunteers so it is too early to draw any conclusions from this. It will be important for me to record demographics of all participants, and discuss reasons of those who may refuse to take part after being informed about the research (i.e. decline to give consent). As a new technology, it may be that volunteers so far are "early-adopters" and different to the wider population in terms of ability and comfort with such devices. However, it is worth noting that previous SenseCam studies in computing and memory have successfully employed SenseCam in older populations and in those with cognitive impairments [138].

Finally, another important point is that for my thesis I have an objective of assessing the error on self-report. Therefore my "population" is defined by people who would consent to a self-report travel study (e.g. the NTS) already. While the volunteers for SenseCam research may not be representative of all groups, they may not be so different to populations who would consent to travel surveys.

Conclusion 7 – External validity

Considering the external validity of my SenseCam method, it is too early to draw conclusions for wider populations based on the small sample sizes used so far. Demographics and differences between groups should be assessed in future SenseCam research.

Reliability

Inter-rater agreement and reliability

Inter-rater	8. Assessment of correlation between two independent researchers coding the same data for mode and duration	Can we develop a coding protocol so that different researchers determine the same journeys and durations?

Can I develop a coding protocol so that different researchers independently determine the same journeys, frequencies, modes and durations from the image data? This is assessed by correlation between two independent researchers coding the same data for frequency (percentage agreement), mode (Kappa statistic) and duration (ICC). The coding protocol is presented in Chapter 6.

Pilot 2 – The School Food Journey Study

I coded the data from the School Food Journey Study according to the coding protocol developed for Pilot I. In order to assess inter-rater agreement/reliability, a second researcher (Alex Hamilton, DPhil student in my group) independently coded the data using the same protocol after a brief training session in which I took him through the research objective, the coding protocol and 10 example journeys from the field testing data.

The inter-rater reliability for the raters for journey mode (categorical) was found to be perfect agreement (Kappa = 1.00). Alex also identified the same 135 journey stages from the data that I had previously. The inter-rater reliability for duration (continuous) assessed by image was also very high (intraclass correlation coefficient = 0.989 (0.985-0.992)) suggesting very good agreement.

I also used feedback from Alex about unclear or unspecific aspects of my protocol and refined it for future work. This predominantly surrounded the language used, although during the 10 training journeys it became apparent I needed a clearer definition of when a journey began. Specifically, when there were prolonged short movements (e.g., walking for two frames, stopping for three frames, walking for three frames, stopping for eight frames, before beginning to walk for a continuous period).

Criterion experiment

Inter-rater reliability (agreement)

From the 20 Criterion experiment journeys I was able to assess how well my coding of data (AD; Camera A) compare to AD's coding of the same data from the same camera. This was assessed by the paired samples t-test and the ICC statistic;

Mean difference = 0:02 minutes (95% CI = -0:01 to 0:05)

Standard deviation = 0:07 minutes

Intra-class correlation coefficient = 0.999 (95% CI = 0.998 to 1.000)

The very small mean difference and standard deviation, and very high intra-class correlation coefficient suggests very good agreement and reliability between coders using my protocol for these 20 journeys.

Conclusion 8 - Inter-rater agreement and reliability

When the same data (from Pilot 2 and the Criterion experiment) was independently coded by two researchers, the inter-rater reliability was very high.

Inter-instrument agreement and reliability

Inter-instrument	9. Assessment of correlation between independent devices recording the same durations	Do different devices record consistent data?

Do different devices record mode and duration for the same journeys consistently? This is assessed by correlation of data from different devices.

Criterion experiment

How well do Camera A, B and C and their coded data (by PK) compare to each other? These data were from three cameras (on two wearers) for 13 journeys.

Intra-class correlation coefficient = 0.997 (95% CI = 0.993 to 0.999)

Conclusion 9 - Inter-instrument agreement and reliability

This result suggests there is very good agreement between devices and high inter-instrument reliability. As this incorporates my coding of the three sets of 13 journeys it also suggests very good intra-rater reliability.

Test-retest reliability

Test-retest	10. Comparison to direct observation for repeated journeys	Does SenseCam record the same modes and durations over consecutive tests?

Are the data collected from SenseCam stable over multiple tests?

Criterion experiment

In the literature this can mean if the same thing (e.g., adult height) measured twice gives consistent results. However, travel behaviour varies temporally as previously mentioned, so the "same thing" does not exist in this way. As previously mentioned the measurement of varying things (the 20 controlled journeys) showed very consistent differences between direct observation and SenseCam results. While this is not strictly test-retest, it is analogous and indicates good reliability.

Another important aspect is whether the same researcher interprets or codes the same data the same way at different time points. This intra-rater reliability is an important component of test-retest reliability. I have done this for the same 20 criterion experiment journeys from Camera A.

Intra-rater reliability (agreement)

From the 20 experiment journeys I was able to assess how well my coding of data (AD; Camera A) compared to my coding of the same data three weeks later This was assessed by the paired samples t-test and the ICC statistic;

Mean difference = 0:02 minutes (95% CI = -0:02 to 0:06)

Standard deviation = 0:10 minutes

Intra-class correlation coefficient = 0.999 (95% CI = 0.998 to 1.000)

The very small mean difference and standard deviation, and very high intra-class correlation coefficient suggests very good agreement and intra-rater reliability using my protocol for these 20 journeys.

Conclusion 10 - Test-retest reliability (Intra-rater reliability)

When the same data (n=20 journeys) were coded by the same researcher (PK) three weeks apart the agreement was very high and this suggests good test-retest (intra-rater reliability) for my method.

Limitations to Pilot tests and Criterion experiment

There are a number of limitations to my methods for testing the validity and reliability of my SenseCam method for measuring travel behaviour. As previously discussed, the samples in both studies were very small, so the generalisability of the findings e.g. that SenseCam is feasible, are not fully clear. There may be certain groups or certain places not yet tested where SenseCam is not acceptable. It is also not clear how scalable the method might be based on these two small pilots. In terms of reactivity I cannot be sure how much wearing the camera changes the behaviour (reporting or travel) of the participant.

There are also a number of limitations to the criterion experiment. While I included car driving, car passenger, walking and cycling travel modes, there were no train, tram or bus journeys. All journeys were also undertaken in perfect daylight levels. In darkness there can be some more ambiguity (2-6 frames, or 20-60 seconds) about exact start and end times. There are also only 20 journeys, and despite the very good scores for validity and reliability it may be the method fails or errs once per 100 journeys for unknown reasons.

AD and I conducted the coding. This was not unseen coding as we took part in the journeys the previous day. While this should not affect our adherence to the coding protocol, it potentially gave us an advantage on coding the exact start and end times. The experiment was also controlled, so only simulated free-living. A real day would include more incidental movement, broken journeys and potentially ambiguous activities (e.g. walking to, and then around, a shopping centre).

The limitation of this approach is that it is only applicable to the individual journeys. When it comes to assessing daily journey frequency and daily summary travel, it will be assumed that the validity and reliability of SenseCam holds. This is not strictly true for reasons of compliance, device failure and wear-time. It would be beyond the scope and resources of this thesis to test multiple whole days of free-living by SenseCam and direct observation.

Despite the claim that SenseCam is an objective measure, there is an element of subjectivity in my coding protocol. I have defined the start and end of the journeys for my direct observation from existing NTS definitions, but have not validated this procedure in anyway. Finally, I intended to use four cameras for this experiment (two on AD and two on JR), but have only presented results from three. This is because one camera failed entirely due to hardware malfunction.

Future validation and reliability testing

I have made some suggestions for how future criterion assessment of SenseCam could improve on my experiment;

- Including other transport modes
- Including travel at night and in darkness
- Incorporating unseen coding; the coders did not conduct the journeys

- Including more free-living type journeys (with variable breaks)
- Validating my direct observation protocol

Chapter conclusion

Through a battery of field testing, pilot tests and a criterion experiment against direct observation I have demonstrated the validity and reliability of my SenseCam method to measure travel behaviour. The results from the comparison to direct observation suggest that SenseCam is well suited for assessing the error on self-reported travel behaviour (my third thesis objective). However, these conclusions are based on certain assumptions, and the validity and reliability of SenseCam in free-living remains untested.

CHAPTER 5 – DATA HANDLING AND THE CODING PROTOCOL

Summary

The previous two chapters outlined a methodological framework for establishing the feasibility, validity and reliability of my SenseCam method. While presented sequentially, in practice, this was an iterative process and findings from each stage informed the next stage of testing and development. The methods I used to generate travel data from SenseCam were refined and developed throughout this process.

This chapter presents the developed procedures and protocols. This includes information on how the device and software operate, the coding protocol, and human factor inputs to interpretation that go into generating travel data from SenseCam images.

This chapter can be considered the final outcome and the practical part of Thesis objective 2: Can I develop a feasible, valid and reliable method for generating travel data from wearable automated cameras? I consider this an "original contribution" of my thesis.

Generating travel data

Once a participant has worn and returned their device (hopefully full of images), I need to generate travel data in a systematic way. These are the data processing methods and coding protocols referred to in previous chapters. This incorporates (1) downloading images, (2) viewing images in custom software and (3) annotating and coding the images. Existing protocols from computer science and memory studies existed for (1) and (2) and I was able to use these for my own travel research [130, 180]. For (3) I had to create my own coding protocol for generating journey data from the images. I developed this for Pilot 1 and after refinement tested the validity and reliability in Pilot 2 and the Criterion experiment, according to my framework.

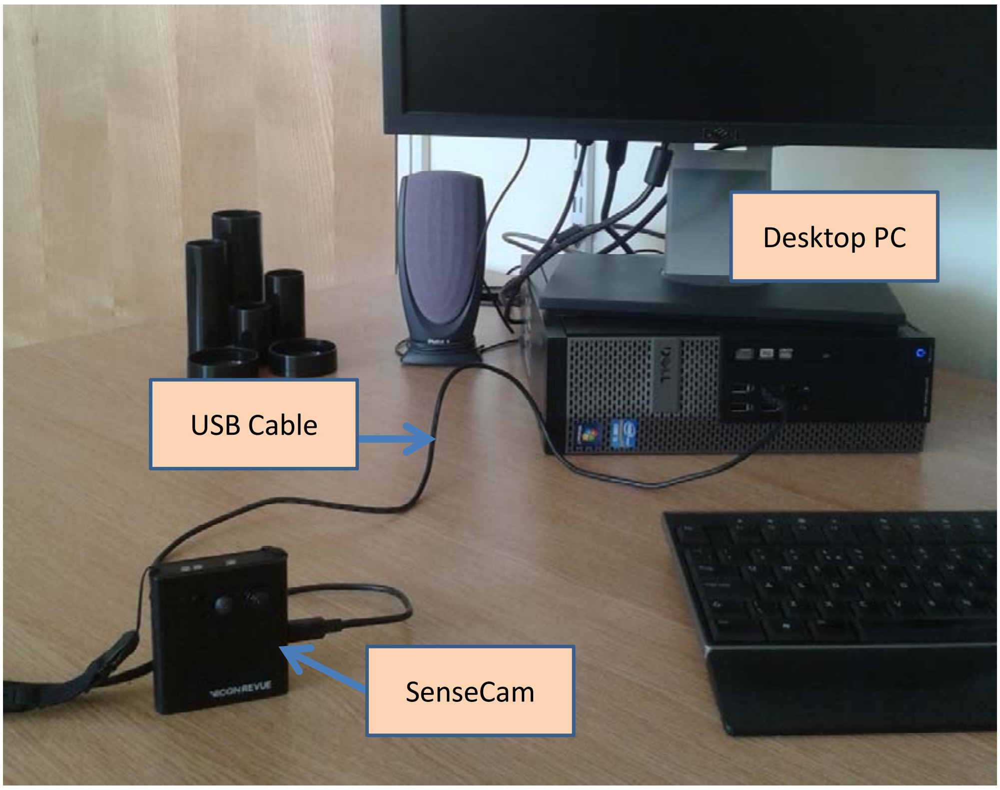
Data processing: downloading images, viewing in custom software and coding the data

How do images become journey data?

This section describes the process of generating useable travel behaviour data from returned SenseCam devices. It refers to some ethical issues not discussed yet. These are covered in the following chapter.

1. Images (and other sensor data) are downloaded from returned SenseCam to local machine by USB cable. The SenseCam should be turned on before download and connection to computer.

Figure 5.1 SenseCam plugged into desktop computer for data download



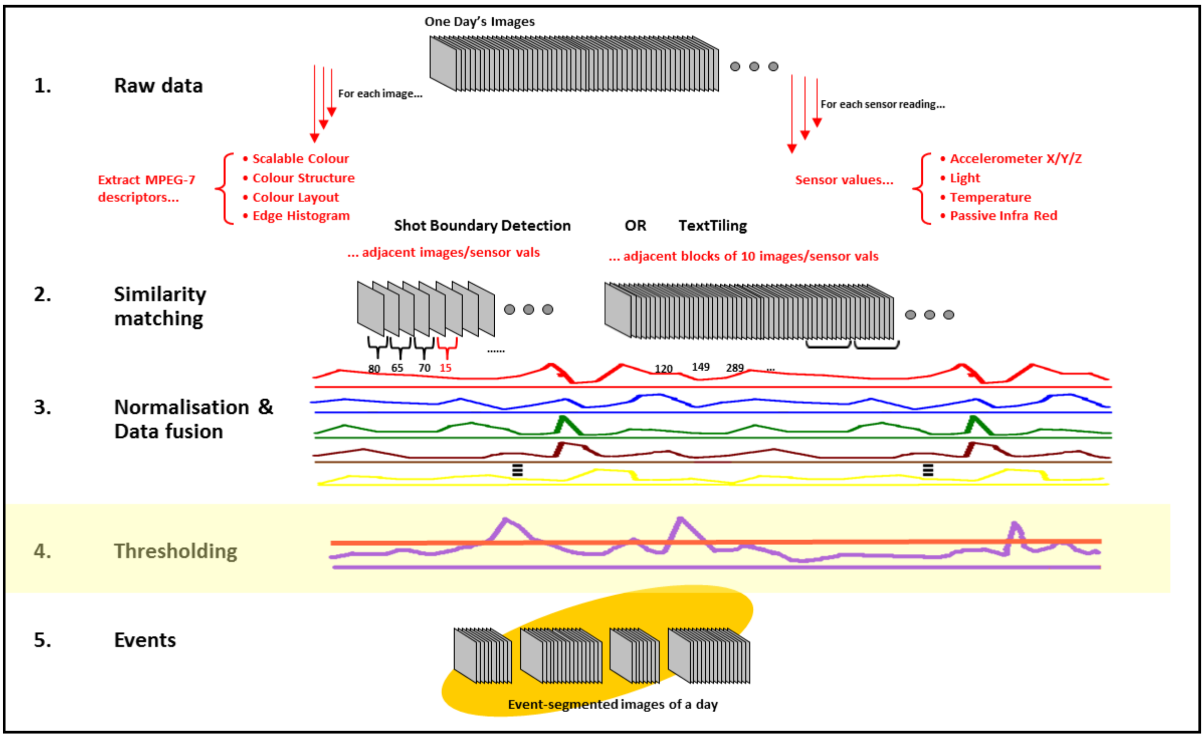
2. Images are imported using the SenseCam Browser software developed by Dr Aiden Doherty (Oxford University and Dublin City University) [180, 181]. The browser will assist the user

to begin data download. The destination folder for storing images on the local machine should be secure and password protected in accordance with ethical research practice.

The free, open-source SenseCam Browser can be downloaded at this web address: (http://sensecambrowser.codeplex.com/). Full documentation on installing and operating the software is published on the same website.

3. Once downloaded, the SenseCam Browser software uses the image, heat, light and motion sensor data to group images “intelligently” into sematic events of the same activity[18] (see Figure 5.2). Examples of activities the software can automatically detect with good results include watching television, sitting in a meeting or travelling in a car [180].

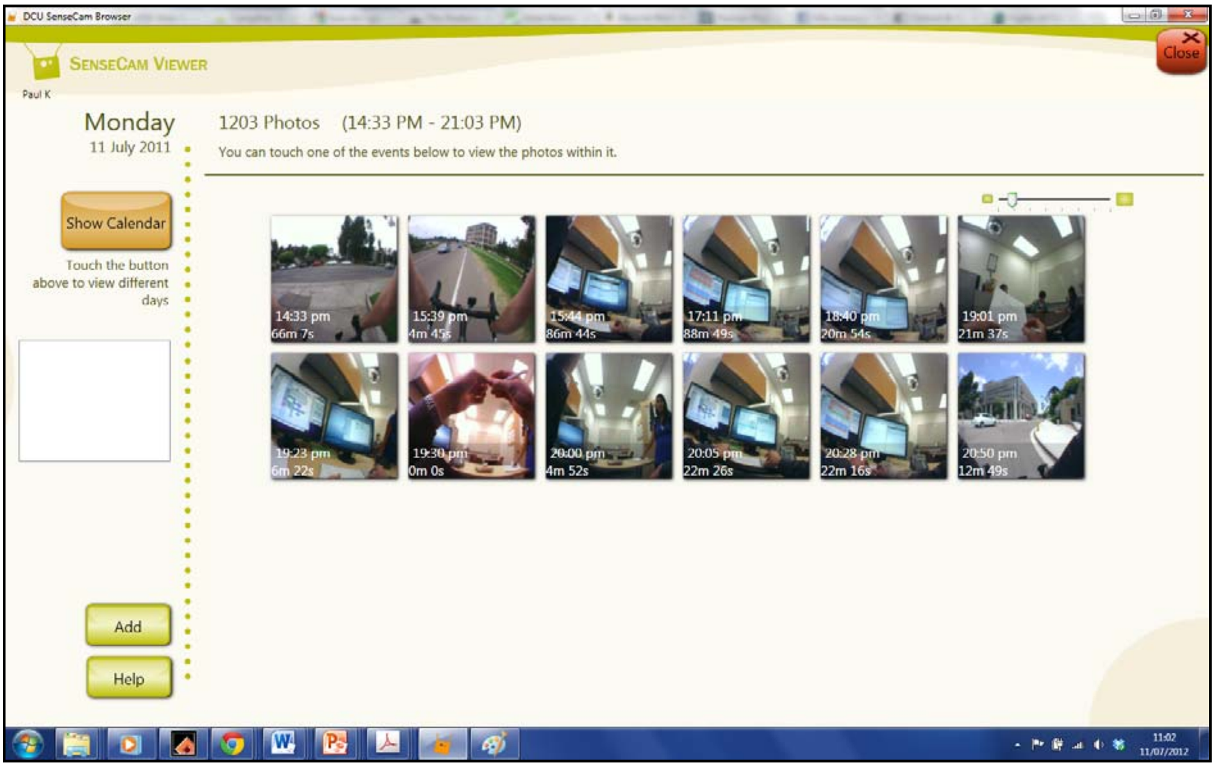
Figure 5.2 Representation of the stages of event segmentation automatically conducted by the browser software (Courtesy of Dr Aiden Doherty)



[18] This option can be deactivated according to preference. If deactivated the browser will display the entire days images un-segmented as a single event ready for coding

4. The software then displays the images on the computer screen (see Figure 5.3).

Figure 5.3 Screenshot of images segmented in events by software. On this test day the Camera was recording between 14.33 and 21.03 and captured 1203 images (one image every 10.4 seconds).



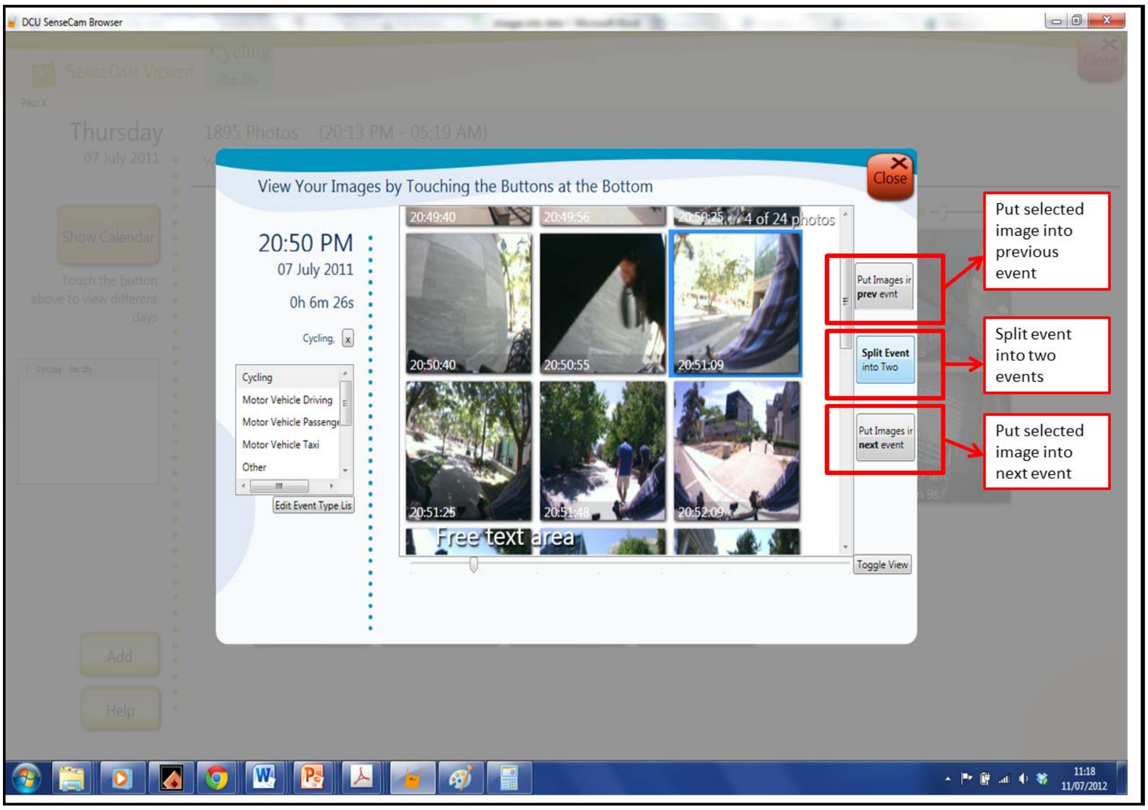
Based on the sensor data the software has segmented the images into 12 events, each shown by a single image. The viewer (researcher or participant) can view the event by clicking on an image.

5. In accordance with ethical research practice, the participant is shown how to use the browser and given the opportunity to review and optionally delete any unwanted or unflattering images [139]. This can be done by the participant in confidence, or with researcher assistance if requested. When the participant has removed any unwanted images, the researcher can begin annotation (labelling in the software) and coding (categorisation of data and meta-data prior to analysis).

6. The researcher can browse the segmented events as "miniature picture book movies" either frame by frame or as continuous playback of images (with varying speed options). Event segmentation can be corrected or edited by the researcher using a simple interface in the

browser software (see Figure 5.4). While the automated segmentation has shown considerable improvement [181] for the purposes of coding travel and travel duration, this manual step is necessary as the algorithms cannot yet implement the coding protocol with the desired level of detail for my research objectives. This step should be considered a "first pass" with the aim of grouping images into the appropriate general events; the exact start and end points will be corrected later.

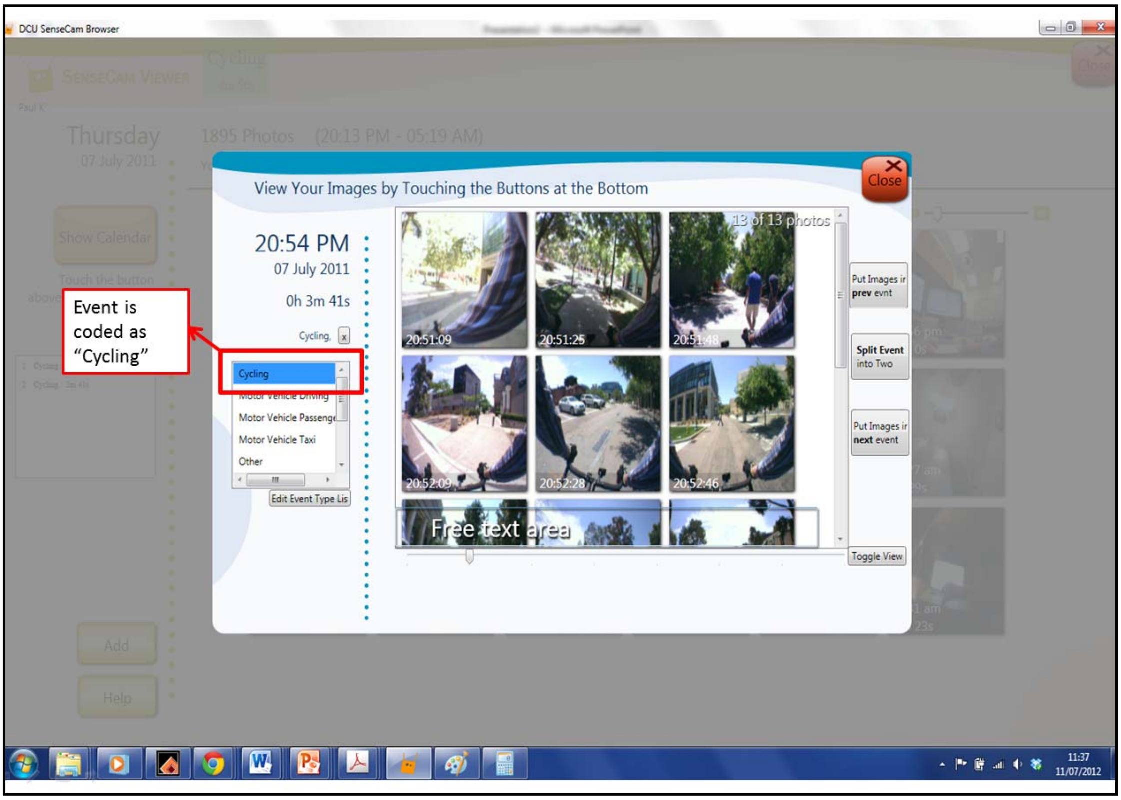
Figure 5.4 Single events (as segmented by software) can be manually corrected by the researcher



In this example the viewer has clicked on an event that includes walking and cycling which were grouped into one by the software. It can be split into two separate events by selecting the transitional image and clicking on the "Split events into two" option.

7. According to my coding protocol (see next section) the researcher can identify the events that correspond to travel and annotate them by mode accordingly from a modifiable list (See Figure 5.5). These annotations are saved by the software as meta-data.

Figure 5.5 Event being coded as “Cycling” from a list of researcher defined options



8. According to the coding protocol (see next section) the researcher can now identify and correct the first and last image of each travel event (journey) to determine journey duration. This is done as before (see Step 5) using the “Split” and “Group” options in the software. The software automatically calculates duration from the time-stamp on the first and last image for all coded events (see Figures 5.6 and 5.7).

Figure 5.6 Event display with four journeys coded

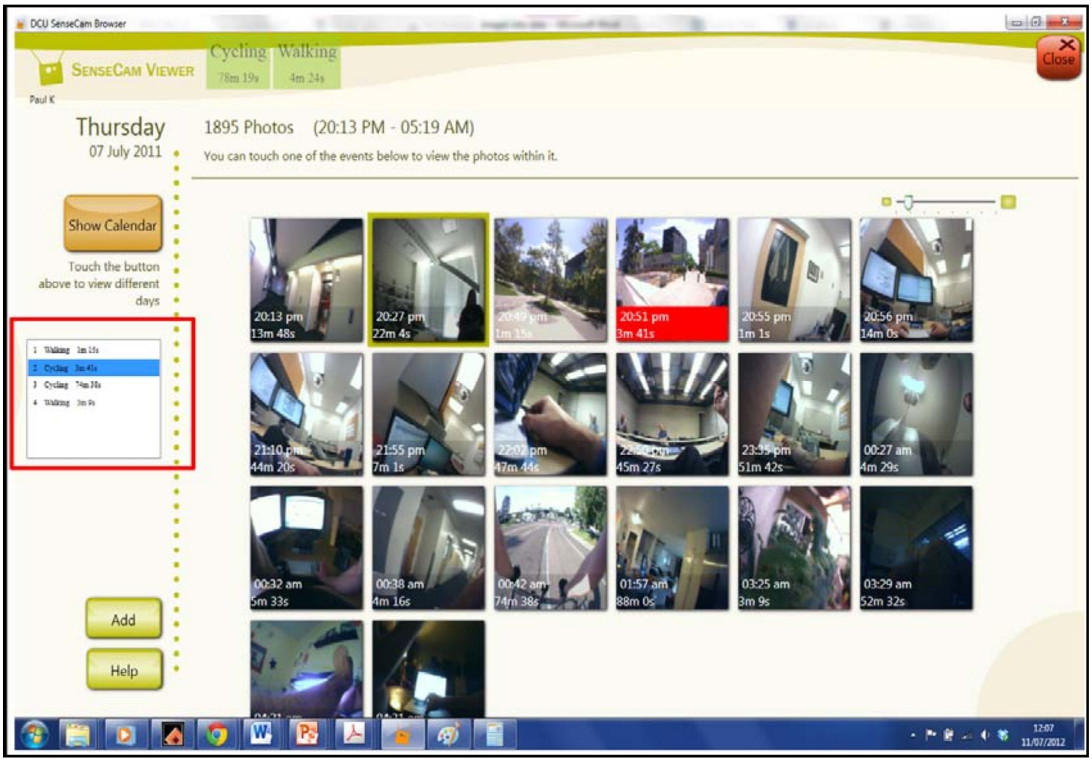
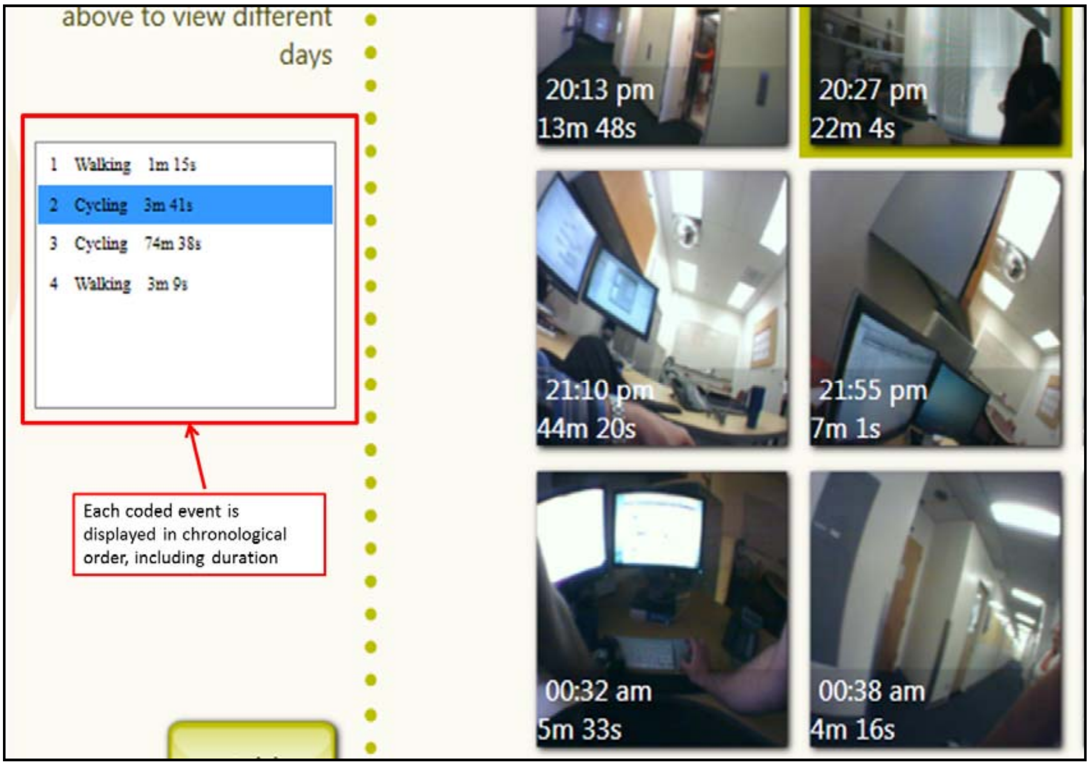


Figure 5.7 Close up of journey display



9. Once annotation in the software is complete, coded data for all participants can be exported in electronic format for analysis. The data can be extracted from the software's database using Microsoft SQL Server Management Studio Express.

Researchers in San Diego have developed a full instruction and training manual for using the SenseCam browser. This also includes coding protocols for other behaviours such as moderate to vigorous intensity physical activity (MVPA) and sedentary behaviours. This manual is available to any researcher on request.

The SenseCam Journey coding protocol

As mentioned in Chapter 1, work with GPS has raised questions about what a trip or journey actually is [86]. Text box 5.1 shows the detailed NTS definition of a journey or trip [120].

Text box 5.1 Detailed definition of a trip and journey as defined for the NTS

> The basic unit of travel, a trip, is defined as a one-way course of travel with a single main purpose. Outward and return halves of a return trip are treated as two separate trips. A trip cannot have two separate purposes, and if a single course of travel involves a mid-way change of purpose then it, too, is split into two trips. However, trivial subsidiary purposes (e.g. a stop to buy a newspaper) are disregarded.
>
> Note that in earlier publications the word 'journey' has been used. 'Trip' is now used for clarity, as the word 'journey' is often used in travel literature to mean a sequence of trips.

For this thesis I have used trip and journey synonymously, and used stage to represent component parts. The distinction for stages is important because I am attempting to match SenseCam coded data to self-reported data. I have found self-reported data is more likely to group stages into journeys, while the SenseCam can reveal independent stages. While I will therefore conduct analysis at the trip or journey level, by coding stages I can investigate grouping or chaining of stages as a potential variable influencing error. The coding protocol was designed to identify the following aspects of journey behaviour;

- Mode
- Duration spent travelling from start and end time
- Purpose where possible from images at origin and destination
- Any breaks in the journey (and whether trivial or change of purpose)

- Journeys comprising multiple stages

As previously stated this protocol was developed and refined over the course of the pre-testing, the pilot testing and the criterion testing. Some of the most useful refinements came when I was training other researchers to code the images, as I had to explicitly state my own assumptions and situations where subjective judgement might be required.

The coding protocol

As previously discussed, the software allows the researcher to split and group events and annotate them. This protocol describes how journey and trip events are identified, start and end times selected and the event annotated in the software.

The coding researcher interested in travel behaviour has six tasks; (1) to identify travel events (journeys and trips); (2) to code the events by mode; (3) to manually check and correct the start and end images of the event; (4) to code chained journeys; (5) to check for breaks in the event that demonstrate if it should be two or more sub-events (stages) and; (6) code purpose.

(1) **Identifying travel events**

The researcher is looking for purposive travel through the participant's environment as indicated by the visual data. While any given single image may be ambiguous, a number of consecutive images (including images before and after the travel event) will give context and provide insight. The different travel modes will have different visual indicators:

Motor vehicle travel

Motor vehicle or car travel can be identified by the inside of vehicles; characteristics such as seats and windscreens will indicate this. Preceding images may show the vehicle being approached and entered. The presence of a steering wheel will demonstrate the participant is the driver; the lack of a steering wheel that they are a passenger.

Public Transport

This includes travel by bus, tram, train or subway (underground). Visual indicators would be seating, other passengers and internal design of vehicle or carriage. Preceding images may show the outside of the vehicle or carriage, as well as bus-stops, platforms or descending to underground level.

Cycling

Cycling can be identified by the presence of handle bars, and in many cases cycle lanes marked on the road. Preceding images may show the approach to bicycles or bicycle racks.

Walking

This may be the most complex behaviour to identify from SenseCam images. Walking for travel has to be distinguished from other walking such as incidental movement (e.g. in the home), retail walking (e.g. within a department store) or occupational walking (e.g. movement within an office). Walking for travel should be from one destination to another, rather than walking within another context. A strength of SenseCam is that the visual information allows for these different types of walking to be identified and distinguished. The visual indicators of walking include relative movement past (or getting closer to) stationary objects such as trees, buildings or parked cars. The presence of a path, sidewalk or pavement may also indicate walking. The coder is looking for evidence of forward motion.

Other

There are other possible travel modes, for example, tandem cycling, horseback riding, and travel by boat or ferry. In the two pilot tests conducted none of these (or any other outside the existing protocol) were reported or identified by SenseCam. Any journeys that do not fit the existing criteria should be coded as “Other”. Should a previously un-code-able journey mode make a significant contribution to total journey behaviour (e.g., >5%) they can be re-analysed and the protocol amended.

(2) Code travel events by **mode**. According to the above rules, the events can be annotated in the software and coded as:

- **“Motor vehicle driving”**
- **“Motor vehicle passenger”**

- **“Motor vehicle taxi”**
- **“Public transport bus (or coach)”**
- **“Public transport train (or tram)”**
- **“Public transport subway (underground train)”**
- **“Cycling”**
- **“Walking”**
- **“Other”**

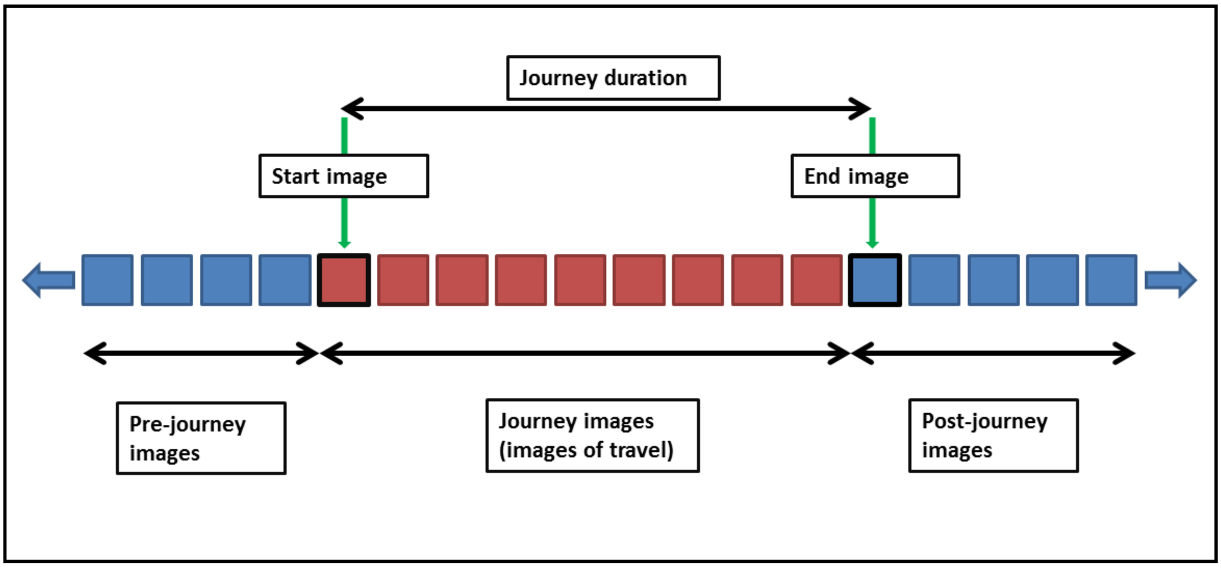
(3) Check and correct the **start and end of travel events** to calculate duration. The aim is to determine the time spent travelling, and to distinguish this from travel related activities such as securing bicycles or sending a text message from a parked car.

The general principle is to identify the first and last image of travel;

- The first image of travel should be selected as the start of the journey.
- The first image of travel having ended (i.e. the image after the last image of travel) should be selected as the end of the journey (see Figure 5.8).

Given that during travel the camera takes an image every 10-15 seconds (discussed in Chapter 7), this means that following my protocol will result in the journey losing between 0-15 seconds from the start, and gaining between 0-15 seconds at the end. This is a limitation to measuring duration by SenseCam, introducing some uncertainty. However, losing time at the start, and gaining time at the end is preferable to gaining time at both ends, or losing time at both ends, which would introduce systematic bias to the calculated duration. Further, the effect will be small.

Figure 5.8 How to select the first and last image of the travel event



From the time stamps on the selected images, the software will calculate the duration of the journey. **All journey events greater than three frames or more should be coded.** Those that last three minutes or longer will be retained at analysis (this is the minimum duration required for a journey to be reported in the travel diary – see Chapter 7).

For the different modes, the start and end images will be identified by differing visual indicators;

- **"Motor vehicle driving"**

  The start image will be the first that shows driving. The vehicle might be entered and activities such as texting, in vehicle GPS setting, or applying make-up conducted before driving commences. The first image that shows one or both hands on the steering wheel for the duration of travel will indicate that driving has commenced.

  Similarly, driving may finish and in-car activities be conducted before exiting the vehicle. The end image will be the subsequent image to the one that shows one or both hands on the steering wheel for the last time indicating that driving has ceased.

- **"Motor vehicle passenger"**

  For motor vehicle passengers, there is less information about whether or not the car is being driven. Therefore the start image is the first that shows the inside of the vehicle. The end image is the subsequent image to the last one that shows the inside of the vehicle. If there is visual information about motion from objects such as trees or buildings this should be used to refine start and end of travel.

- **"Motor vehicle taxi"**

  Same procedure as for motor vehicle travel

- **"Public transport bus (or coach)"**

  The start image is the first that shows the inside of the vehicle. The end image is the subsequent image to the last one that shows the inside of the vehicle. There is usually less visual information about motion as passengers do not always sit facing a window.

- **"Public transport train (or tram)"**

  The start image is the first that shows the inside of the carriage. The end image is the subsequent image to the last one that shows the inside of the carriage.

- **"Public transport subway (underground train)"**

  The start image is the first that shows the inside of the carriage. The end image is the subsequent image to the last one that shows the inside of the carriage.

- **"Cycling"**

  Once cycle use has been identified from handle bars and/or approach to bicycle or bicycle racks, the start image is the first that shows movement through the environment, as indicated by relative passage of stationary objects such as buildings, trees or parked cars. The end image is the subsequent image to the last one that shows movement through the environment, defined by the same criteria.

- **“Walking”**

  Once walking for travel has been identified as defined previously, the start image is the first that shows movement through the environment, as indicated by relative passage of stationary objects such as buildings, street signs, trees or parked cars.The end image is the subsequent image to the last one that shows movement through the environment, relative to stationary objects. With walking there are often indicators such as being stationary at a front door (while opening or waiting for an answer) or entering a building.

  Leaving a building is considered the start of a “walk” and entering a building is considered to be arrival at destination and the end of the “walk”. This is a limitation as walking can continue inside buildings such as hospitals, or shopping centres. Sometimes a participant can arrive at home and continue to walk around their house for many minutes. In my protocol, this walking is considered to be incidental or associated with other activities (leisure, housework, occupational, shopping, etc.) rather than in the domain of travel.

- **“Other”**

  Follow the same general principles to identify start and end images, and note the approach taken as meta-data.

Limitation: The validity of the calculated journey duration will vary slightly according to mode. This is due to the variability by which the exact start or end image can be identified. For example, motor vehicle start and end times can be more exactly identified than for train travel. This is because there is often more visual information about start of travel for motor vehicle travel. Likewise, walking and cycling start and end times can be more exactly identified than for vehicle passengers.

(4) Code **Chained journeys**

A single journey can be made of multiple chained-modes and stages. For example a chained-journey might look like:

6 minute walk – 20 minute bus – 12 minute walk

10 minute drive – 30 minute train – 15 minute walk

Joined stages should be coded to show they are sub-sections of a larger travel journey. This will allow chained journeys to be compared to single event or mode journeys that may have differential reporting error. **A single stage should be coded if it is three or more frames**. Those stages or trips lasting three minutes or longer will be retained at analysis. The dwell, or waiting times in-between (e.g., on platform or at bus stop) are not included in travel time (as per the instructions on the travel diary).

Additionally, chained journeys mean there may not be a "destination" image for any but the last stage (the final image of walking may be followed by the first image of a car travel). In these cases, the last image of walking is the end image for that stage, and the first of the car journey (the next image) is the start image of that stage. The limitation is that the preceding journey is artificially shortened by one frame. Due to the 10-15 second sampling rate during travel these cases should be rare, and introduce minimal error when they do occur.

(5) Check for **breaks in the journey** event that either should not be included in travel time or that demonstrate it should be coded as two or more stages. A single journey can be broken for incidental reasons e.g., stopping walking to talk with someone, or for a change of purpose e.g., stopping a car for fuel or to collect passengers.

A break may be identified by visual information such as the inside of shops or petrol stations. For walking and cycling the lack of movement past stationary objects will indicate a stop or break in the journey. This is also often true for vehicle journeys, though can be harder to see. Often the continuous presence of a third party in front of the camera will show a conversation and broken journey. Waiting in traffic or at traffic lights should not be considered a break.

To be classified as a break, the interruption, or clear indication of no travel movement, must be clearly identifiable for three frames or more (this selection has been made arbitrarily but corresponds to the threshold for coding a journey).

In coding, the following rules will apply;

- If the break is less than one minute there is no change of purpose it will be disregarded

- If the break is greater than one minute and less than three minutes, but there is no change of purpose, the journey or stage will be considered to be paused but not broken and duration will be a sum of the two (one before the pause; one after)

- If there is a change of mode (but not purpose) or the break is greater than three minutes, the travel event will be coded as two independent stages. If there is a change of purpose, it will be considered two independent journeys.

Matching to travel diary data

These rules are used to guide the coding of the SenseCam data. However, the self-reported data is unlikely to follow these rules exactly. For this reason, when it comes to matching there will be some subjective judgement required from the researcher. For example, the images may show a journey broken into four walking stages or trips of 11, 15, 5 and 10 minutes with breaks of more or less than three minutes. However, this may be reported as a single journey (Walking to town; purpose shopping; duration 45 minutes). In these cases, when they come to be matched they should be entered as a single SenseCam journey of 41 minutes (11+15+5+10). It will be recorded that this journey was comprised of four journeys or stages to investigate this chaining as a source of bias.

Participants may sometimes report a return journey as a single journey (Car to school and back; purpose collect children; duration 20 minutes). The SenseCam data will show two journeys of 12 and 14 minutes. In these cases the self-report data will be split into 2x10 minute journeys for analysis. Although this introduces an assumption about the reported behaviour, this is the same approach used in NTS data processing [120].

Lost data

In some cases image data during a journey or trip is not available. A lens may be obscured by clothing, the image may be too blurred, or the light level too low. These are cases of lost data. Often it is only a single image, but when the data is lost for three consecutive minutes or more, the journey should be considered incomplete on the SenseCam. The reason for the loss of data should be coded. Particularly in levels of low light, there is subjective judgement by the researcher on what can be interpreted. A good example is in very low light, in-car lights such as on the car radio or speedometer can reveal car occupancy, while passing street lights can reveal motion.

Occasionally the transition (start or end image) is lost. The journey coding should reflect this. If the "lost images" are 3 frames or less the journey is still valid, if they are more than 3 frames the journey duration is invalid and is coded as no start or no end image.

(6) Coding **purpose** also requires subjective judgement from the researcher. As the ability of SenseCam to observe journey purpose has not been validated, this data will primarily be used to help match trips to the self-report data. Using images before and after the journey (e.g., in the workplace, in a shop, in a restaurant, at home), it will be categorized in into the following purposes (adapted from the NTS); commuting, work related (journey conducted as part of working day), shopping, leisure or recreational (e.g. restaurant, gym, cinema), visit friends or other.

Analysis of journeys and stages

Whether a travel event is a single journey or a series of stages according to the definitions above will be included in the coded data for that event. For analysis all travel events will be considered to be journeys, however it will be possible to investigate those travel events comprised of chained stages independently according to a Chained trip (Yes/No) variable in the coded data.

Limitations

These processes and protocols have been developed to assess journey mode and frequency, with some information on purpose. Primarily, they allow the assessment of travel duration, for comparison to self-report. However, they rely on the images (particularly the transition images) not being lost to obstruction, low light or device failure.

While they will reveal the time spent travelling, they do not provide any information about when the participant perceives their journey to start or end. I am also limited to my "door-to-door" policy, so active travel within large buildings or long corridors is lost to other domains of physical activity.

In the previous chapter I presented evidence for the validity and reliability of this method. However I have not tested the sensitivity of the rules or decision metrics. Other choices may be preferable. For example, researchers in San Diego have created their own coding manual for travel behaviour. This is based on a four frame rule set (compared to my three frames). Future research should investigate how these choices influence the data obtained.

Chapter conclusion

This chapter outlines the processes and procedures for creating travel data from SenseCam images. The performance of SenseCam and the validity and reliability of the coding protocol has been previously reported in Chapter 4. I have presented the data methods and coding protocol here so that the reader could follow my procedures and use SenseCam to create travel data in their own work.

CHAPTER 6 – ETHICAL FRAMEWORK

Summary

This chapter describes how I developed an ethical framework for using SenseCam in the observation of health behaviours. Technological advances mean automated-wearable cameras like SenseCam are only recently feasible for investigating health behaviours in a public health context. However, existing ethical guidelines do not fully legislate for such approaches in health behaviour assessments.

Automated cameras like SenseCam are potentially very intrusive; generating unprecedented levels of image data, some of it perhaps unflattering or unwanted. The research participants and third parties they encounter may feel uncomfortable or that their privacy has been impacted. This chapter describes how I identified new ethical issues associated with observational SenseCam research and integrated suitable ethical guidelines with existing frameworks for visual research. The primary recommendations are presented in a table at the end of this chapter and concern respect for autonomy through appropriate approaches to informed consent and adequate privacy and confidentiality controls.

Published work

This chapter is based on a paper on which I was first author (An Ethical Framework for automated-wearable cameras in health behaviour research; Paul Kelly, Simon J Marshall, Hannah Badland, Jacqueline Kerr, Melody Oliver, Aiden R Doherty, Charlie Foster; American Journal of Preventive Medicine; 2013) [139]. Certain sections and passages are taken directly from the paper. The work was conducted primarily for this thesis, but I decided to write up and publish the paper prior to submission due to the level of interest and number of requests for ethical guidance. I conceived and designed this work, and led the process. SM provided specific ethical guidance; HB, JK and MO contributed previous experience with SenseCam; and CF and AD provided editorial support. In the early stages I was also advised by Professor Michael Parker (Ethox Centre, Department of Public Health, University of Oxford).

Background

The use of SenseCam as described in this thesis is exciting because it allows researchers to view participants' behaviour in a new way. The thousands of first person point of view images per day have the potential to reveal unprecedented levels and detail of information. However, this, allied to other factors raises important ethical considerations for research using such devices [182, 183]. In the course of my two pilot studies I received two ethical approvals from the IDREC at Oxford University. This was followed by successful applications by researchers at Auckland University of Technology (AUT), University of California, San Diego (UCSD), Sydney University, Durham University and Auckland University for related projects[19]. However, the question of ethics was often (if not always) raised at meetings and conferences when I presented SenseCam. Furthermore, I have repeatedly been contacted by researchers considering using SenseCam for advice and feedback on their own Institutional Review Board (IRB) applications.

As my second thesis objective was to develop a new methodology for public health focussed travel research, I determined it was an important step to formalise and make available to the scientific community a framework for ethical research using automated passive image capture devices such as SenseCam.

Methods

I developed this framework through the following steps:

1. Examination of the existing ethical literature on visual research and automated image capture
2. Identifying the aspects most relevant to SenseCam research and similar approaches
3. Consultation with other researchers using SenseCam, to identify relevant ethical issues

[19] Not all ethics applications at other institutions have been successful first time, and IRBs have raised some concerns regarding privacy and research with participants under 18 years of age. Specifically in Durham and Auckland there was an extensive list of concerns from the IRB, and while ethics approval was ultimately granted the process took many months

4. Semi-structured satisfaction interviews with participants wearing SenseCam in field testing and both pilot studies to identify issues associated with participant burden

5. Consultation with ethics specialists at Oxford University

Results

Existing literature

In the social sciences, image based research (e.g., using photography) has existed for over 100 years [83, 184]. In recent years, the British Sociological Association (BSA; 2006) and The Economic and Social Research Council (ESRC; 2008) have both published guidelines on the ethical use of images with human research participants [134]. Both sets of guidelines highlight the importance of informed consent (preferably written) for involvement in research, and for subsequent use of the image data [184, 185]. Within existing guidance there is consensus that when visual data are being used solely for observation or elicitation purposes, it is the dissemination and publication of images that create the most complex issues [185].

These guidelines were developed from the philosophy of the non-consequentionalist approach which is related to prinicipalist approaches, commonly used in public health research [185, 186]. They are based on the following principles: *Respect for autonomy* which relates to issues of voluntariness, informed consent, confidentiality, and anonymity; *Beneficence* which concerns the responsibility to do good; *Non-malificence* involves the responsibility to avoid harm; and *Justice* which encompasses the importance of the benefits and burdens of research being distributed equally [185, 186]. These are the four main principles of ethical research as considered by IRBs.

Prosser et al., subsequently codified these guidelines for visual research as follows:

- researchers should strive to protect the rights, privacy, dignity and well-being of those they study;
- research should be (as far as possible) based on voluntary informed consent;

- personal information should be treated confidentially and participants anonymised unless they choose to be identified;
- research participants should be informed of the extent to which anonymity and confidentiality can be assured in publication and dissemination and of the potential re-use of data [134, 185].

It is these guidelines and this framework that will be used when considering the new issues that SenseCam-type approaches raise.

SenseCam research as a method

Four common types of visual research data are identified: (1) found or existing data (e.g. photo albums not originally created for research purposes), (2) researcher created research data, (3) respondent created research data, and (4) representations; these can take forms such as photos, videos, film, drawings, and graphical representations [134, 185]. The daily capture of thousands of first-person point-of-view images from an automated-wearable camera is a new technique[20].

In traditional participant-created data techniques, researchers gave cameras to participants and asked them to take photographs at certain times of certain things. This requires conscious action from the participant. Automated-wearable camera techniques such as SenseCam elicit a shift from this researcher-centric construction of the social world to that of the participants [134]. Image collection is passive rather than purposive allowing researchers to see aspects of participants' lives they might otherwise not gain insight or access to. While there are similar methodologies now being employed by other research groups in diet and exercise studies [136, 187-189] there had not been any specific ethical guidance published.

Ethical issues identified with the SenseCam approach

The following issues were identified:

[20] Other devices have recently come to market. The Vicon Revue is a variant of SenseCam made under licence by Vicon. They will soon be releasing a new version call the Autographer. A company in Sweden has launched the Memoto and some researchers have started using smartphones as well.

1. Automated-wearable cameras collect many more images than traditional photography (2000-3000 in 12-hours of wear)

2. Image capture is passive not purpose; the camera records images independent of any action by the wearer

3. The participant may forget they are wearing the camera and capture images and information they would not wish or choose to have on record

4. Confidentiality may not always be possible

5. Third parties who encounter the wearer have their image captured without the opportunity to provide informed consent. These include; family members, co-habitants and friends; colleagues and co-workers; and strangers and the general public

6. Images will be seen by a large team of researchers

These issues are discussed below, with reference to existing ethical guidelines and principals.

1. Automated-wearable cameras collect many more images than traditional photography (2000-3000 per 12-hours of wear).

Traditional participant initiated photography may capture 30 images per day, perhaps up to 100 with cheaper digital cameras now available. However, by wearing SenseCam for a day, the participant is likely to capture between 2000-3000 images. By definition this will result in more information about their day being recorded (albeit potentially repetitive or benign), and is a greater intrusion on privacy, and in turn autonomy. Therefore prior to the point of informed consent, this issue must be made explicit to protect participant autonomy.

With more participant information recorded, the potential harm from a breach in confidentiality is far greater. Enhanced data security measures are therefore warranted. As lost devices pose a risk to confidentiality they should be configured so that only the research team may access them using specialist viewing software.

As with any visual research, collected images should be stored securely and password protected, according to national regulations (e.g., the UK Data Protection Act) [184] and the use of the collected images must be in accordance with protection of privacy, confidentiality, and anonymity. No image that identifies a participant by place, context, or other should be disseminated or shared without participants' express consent. Faces and identifying features should be obscured if images are used in publications or other dissemination. This is especially important in light of recent technologies like Facebook's facial recognition software [190]. To protect autonomy, participant information should be explicit about data retention, and the possible secondary analyses that may take place. Participants often enjoy reviewing their own images and request copies. Images should not be given to participants, as once the image has been "given away" by the researcher they cannot control its use (e.g., on social networking sites), or the context in which it is displayed.

It is worth noting that the nature of observation with SenseCam means that the images are aggregated to determine the health behaviours; they do not need to be displayed to support the findings. This is in contrast to traditional visual research methodologies that may use photos for illustration of social norms or events. It may not be true for future image based research or intervention using guided reflection.

2. **Image capture is passive not purpose; the camera records images independent of any action by the wearer.**

An automated camera records images without any affirmative action from the wearer. This differs from traditional photographic methods when a photo is chosen or posed. This reduces the control the participant has over the content and unflattering or unwanted images may be recorded. Examples include adjusting clothing in a mirror, grabbing a handful of food from the fridge, or browsing social networking websites while at work. Autonomy is threatened by this loss of control and this must be made clear in participant information prior to informed consent.

There are some steps that can be taken to preserve autonomy. SenseCam is fitted with a privacy button that stops image capture for seven minutes [83, 130]. Participants should be given full instruction in this procedure. Participants should also be told they can remove the device at any time or in any situation they wish. Participant information should inform them that they will not have to provide any explanation to the research team. Such situations may include trying on clothes, while doing online banking, or in social situations such as waiting outside school to collect children. This will also protect participant autonomy, even though it may affect the amount of data collected or the context of certain behaviours.

Participants should also be advised that they will get the opportunity to review and delete images at the end of data collection. This is especially important as they have consented to the collection of images, prior to knowing exactly what images and information will be recorded.

3. **The participant may forget they are wearing the camera and capture images and information they would not wish or choose to have on record.**

Despite the steps to avoid unwanted or unflattering images, it is not uncommon for participants to forget they are wearing the device. This can lead to recording of images they would not want, such as going to the bathroom, internet habits, reading confidential documents, getting undressed, or being in the presence of family members who are dressing. For autonomy, participant information should make this risk clear, and remind participants that they will get to review and delete images (in private if requested) before anyone else sees them.

4. **Confidentiality may not always be possible.**

Images retained after participant review and deletion will pass to the domain of the researcher and be classified as research data. If these images depict any illegal activities the researcher may be under legal and professional obligation to breach confidentiality and pass image data onto appropriate authorities. Plausible examples include texting while driving, drinking in a bar prior to driving home[21], or taking illicit substances. Again these are images that are more likely with a SenseCam-type technique, than traditional photographic methods. Importantly, images of criminal damage, sexual violence, and hate crime do not have the privilege of confidentiality and may be liable to subpoena by

[21] In most countries, exceeding specified blood alcohol limits while driving is the criminal act rather than the consumption of alcohol

a court [184]. Allen et al., noted this when she said “the law may not permit privacy” [191]. As stated above, participants have less control over data collected, so there is a risk of them incriminating themselves by wearing SenseCam. Participant information should reflect this.

5. **Third parties who encounter the wearer have their image captured without the opportunity to provide informed consent.**

Typically a participant will encounter many people or third parties over the course of a day. These third parties can include family members, co-habitants, friends, colleagues, co-workers, acquaintances, and individuals unknown to the participant.

These people will often be recorded in the images, either knowingly or unknowingly. The nature of the device also changes the traditional construct of taking a photograph that most people would understand. Importantly they will not have had the same opportunity as the research participants to provide informed consent to be photographed. A previous study in digital life-logging reported third parties were primarily concerned with a valid purpose, protecting themselves, their images and their identities [183].

Family members, co-habitants and friends: these people are most likely to be encountered (and therefore photographed) in their own home. Existing guidelines for visual research recognise privacy includes reasonable expectations of where you will not be unknowingly photographed and this should be considered the case for people in their own home [184, 185]. It may not be practical to collect written consent from all family members and cohabitants, but to protect their autonomy verbal permission should be sought by the participant permission prior to wearing an automated-camera in the home. The participant should also be provided with the necessary information to explain how images will be recorded, securely stored and not disseminated without consent. When prior verbal permission is not possible (e.g. for friends and acquaintances), verbal permission upon first contact with explanations and the offer for device removal if requested, will respect their autonomy. When permission is not given, research should not proceed.

Colleagues and co-workers: the appropriateness of an automated-wearable camera will differ between workplaces. Numbers of employees, work context, relationships with co-workers and interactions with the public will differ by, for example, offices, shops and hospitals. To protect the privacy of all parties each setting should be assessed by the researcher and potential participant before data collection starts; when the workplace is not suitable for image capture the participant should be excluded or directed not to wear at work depending on the research question. Parents should be informed of the research when the workplace involves interaction with children (e.g. teachers, other school-workers or day care workers). When the device is worn at workplaces, participants should seek verbal permission from managers or supervisors, and should inform direct co-workers about the device. If requested, the device should be removed.

Strangers and the general public: existing guidelines state that when taking images of individuals and groups in public spaces it is not practical or necessary to obtain informed consent unless the images are published or disseminated in such a way that they can be recognised [184, 185]. However, the participant should be advised to remove the device if requested to do so by a member of the public to protect both the participant and the third party.

There is a small risk of burden or harm to the participant (contradicting the principle of non-malificence) when wearing SenseCam type devices in free-living. For example, the participant could be questioned by (potentially hostile or suspicious) third parties who object to unsolicited image recording. Or they may attract unwanted attention from people who notice and possibly want to take the device. In field testing and the pilot studies participants reported that they felt more comfortable going out in public when provided with a short statement explaining the study purpose and device and concluding with the offer to remove. This reduces burden for the participant and the third party. Furthermore, instructing the participant to remove the device if they ever feel uncomfortable or threatened (e.g., walking home late at night) is recommended to reduce the risk of related harm. Participants should also be told to inform third parties that they can request deletion of images they are in by asking the participant to inform the research team or contacting them directly.

Research with SenseCam has certain parallels to covert research insomuch as some third parties will not know they have been photographed and this type of research has very complicated and serious ethical and legal issues [184]. The distinction is that automated-wearable camera research is not about the third parties or their behaviours, and their recording in images is largely incidental. Regardless, respect for autonomy directs that the privacy of third parties must also be protected, and images containing them should be treated appropriately. For example, if their presence is investigated as influencing the behaviour of the participant (for example eating or drinking in a group setting) their anonymity should be preserved and no image that identifies them should be published without consent.

In certain settings photography may be prohibited, for example some workplaces, banks, swimming pools, and airport security. Participants should be advised to remove the device before entering these settings to avoid challenge or incident. Pocket sized information cards can be provided to help participants explain the device if needed. There may be some populations that are not suitable to be included in this type of research as a result of these and other settings. Specific examples include medical or care providers interacting with patients, workers at military institutions or those at workplaces with high levels of commercial secrecy. Consideration should also be given to the appropriateness of research in certain cultural settings that do not approve of photography (e.g., aboriginal communities) [184].

6. Images will be seen by a large team of researchers.

An automated-wearable camera study with 50 participants collecting data for 3 days will generate nearly 500,000 images and team of researchers may be needed to code and process these data. While a participant may prefer that only the researcher(s) they have met and spoken to will see their images, at informed consent, participants should be made aware that a team of trained researchers may be viewing their images. Appropriate training and instruction should be given to all researchers that come into contact with the images as this is unlikely to be specifically covered in existing institutional ethical training; researchers coding image data should be reminded to not discuss the content with anyone outside of the team, to not identify anyone they recognise in the images, to store the images safely, and to be aware of how sensitive the data are.

Operating Ethical Framework for automated-wearable camera research

Based on the ethical issues 1 to 6 discussed above, I formulated a checklist for free-living observational studies using automated-wearable cameras. This was done in collaboration with researchers from University of California San Diego and Auckland University of Technology. This is shown in Table 6.1, taken directly from the paper published in 2013 [139].

Table 6.1 Ethical guidelines for the use of automated-wearable cameras in observational health behaviour research (reproduced from Kelly et al, 2013)

Informed written-consent of participant	Participant information should explicitly detail the following: • how many images and how much information will be collected • the nature and type of data that can be collected by wearing an automated-wearable camera (images will depict where you go, what you do, and for how long) with examples • participants can forget they are wearing the device and record unwanted and unflattering images with examples provided (e.g., bathroom visits, online banking) • data of illegal activities may not be protected by confidentiality and may be passed to law enforcement depending on the national law and nature of the activity • no individual will be identifiable in any research dissemination without their consent • participants will have the opportunity to view (and delete if necessary) their images in privacy • participants are able to remove the device or temporarily pause image capture whenever they wish • participants will not get copies of their images • a team of specifically trained researchers will have access to the image data
Privacy and confidentiality	Devices should be configured so that data can only be retrieved by the research team. It should be impossible for participants or third parties who find devices to access images Data should be stored according to national data protection regulations Identifying images should not be used without express consent of those individuals who are depicted Devices should be configured to allow participants to cease recording for short periods. Participants should also be allowed to remove the device at any time, with examples of where this might be appropriate (e.g., airport security) Appropriate training should be provided for all those in the research team who have contact

	with the image data
Non-malificence	Participants should be prepared for questions by the public with a short sentence that explains the device and concludes with an offer to remove if they are feeling uncomfortable Participants should be instructed to remove device in any situation where it is attracting unwanted attention, or they feel threatened or uneasy wearing the device
Autonomy of third parties	Participants should seek verbal permission from family members and co-habitants prior to study commencement Participants should seek verbal permission of workplace managers or supervisors. If possible this should be prior to study commencement, but in reality may be a rolling process. Appropriateness of device to work setting should be assessed by researcher Participants should inform friends and acquaintances of device when encountered. Offer to remove device if they are uncomfortable Participants should be told to inform third parties that they can also request image deletion by asking the participant to inform the research team, or contacting them directly The privacy and anonymity of third parties must be protected; no image that identifies them should be published without their consent Photography is inappropriate in some cultural settings and automated-wearable cameras should not be used in these instances

Strengths and limitations

The strength of this work is that the recommendations are derived from real life experiences of participants and researchers in field testing and pilot studies. This framework addresses issues raised by research participants during device-collection interviews, as well as theoretical issues. It is also important that the framework is a development and expansion of existing and well-accepted frameworks for visual research. The framework is a timely contribution as visual methods are becoming more prevalent as costs come down and technologies improve.

It is a limitation that the experiences used are from a small number of studies; this framework should be considered a first attempt to codify the issues and solutions. It is possible that there are issues with these techniques that have not arisen or occurred to the SenseCam community yet. It is likely the framework will need updating and reworking in the future as these issues emerge.

As an example, this framework does not consider studies in vulnerable populations. In these cases the extent to which participants can give informed consent is debatable and the consent of carers or significant others may be required [192]. A development in Oxford University is increased awareness of the sensitivity of religious data and there are recent changes to the rules on what can and cannot be stored. Regarding SenseCam, it is possible to collect images of visits to places of worship and the framework does not currently cover this.

Conclusion

This chapter attempts to address a key aspect of developing a new methodology in observational research; namely ensuring that the method adheres to ethical standards for research with human participants. Automated-wearable camera research generates more image data than ever before, some of it potentially unflattering or unwanted. This can be intrusive and participants and third parties may feel uncomfortable or that their privacy has been invaded. I have attempted to formalise the protection of all according to best ethical principles through the development of an ethical framework.

It has been stated elsewhere that technologies often evolve faster than legal and ethical systems are able to respond and unforeseen ethical issues emerge [182, 193]. In this chapter I describe how, with other researchers, I have tried to provide ethical guidance that keeps pace with a developing technology.

Part 3 – Testing my SenseCam method

CHAPTER 7 – DESIGNING MY MAIN EXPERIMENT

Summary

Having established the feasibility, validity, reliability and ethics of my SenseCam method in Chapters 3-6 (in response to my second thesis objective), I designed a study to achieve the third objective of this thesis; to use SenseCam to investigate error on self-reported travel behaviour on a modified version of the NTS.

In this chapter I state the four research questions I formulated to achieve this objective and then outline procedures appropriate for answering them. This refers to the way I will include, filter and exclude journeys and data, and how I will seek to enhance internal and external validity by reducing bias and data loss.

Thesis objective 3

"Test my method in a suitably powered sample, and investigate the error on self-reported journey durations (and compare to results from GPS review)"

Study design

The study design for the final experiment will be a measure of agreement investigation (similar to Pilot study 1), as described by Bland and Altman [161]. It will include elements of concurrent validation (crude SenseCam data compared to crude travel diary data) and criterion validation (criterion SenseCam data compared to corresponding travel diary data).

Research questions

RQ1 – how does travel diary and SenseCam data collected over a three day period compare (comparison of measures)?

The answer to this RQ will reveal if SenseCam is a feasible measure of travel behaviour in a sample of this size, and how comparable the data from the two measures are. I will use the crude eligible data from both SenseCam and the diary. Journeys will be deemed eligible if they meet the inclusion criteria for length. These are greater than 3 minutes and less than 120 minutes (recorded for SenseCam; reported for travel diary).

The rationale for RQ1 is to test SenseCam in the largest sample yet used. It is unclear if it is feasible for use in a study beyond 20-30 participants. The three day data collection has been chosen as many health behaviour investigations use periods similar to this, and it is unclear if SenseCam is suitable for such research in a sample this size. I have chosen not to have a longer collection period to control participant burden.

RQ2 - How well does the travel diary measure individual journey duration compared to SenseCam (criterion validity)?

Taking SenseCam as the criterion measure of journey duration, this will show the error on the travel diary when assessing individual journey duration. The answer to this RQ will tell the reader how well the duration of individual bouts of travel are reported.

I will identify matched pairs of journeys as discussed in Chapter 5 – Coding protocol. A matched pair will be eligible if there is complete diary and SenseCam data available for that journey, and the reported duration is eligible (between 3-120 minutes). For journey duration, bias between methods (group-level) will be assessed using the paired t statistic providing the mean difference between methods and its 95% confidence interval. Individual journey-level agreement between methods will be examined using Bland-Altman 95% limits of agreement [194]. The correlation coefficient will be calculated for the relationship between the journey durations estimated from the two methods using both within-subject [195] and between subject methods [195].

Using the Bland-Altman approach I will be able to distinguish systematic (bias) and random error (limits of agreement). I will also be able to use this figure to estimate error or bias at a daily summary level based on journey frequency. I will also assess how well mode is reported, compared to the criterion results from SenseCam.

The rationale for RQ2 is to test the findings from the pilot studies. Pilot 1 suggested systematic over-reporting of travel duration in adults [154]. Figure 7.1 presents a Bland-Altman plot for Pilot 1 data (between-method differences against the mean journey duration for the two methods). The plot illustrates the substantial fixed bias (over reporting of journey duration independent of journey length) revealed by the paired t-statistic. This was found to be +154 seconds (2:34 minutes) per journey. It also shows a large random error at the individual level (95% limits of agreement = 154 s ± 598 s (-444 to 752 s)). Of the 88 journeys, 62 journeys (70%) were over reported and appear above the y=0 line, while just 26 journeys (30%) were under reported.

In contrast, Pilot 2 showed no significant over-reporting bias in children. However, this pilot used a researcher led interview rather than a diary and only included school journeys. RQ2 is designed to investigate in a suitably powered sample whether or not there is an over-reporting bias for journey duration.

Figure 7.1 Limits of agreement (Bland-Altman) plot for self-reported journey duration and SenseCam journey duration (n=88) (reproduced from Kelly et al., 2011)

Image subject to copyright

Original available at

http://www.ijbnpa.org/content/8/1/44

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.

RQ3 - Is the bias on self-reported journey duration different according to certain variables?

The answer to this question will alert the reader to which types of self-reported journeys and participants are more or less likely to provide biased journey times. I will conduct the same analysis as in the previous research question but this time explore differences by variables hypothesised or previously shown to influence self-report error or bias. These variables will be grouped as study, journey and participant characteristics and will include: gender and age shown to influence reporting of journey duration when compared to GPS [61, 89, 196, 197]; day of week shown to be potential factor due to reporting fatigue [88]; integer preference, a suspected factor [95]; and mode suggested to influence over-reporting in Pilot 1 [189]. By examining variables such as educational status of participants, study location and weekend or weekday I will be able to assess possible threats to internal and external validity.

The rationale for RQ3 is to see if there are any variables that have a greater or lesser influence on the error on self-reported durations. For example, both Pilot 1 and 2 suggested mode and gender may have an effect, but the studies were not sufficiently powered to draw any conclusions.

RQ4 - How well does the travel diary measure daily summary travel behaviour compared to SenseCam (criterion validity)?

The answer to this question will tell the reader whether daily summary travel durations and journey frequencies are biased. I will return to the crude SenseCam data and select only the full (criterion) days of data. A day of SenseCam data was not considered full if one or more journey was missing or incomplete. This usually occurred if the camera was obscured for the start or end of a journey, not worn, or switched off for a section of the day. I will use this data to calculate criterion daily summary travel time and compared these summaries to the diary data.

The rationale for RQ4 is that summary travel durations are often used in public health and epidemiology. It is also driven by the finding from the GPS Review (Chapter 2) that the choice of matched-trip or unmatched-trip analysis might influence the findings. Doing this this will allow me to incorporate a second dimension of travel behaviour into my investigation (frequency as well as duration).

Defining an eligible journey

The coding protocol defined in the previous chapter allows me to code any SenseCam journey with three or more images. To be eligible for inclusion in this study I determined a minimum duration of three minutes and a maximum duration of 120 minutes.

Minimum duration

The National Travel Survey only collects detailed walking data on Day 7 of data collection (on Day 1-6 only walks greater than 1 mile or 20 minutes are included) [120]. As walking is of particular interest to my thesis, my experiment is based around investigating the Day 7 of the NTS.

On Day 7 the NTS asks participants to record all walks over 50 yards (45 meters). At a walking speed of 3.5 mph this would be a 50 second walk. However, the participants are only asked to complete one day of this as it is considered onerous. Therefore, to reduce participant burden for the multiple days of my data collection I will ask participants to record journeys over 3 minutes.

This is a limitation to my study design as it restricts journeys of 1-3 minutes from being reported. As SenseCam has been shown to be sensitive to journeys of 1-3 minutes I will be able to assess if this is an important bias. It should also be noted I asked by minimum duration, whereas NTS asks by minimum distance. I chose duration for standardisation between country locations which may use yards or meters by preference.

Maximum duration

The NTS defines a long journey as one of over 50 miles (within the UK) [120]. I am going to assume this only applies to vehicle journeys (not walking or cycling). Using rough speed approximations the extremes of this would be 40-45 minutes on a motorway (at 70 mph), or four and a half hours in city traffic (at 11 mph[22]). I have decided to use 120 minutes (2 hours) as an arbitrary cut off between these extremes to include journeys that could be considered usual or regular. This is a limitation to my method, and I will review if the data shows this to exclude a large proportion of journeys and potentially bias the data.

Sample size calculation

The sample size calculation is designed for RQ2 (How well does the travel diary measure individual journey duration compared to SenseCam (criterion validity)?). This is the primary research question. I also have data for similar research questions in Pilot 1 to base my estimates on. It is an assumption that these sample sizes will also be adequate for RQ3 and RQ4. However, as RQ3 will investigate different variables, samples within those variables will be smaller and this may affect the findings. Any sample over 40 will make this the largest health behaviour SenseCam study to date, satisfying RQ1. I have no pilot data to make estimates for RQ4.

[22] http://www.thisislocallondon.co.uk/news/topstories/804876.london_cars_move_no_faster_than_chickens/

In RQ2 I am primarily interested in bias at the group level; therefore I estimated the sample size required from the desired 95% confidence intervals around the mean difference between methods. In other words, the uncertainty on my estimate of bias on self-report. This approach is discussed by Hopkins et al., [172].

As a lower limit (widest acceptable), this 95% confidence interval should be narrow enough to give the estimate of bias epidemiological relevance. As an upper limit (narrowest still useful) it would be meaningless to have a 95% confidence interval less than the uncertainty on the SenseCam method itself. At the time of designing this study, I had not conducted the criterion testing of SenseCam[23]. Therefore, I did not have any experimental indication of the uncertainty inherent in the device and method beyond the sampling rate of the camera.

Image sampling rate

I was able to investigate the sampling rate of the device as an indicator of the uncertainty in the SenseCam method. Analysis of data from the field testing and Pilot 1 has shown the sampling rate of at least every 20 seconds published by Microsoft [83] to be an under-estimation for travel behaviour. Using software developed by Dr Aiden Doherty [180] I was able to show the average sampling rate over 70,000 pilot data images was actually 9.3 seconds and varies by journey mode (see Table 7.1). This is likely to be a reflection of the different way the in-device sensors (light, heat and motion) are triggered by the different actions of walking, cycling and vehicle travel.

Table 7.1 Image sampling rate for different modes of travel for n=70,000 travel images

Journey mode	**Sampling rate (seconds per image)**
Vehicle travel	7.6
Cycling	11.6
Walking	16.8
All Journeys	**9.3**

[23] It had not even occurred to me to conduct such a test at the time of designing my main study

The sampling rate of every 10 seconds (rounded from 9.3 seconds) means that without further information, an uncertainty less than 20 seconds (10 seconds at start and 10 seconds at end) would be meaningless.

Based on the formula for the standard error of the difference between means (d);

$$\mathbf{d} = \sqrt{(\boldsymbol{s}^2/\boldsymbol{n})}$$ [161];

where n represents the number of measurements; and the formula for the 95% confidence interval (95% CI);

95% CI = +/- 1.96 * d [161];

and the standard deviation (s) of the individual differences (bias) found in Pilot 1 of 305 seconds, I calculated the required measurements (matched journey pairs) for a range of potential 95% confidence intervals (see Table 7.2). I also used the number of eligible journey pairs returned per day in Pilot 1 (4.4 per day) to estimate the days of data collection required, and the three day proposed protocol to estimate the number of participants required. I made similar estimates for active journeys (walking and cycling; 2.7 per day in Pilot 1) to reflect the public health focus of this thesis.

Table 7.2 Estimates for required measurements, days of data collection and participants for a range of 95% confidence intervals on the estimate of bias between measures

		All travel		**Active travel**	
Predicted 95% confidence interval on bias (seconds)	**Measurements required (n)**	**Days of data collection (n)**	**Participants required (n)**	**Days of data collection (n)**	**Participants required (n)**
20	3573	812	270	1324	441
30	1588	361	120	588	196
40	893	203	68	331	111
45	706	160	53	261	87
50	571	130	43	218	71
60	397	90	30	147	50

Using these results I chose the sample size based on a predicted 95% confidence interval on the bias of 45 seconds. Therefore I aimed for 85-90 participants. This was a balance of achieving an epidemiologically relevant uncertainty (on all travel and active travel estimates) and a pragmatic number for a thesis data collection. It is also only an estimate as the standard deviation in my main experiment may be different to Pilot 1.

Experimental validity

The following section describes how I attempted to maximise experimental validity and minimise potential sources of bias.

Internal validity

Internal validity is the extent to which conclusions drawn from the experimental data are free from confounding issues which cause bias such as reactivity and missing data. They are sometimes described as the methodological quality of the study.

Reactivity

Reactivity is a change in normal behaviour as a result of being measured [198]. It is a potential limitation of my study that being measured, in this case wearing a SenseCam, makes the participant behave differently to their normal behaviour. The normal behaviour in this instance is diary reporting of travel behaviour with no device. It is also possible that participants may 'react' to wearing SenseCam, possibly changing their travel behaviour and making them more aware or conscious of journeys. This in turn may change the way they report journey durations and influence my findings. We do not know to what extent wearing SenseCam influences self-reported travel behaviour. This is an important question when trying to investigate the error on self-report with SenseCam.

To truly measure the extent of reactivity, participants must be both knowingly and covertly measured, with the results compared. However, this is both practically and ethically unfeasible, so studies often compare first and subsequent days of measurement to investigate this phenomenon as it is assumed participants get familiarised to devices and reactivity decreases during the study period [198]. Unfortunately, a 2010 review of pedometer trials found the reactivity question is often overlooked when reporting results [198]. They identified just three studies [199-201] that assessed reactivity in children; in this case defined as the difference between first and last days measurement. Each of these studies suggested no reactivity to wearing either sealed or unsealed pedometers.

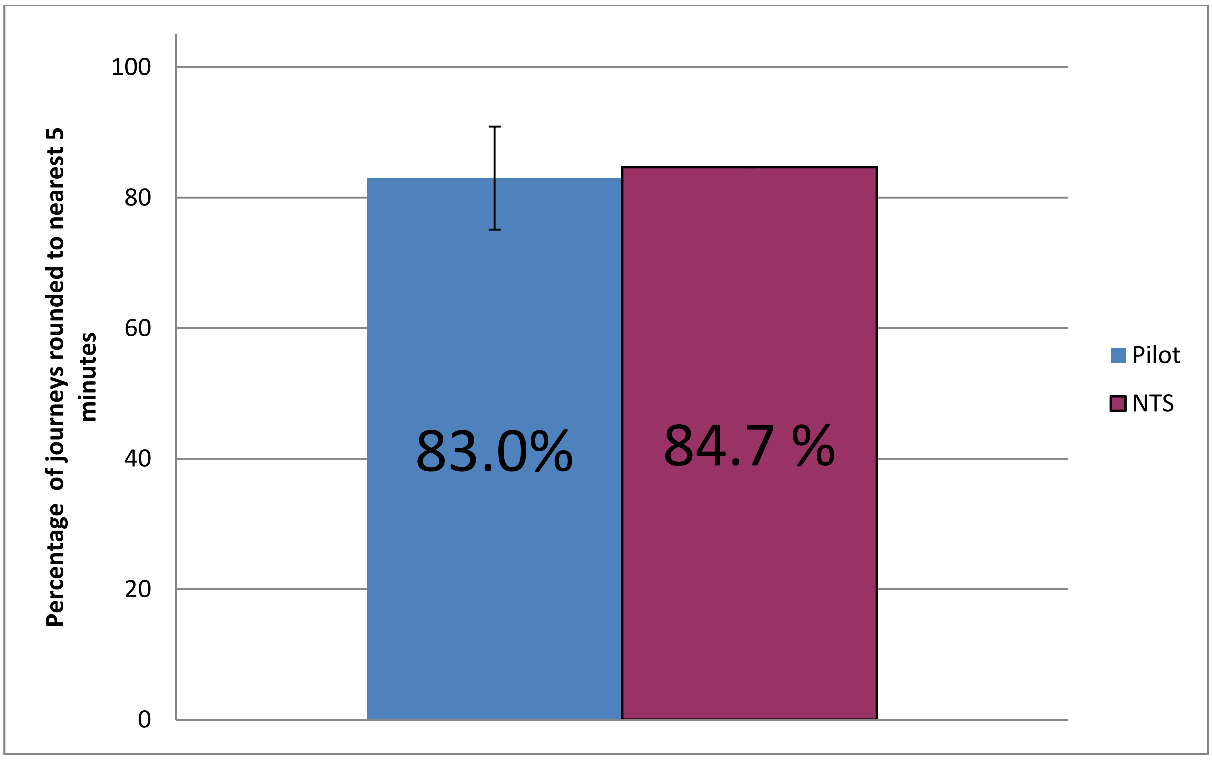
In Pilot 1 and 2, I only collected one day of travel data so was unable to assess potential differences across days. However, I have compared one reporting characteristic of the NTS to the data from Pilot

1. As previously discussed, a suspected source of error in self-reported data is rounding journey duration to the nearest five minutes – termed integer or digit preference. From the publically available 2008 NTS data I analysed 2,052,864 journeys and found that 84.7% (95% CI 84.65-84.75) of journeys were rounded to the nearest five minutes [202]. In comparison, in the pilot 80.3% (95% CI 75.9-90.8) of the 88 journeys were rounded to the nearest five minutes (see Figure 7.2).

This suggests that there was a small non-significant lower difference in the proportion of rounded journeys in the pilot, possibly suggesting a marginal reactivity to wearing the device. However, whether this apparent reduction in rounding was caused by the device, the one day of data collection (with device) compared to six days (NTS) or the non-representative nature of the pilot sample (minimum educational level university degree) is unclear. I searched for evidence of reactivity in the same way, and across days in the main experiment. In an attempt to minimise bias I did not explicitly tell participants I was assessing how well they report behaviour.

It is worth noting that participants completing the diary may have already been subject to the reactivity from the diary itself. Therefore, introducing the camera might not have any substantial extra effect on awareness of journey behaviour.

Figure 7.2 Percentage of journeys reported to nearest five minutes



The 95% confidence interval on NTS data was too small to be displayed

Missing data

Missing data can bias the results and it is important to assess this [174]. Altman and Bland suggested that data can be "missing completely at random", "missing at random" and "missing not at random". In the first case the fact the data are missing is unrelated to any unobserved value and the outcome. In the second case the data are missing in a predictable way. Both these cases are less of a threat to bias. However, in the third case the data may depend on an unobserved value, and can lead to bias [203].

In terms of recording travel behaviour, this depends on why the data are missing; are they missing for a reason (e.g. SenseCam cannot record train journeys well) or are they missing completely at random (SenseCam fails randomly for 1 in 100 journeys)?

In terms of investigating bias on self-report (RQ2 and RQ3) this depends on why matched-pairs were unavailable; did SenseCam systematically miss the journeys most likely to give high or low errors? Further, what was missing from the diary? These missing journeys will also influence the summary analysis (RQ4). To minimise these biases, it was important to attempt to minimise the data loss. While developing my methods I was able to identify factors that lead to SenseCam missing data. These include forgetting to wear (or forgetting to replace when removed), lens obscured by clothing (especially during inclement weather), and only remembering to turn on device once the first journey of the day has started. All of these issues were added to the participant information documents used in the pilot studies and were covered during device demonstration and hand-out. It is well known how important high quality instructions are to participant compliance [53]. In Pilot 2 participants reported that the text message sent to remind them to charge the device was helpful [157]. This agrees with previous findings [204] was added to the data-collection protocol – I asked participants if they would like an optional email or text reminder each evening to charge the device. Regardless of this, there were missing data and I reported these in my results. Data imputation of missing values is not appropriate for a method comparison study such as this.

External validity

External validity is the extent to which conclusions drawn from the data are generalisable to the larger population (e.g. do not exhibit sample bias). Age, sex, ethnic origin, socio-economic status may all limit generalisability [163]. In this case I am considering whether any reporting bias I detect is generalisable.

I conducted data collection in different locations to give a range of travel environments and travel behaviours. However, the requirement to wear a camera for multiple days may have resulted in a selection or recruitment bias. I assess these issues in my main experiment.

Chapter conclusion

My main experiment was designed as a comparison of measures study with concurrent and criterion validity components. I state my four research questions and the rationale for each. A sample of 85-90 participants, recruited from four locations, and asked to wear SenseCam and complete a travel diary for three days, was estimated to give meaningful confidence on the results. I identified various sources of bias and threats to validity which will be assessed. Specific methods will be presented in the next chapter.

CHAPTER 8 – MAIN EXPERIMENT

Summary

This chapter describes the methods, results and discussion from the main experiment of this thesis. Using the previously developed methods, protocols and procedures I compared data collected from a self-reported travel diary (based on the NTS) to that recorded by SenseCam. The study was designed to answer four research questions (RQ1-4). These research questions focus on the feasibility of SenseCam and the bias or error on self-reported travel data and relate to the third and final objective of my thesis.

I recruited a semi-convenient sample of 88 participants and asked them to wear SenseCam and complete a travel diary for 3 or 4 days (depending on study wave). From 84 participants who complied with the study, I collected 1864 journeys (1369 in the diary; 1629 recorded by SenseCam) in four locations (Oxford, UK; Romford, UK; San Diego; USA; and Auckland, New Zealand).

After filtering for eligibility (by duration), comparison of the SenseCam and diary data suggested the diary collected slightly more eligible journeys (0.14 per day) that were marginally greater in duration (approximately 2 minutes). As a result total travel time per day was approximately 12 minutes longer according to the diary data. However, generally the data were similar from both measures.

Taking SenseCam as a criterion measure of journey mode and duration, and comparing to the self-report data, I demonstrated journey duration was over-reported by 2:08 minutes per journey. I analysed this finding with respect to descriptive variables that may influence this bias.

Analysis of summary daily summary travel duration contradicts my estimates of daily over-reporting based on individual journey bias. It is unclear if this finding is a result of including frequency in analysis or that applying filters to the data introduces biases that confound any real effects. The implications of all these findings are discussed with relation to population and individual level investigation.

Methods

Study locations

Volunteer participants were recruited in four geographic locations to give a range of developed world travel environments and contexts. The locations were chosen for pragmatic reasons where I had contacts in partner institutions willing to host and support data collection.

The locations were as follows:

- Oxford, Oxfordshire, UK (home institution)
- Romford, Essex, UK
- Auckland, New Zealand
- San Diego, California, USA

Recruitment

All participants were volunteers, although recruitment approach varied slightly by location for practical reasons. In Oxford, Romford and Auckland recruitment was active in nature and participants were approached from existing networks. Potential participants were approached by the research team by email or in person. We asked selected individuals e.g. college administrators, sports club captains, social club leaders to circulate emails asking for volunteers. 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[24]. In San Diego a $50 compensation for the participants' time was also offered, in-line with institutional policy.

[24] We advertised for cycle commuters as we wished to ensure a sufficient number of active journeys in the data in order to give the results more public health relevance

Study protocol

Participant information

Having volunteered to take part, participants were provided with participant information sheets, and given the opportunity to ask questions before providing informed consent. The relevant documents can be found in Appendix 7.

Study visit

Post consent, participants attended a study visit to be taken through the study protocol. In Auckland and San Diego sessions were conducted in groups on the University site. In Oxford and Romford the researcher(s) visited the participants at their home or work according to preference. Appropriate risk assessments were conducted prior to these study visits. Study visits took between 15-20 minutes per group or participant.

SenseCam protocol

During the study visit participants were given a demonstration of SenseCam operation and asked to wear the device for between three and four days, depending on location and data collection wave. This difference existed when the data collection period was over a weekend because there was no weekend study visits. Start days were varied (non-randomly by collection wave) so a range of weekdays and weekend days were covered. In Oxford, Romford and Auckland, collection started on the following morning. In San Diego, due to the time available, data collection started as the participant left the study visit, and concluded when they returned resulting in some "partial days". Brief user guides (see Appendix 8) were provided with instructions on the practical issues with wearing SenseCam (e.g. wearing over bulky clothing, weather and securing the camera). They also received an optional text message or email to remind them to charge the device each evening.

In the pilot studies I attempted to control device reactivity known to be an issue in physical activity assessment [198]. Similarly, in this study, participants were asked to wear the device throughout the whole day (when possible) to become acclimated to the novelty of wearing a camera. However, participants were told that they could remove the device at any time (e.g. if feeling uncomfortable or if asked to do so by a colleague or family member). They were also instructed in how to stop image recording for a seven minute period by using a privacy button. Additionally, the device was configured to scramble images, which could only be unscrambled on the laptops of the researchers administering this study. Furthermore, participants were not explicitly informed the images would be used to assess "accuracy" of travel reporting; they were told the images gave us a visual record of their travel that would give us further information to add to the diary data. All of this was in accordance with the ethical framework developed for this thesis (see Chapter 6) [139].

Travel diary protocol

Participants were given a three or four day travel diary (depending on collection period) which was a modified, un-validated version of the National Travel Survey (NTS)[25], the continuous annual UK travel survey. [60] As in pilot work, the diary asked participants to record journey mode, purpose and duration for the same period as they wore SenseCam. Questions about ticket price were removed (see Appendix 9). In accordance with the actual 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 transportation between any two locations of greater than three minutes (as discussed in the previous chapters). In each location, an example of a completed day was provided to demonstrate how the survey should be completed. This was modified to be relevant to that study location. As with the NTS, 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 (Appendix 9).

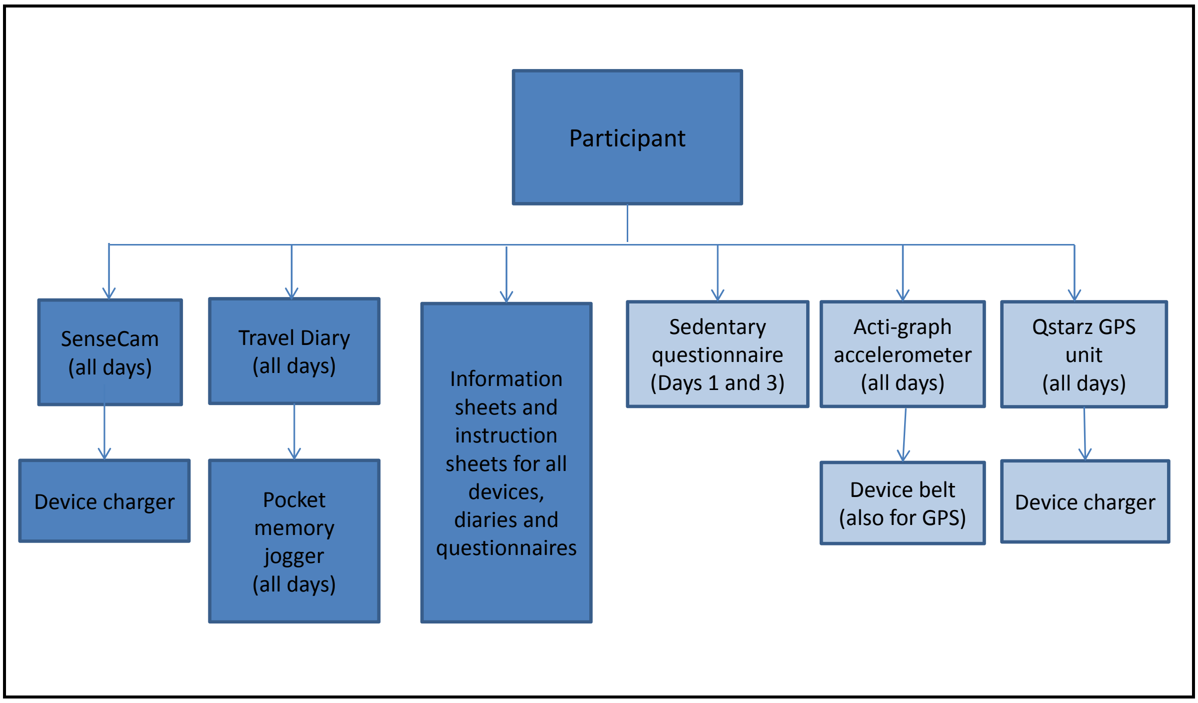
Additional devices and measures

Participants also completed a short demographic survey (see Appendix 10). This study presented an opportunity to collect GPS and accelerometer data, as well as sedentary time questionnaire data that would be used in collaborator studies. Participants were therefore asked to wear a device belt with a QStarz GPS unit and Acti-graph accelerometer, and complete a short (two page) sedentary (sitting

[25] As discussed in Chapter 1, the NTS has never been objectively validated in this way. Similarly, my modified version has not been validated. All questions in my modified version appear in the real NTS. It is possible that by removing questions I made the diary less burdensome and introduced some bias on the self-report data I collected.

time) questionnaire about Day 1 and 3. These data are being independently analysed by a collaboration of researchers from University of California San Diego, Auckland University of Technology and Oxford University, with a view to investigating what SenseCam images can reveal about sedentary behaviour, accelerometer based physical activity recognition and travel environments. Details of resulting papers are included in Chapter 9. All devices and materials administered are shown in Figure 8.1.

Figure 8.1 Devices and materials given to participant at research visit



Dark blue represents resources used for this thesis; light blue used for related projects.

Return of devices and data

Return study visits were always conducted by a researcher, in person, and one-on-one, as directed by the ethical framework; 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 [180]. Other devices and all questionnaires were also collected at this time. A short satisfaction interview was conducted, exploring the participant's experience of wearing the device, and any issues they had during the collection period (see Appendix 10). This survey was based on the one I used for my pilot studies [157, 189]. Return study visits took between 25-30 minutes per participant.

All image data were anonymised and stored on password protected machines in accordance with procedures of the institutions where data collection took place. Questionnaire data was scanned to make electronic copies which were treated in the same way. Data were shared between institutions by secure file transfer.

Ethics approval

The primary approval for this thesis was granted by the Social Sciences and Humanities Inter-divisional Research Ethics Committee (IDREC) in accordance with the procedures laid down by the University of Oxford for ethical approval of all research involving human participants (IDREC Reference Number: SSD/CUREC1A/10-092 – see Appendix 4). This study was also approved by the IRBs at the other participating institutions (UCSD and AUT; see Appendix 11).

Data coding

Data coding was conducted according to the validated protocol described in the previous chapters. Once coded, the software automatically calculated journey duration. These data were entered into an electronic database, at the same time as the data from the travel diary. Journeys from the images and the diary were manually matched at this stage using; time of day; sequence; mode; and purpose; to identify "matched pairs". The following data and variables were recorded: participant number, study location, date, SenseCam mode, SenseCam duration, diary mode, diary duration, trip chaining, journey breaking and reason missed (if applicable). Purpose was coded using SenseCam and travel diary data, based on origin and destination. When there was disagreement, SenseCam purpose was used.

Where journeys could not be matched (because the images did not show it, or the diary did not record it) this was recorded as an unmatched journey, with the reason for the miss noted. For example, "camera off", "lens obscured", or "not recorded in diary". All the same variables as with "matched-journeys" were recorded. In some instances, journeys were noted in the diary but analysis of the images showed that no such journey occurred on that day. These journeys were recorded as "Phantom journeys".

I then made a second "coding pass" to check segmentation fidelity, missed journeys and data entry errors.

Data analysis

All statistical analyses were conducted using *PASW* Statistics 18.0 (SPSS Inc., Chicago, IL, USA) software package. Graphs were created using Microsoft Excel (2010).

Compliance

The compliance of the study participants was assessed by the amount of data returned; specifically how many days the SenseCam returned images for, and the travel survey was completed on, compared to the total numbers expected. This was assessed at a whole group level, and also by study location.

Examining the data set

I conducted the following analyses of the data set;

- Assessed evidence for reactivity caused by wearing the device (comparison of self-report data to "real" NTS data) – this was to assess a potential source of bias
- Assessed whether the differences between device and diary were normally distributed (visual inspection and Shapiro-Wilk test) – the parametric tests planned required normally distributed data, so this was to assess if the dataset required transforming
- Tested for heteroscedasticity or proportional bias (visual inspection and Kendal's Tau) - the parametric tests planned require homoscedastic data, so this was to test if the dataset required transforming

The results of these analyses are presented in the appropriate parts of the Results section.

Data quality control

I conducted checks on coding and data entry on a randomly selected 10% of the data set.

Data treatment

This refers to the way I included, filtered and excluded journeys and data. I had a number of research questions for this study, and the data were treated according to the specific research question, as outlined in the previous chapter (Chapter 7 - Study design).

Results

Data collection

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). The primary reason for this was the availability of devices although it gave a range of seasons too. We were able to secure 20 loaned units from Microsoft Research for the Auckland and San Diego phases and this allowed larger groups to wear devices simultaneously; in Oxford and Romford we used the three working devices owned by our research group. In each location, the research team comprised 2-4 researchers following the protocols developed through my pilot studies. I led the data collection in Auckland and San Diego and trained and supervised a researcher (Anja Mizdrak; Research assistant in our group) who conducted the data collection in Oxford and Romford.

Data coding took place between October 2012 and January 2013. The initial "coding pass" took approximately eight weeks, or 300 hours (corresponding to 1-1.5 hours per participant day of images). The second "coding pass" took a further four weeks, or 150 hours. Data analysis took place in January and February 2013.

Description of locations

The four study locations provided a range of high income country (HIC) travel environments. Oxford is a small UK city with high levels of cycling and a pedestrianised centre. Recent data from Sport England showed 14% of adults cycled five times a week (second in the UK to the 25% in Cambridge) [205]. It also has good public transport services. Romford is a suburb on the commuter belt of London. It is well connected to the London over- and under-ground rail network and has multiple bus services [206]. Auckland is the largest city in New Zealand. It is coastal and has a high levels of car use and low levels of cycling [207]. Public transport is less common than in the UK sites. San Diego is an American coastal city with high car ownership. It has a well-developed network of cycle paths, bus services and an over-ground train link [208, 209].

Participants

Table 8.1 shows the participant demographics for those who followed the data collection protocol (n = 84). San Diego provided the highest proportion of participants. Due to much recruitment taking place in university settings, the educational status of the sample was relatively high. The ages ranged from 19-60 years, and there were almost equal proportions of males and females.

Table 8.1 Participant demographics for n=84 who followed the study protocol

		Overall	Auckland, New Zealand	San Diego, USA	Oxford, UK	Romford, UK
Characteristics						
Participants (n)		**84**	15	37	25	7
By gender; Male; Female		 **44** **40**	 3 12	 26 11	 11 14	 4 3
Age (mean in years); Range; Standard Deviation		**33.3** **19-60** **12.3**	38.3 21-57 10.3	37.7 21-60 12.6	26.0 19-51 8.5	25.3 19-49 10.8
Educational status[26]	1 2 3 4	**13** **19** **16** **36**	0 0 0 15	3 2 15 17	5 16 0 4	5 1 1 0

[26] Educational status was derived from highest qualification: 1 = completed high school; 2 = current undergraduate student; 3 = completed undergraduate degree; 4 = completed graduate degree. This was the question approved by the IRB in San Diego. In Auckland and both UK sites we asked questions about job type and inferred highest qualification from these answers based on accepted required qualifications and age of participant. For example, 50 year old senior research scientist would be scored as 4 (completed graduate degree); 19 year old office administrative assistant would be scored as 1 (completed high school). Unemployed (n=3) was also scored as 1.

Reasons for declining to take part in study

Between the Oxford and Romford data collection, 11 volunteers declined to take part in the study after receiving the participant information. This represented 11% of the overall sample engaged. Their reasons are listed in Table 8.2. No one from Auckland or San Diego declined which is probably a function of the known research network approached (Auckland) and the passive advertisement which the participant could choose to answer (San Diego). A further four participants in Oxford and Romford gave informed consent, but failed to start the study protocol.

Table 8.2 Reasons given by n=11 potential recruits for declining to give informed consent (actual quotes)

Volunteer	**Reason given**
Analytical Chemist	Not allowed to have a camera at work
Laboratory scientist	Would not be allowed to wear at work [laboratory]
Senior Sales Adviser	Feel device would make customers feel uncomfortable
Civil Servant	Cameras not allowed in the office
Bank worker	Not allowed to have a camera at work
Primary School Teacher	Not sure head teacher would allow camera at work
Student	[I am] a very private person and do not like the idea of someone knowing what I do
Student	Not enough reward for participation to be worth the hassle
Not known	Invasion of privacy
Not known	Can't be bothered to fulfil study requirements
Not known	Timing of study not convenient

Compliance

Table 8.3 shows the compliance of the study participants as assessed by the number of days for which data was returned. Overall both measures performed well and were highly matched in terms of number of days data returned (SenseCam = 89.8%; Travel diary = 90.6%). Matthews et al., state that 79% accelerometer compliance in NHANES (National Health and Nutrition Examination Survey) in the USA was considered a success (though this would represent a far larger total sample) [53]. There was lower compliance in both UK locations, particularly Romford, compared to Auckland and New Zealand. The four participants who did not start the study should be considered dropout after recruitment. They represent 4.5% of the total number recruited. Hopkins et al., suggest that dropout greater than 10% would introduce substantial bias [172].

Days of SenseCam data (n = 31) were missed due to device malfunction, failing to charge the device or forgetting to wear the device. Days of survey data (n = 27) were missed due to the participant failing to record any data in returned surveys (i.e. returned blank surveys when SenseCam data for same period showed journeys occurring). One participant reported they had lost their diary.

Table 8.3 Compliance of n=84 participants (by location and measure) in terms of days of wear time and data returned

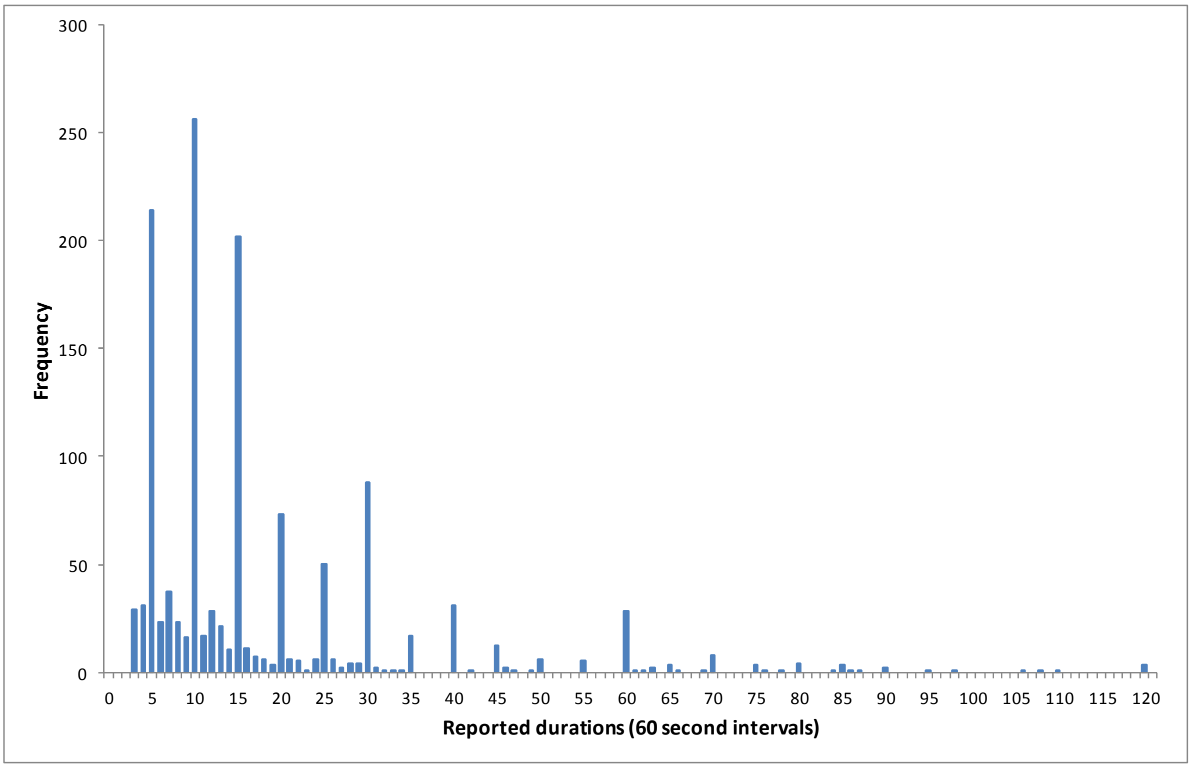
	Overall	Auckland, New Zealand	San Diego, USA	Oxford, UK	Romford, UK
Participants (n)	**84**	15	37	25	7
Days of data expected	**305**	46	156	80	23
Days of partial or full SenseCam data returned (% of possible total)	**274 (89.8%)**	42 (91.3%)	150 (96.2%)	66 (82.5%)	16 (70.0%)
Days of partial or full travel diary data returned (% of possible total)	**278 (90.6%)**	38 (82.6%)	154 (98.7%)	72 (90.0%)	14 (56.0%)

Reactivity

As previously discussed, it is important to consider if wearing SenseCam changes "usual" reporting behaviour. One test I have devised is assessing the proportion of journeys that are "rounded"; a suspected contributor to error on self-reported travel durations (which I will test in RQ3). I am assessing whether participants who wear SenseCam are less likely to round than those that do not wear the device.

Figure 8.2 shows a histogram of reporting at each minute interval. It has clear peaks at each five minute interval and this suggests even though participants were wearing SenseCam they still had a tendency to “round” their self-reported durations.

Figure 8.2 Histogram showing frequency of reporting at each minute interval in the travel diary for n=1369 journeys from n=287 days of data collection



When considering the proportion of journeys rounded to the nearest 5 minutes I found that of 1328 eligible journeys (3-120 minutes), 1011 had been rounded; this equates to 76.1%. (95% CI = 73.8-78.4%).

Data quality control

In the Criterion experiment reported in Chapter 4, Dr Aiden Doherty conducted some data coding to assess inter-rater reliability. Prior to this I conducted some training and coaching with Aiden in data coding protocol, and used 114 randomly selected journeys from the main study data set. We used the matched-journey data set presented later in this chapter so this represented 10.1% of the data. I used this as an opportunity to check my own data entry. I found one mistake on entry of duration and a different mistake on entry of mode representing an error rate of 2/228 or less than 1%.

Research question 1 – how does travel diary and SenseCam data collected for the same period compare (comparison of measures)?

The data were treated as proposed (see Figure 8.3 Study flow chart). This flow chart was adapted from guidelines proposed by Matthews et al., for reporting of methods and data handling [53]. Table 8.4 shows the number of journeys collected by each measure and how many were eligible according to the 3-120 minutes cut offs. 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 minutes and should not have been reported, or in some cases never happened at all (phantom journeys).

Table 8.4 Filtering procedure for eligible SenseCam and the travel diary journeys

	SenseCam	**Travel diary**
Days of data returned	274	278
Total number of journeys recorded	1629	1369
Excluded as over 120 minutes	5	6
Filtered out as less than 3 minutes	354	35
Total eligible journeys	1270	1328

Table 8.5 shows journey frequency and average duration by mode for all participants. Again, there is broad agreement between the two measures, though average travel diary durations are consistently slightly higher. Both measures suggest that two-thirds of trips are active (walking or cycling), and that on average walking trips are the shortest by mode of travel.

Figure 8.3 Study flow chart showing data treatment for Research questions 1-3

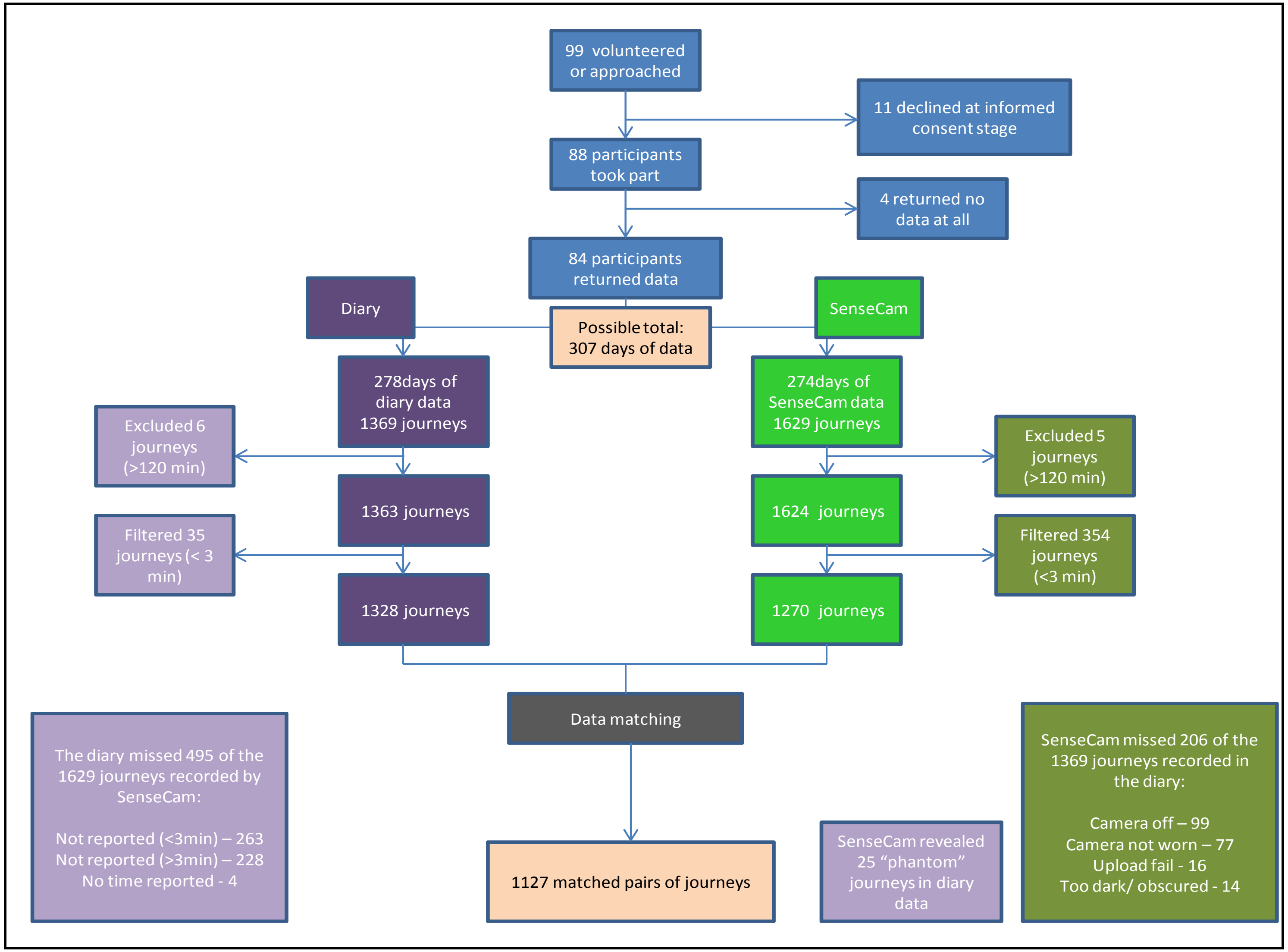


Table 8.5 Journey frequency, percentage of total and mean duration by mode and sedentary or active modes for eligible journeys (crude) by SenseCam (n=1270) and travel diary (n=1328)

	SenseCam				**Diary**			
	Journey frequency	**%**	**Mean duration (minutes)**	**Standard deviation (minutes)**	**Journey frequency**	**%**	**Mean duration (minutes)**	**Standard deviation (minutes)**
All	1270	-	15:39	15:03	1328	-	17:46	16:44
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	15:51	12:32	537	40.0	17:59	14:48
Active	783	61.7	15:31	16:26	791	60.0	17:36	17:57

Table 8.6 shows average daily journey frequency for all participants. Both measures suggest that on average each participant conducted just over 4.5 journeys per day. However, due to the greater average durations in the travel diaries this leads to higher estimates of total daily travel time when looking at the travel diary data. The travel diary data suggests 84:55 minutes of travel per day compared to 72:37 from the SenseCam data. This is a difference of 12:18 minutes (17.0%) per day for all travel, and 6:29 minutes (23.0%) and 5:44 minutes (12.9%) for sedentary and active travel respectively.

Table 8.6 Journey frequency and mean travel times per day by SenseCam and the travel diary

	SenseCam	**Travel Diary**
Days of data	274	278
Total valid journeys (n)	1270	1328
Mean journeys per day (n) Sedentary Active	4.64 1.78 2.86	4.78 1.93 2.85
Mean travel time per day (minutes) Sedentary Active	72:37 28:13 44:23	84:55 34:42 50:07

Missed journeys

There were instances where SenseCam recorded journeys that were missed by the travel diary. These are shown in Table 8.7. Across 274 SenseCam days, the travel diary missed 495 journeys (232 eligible journeys from 3-120 minutes and 263 short journeys <3 minutes that participants were not asked to record). This suggests that the diary should have recorded 1560 eligible journeys (3-120 minutes) during the study period; the data show it recorded 1328 (85.1%) and missed 232 (14.9%).

Table 8.7 Journeys recorded by SenseCam and missed by the travel diary (n=495)

	All missed journeys n (%)	Missed eligible journeys (3-120 minutes)	Missed ineligible short journeys (<3 minutes)
Missed journeys	495	232	263
By mode			
Car	52 (10.5)	41 (17.7)	11 (4.2)
Car passenger	17 (3.4)	16 (6.9)	1 (0.4)
Bus	7 (1.4)	7 (3.0)	-
Train	-	-	-
Boat	-	-	-
Walk	391 (79.0)	145 (62.5)	246 (93.5)
Cycle	28 (5.7)	23 (9.9)	5 (1.9)
Sedentary	76 (15.4)	168 (72.4)	251 (95.4)
Active	419 (84.6)	64 (27.6)	12 (4.6)
By purpose			
Commute	92 (18.6)	39 (16.8)	53 (20.2)
Work related	160 (32.3)	65 (28.0)	95 (36.3)
Shopping	97 (19.6)	31 (13.4)	66 (25.1)
Leisure	88 (17.8)	62 (26.7)	26 (9.9)
Visit friends	9 (1.8)	4 (1.7)	5 (3.0)
Other	48 (9.7)	31 (13.3)	18 (6.8)

The SenseCam also missed journeys recorded by the travel diary. Across 278 travel diary days SenseCam missed 206 eligible journeys. These are shown in Table 8.8 with the reason they were missed. 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%).

Table 8.8 Journeys recorded in the travel diary and missed by SenseCam (n=206)

Total missed	**206**
By reason:	
Camera off	99
Camera not worn	77
Upload fail	16
Too dark/ obscured	14

Principal finding 1

SenseCam is a feasible tool for investigating travel behaviour in a sample of 80-90 participants. Data loss from both measures was approximately 10% in terms of days when no data was returned, and 15% in terms of missed journeys (as revealed by the other measure).

SenseCam and the travel diary give comparable results in terms of travel behaviour. Journey frequencies are very close, but the travel diary gives slightly longer journey durations, and as a result greater daily estimates of travel time (overall and by sedentary or active).

Research question 2- how well does the travel survey measure individual journey duration compared to SenseCam (criterion validity)?

The data were matched as proposed (see Figure 8.3 Study Flow Chart). Table 8.9 shows the matched pairs of journeys by mode.

Table 8.9 Characteristics of matched pairs of journeys (n=1127) from n=227 days of data collection

Days of data	227	
Total number of eligible "matched pairs" of journeys	1127	
Car	312	27.7%
Car passenger	84	7.5%
Bus	34	3.0%
Train	13	1.2%
Boat	1	0.1%
Walk	358	31.8%
Cycle	325	28.8%
Sedentary	444	39.4%
Active	683	60.6%
Eligible journey pairs per day	4.96	
Sedentary journey pairs per day	3.01	
Active journeys pairs per day	1.96	

These data compare well to the crude data in terms of journey frequency (see Tables 8.6 and 8.9) and average journey duration (see Tables 8.5 and 8.10), suggesting the process of selecting matched journeys has not introduced significant bias. The journey frequency has increased as a result of looking at data on the 227 days when journeys could be matched. Days with partial data and fewer journeys were less likely to produce a matched pair.

Overall agreement (assessment of bias) was tested using the paired-samples t-test. The results are displayed in Table 8.10. Figure 8.4 shows a Bland-Altman plot of difference against average for all matched pairs; the mean difference (bias in minutes), 95% confidence interval and 95% limits of agreement are displayed on this figure.

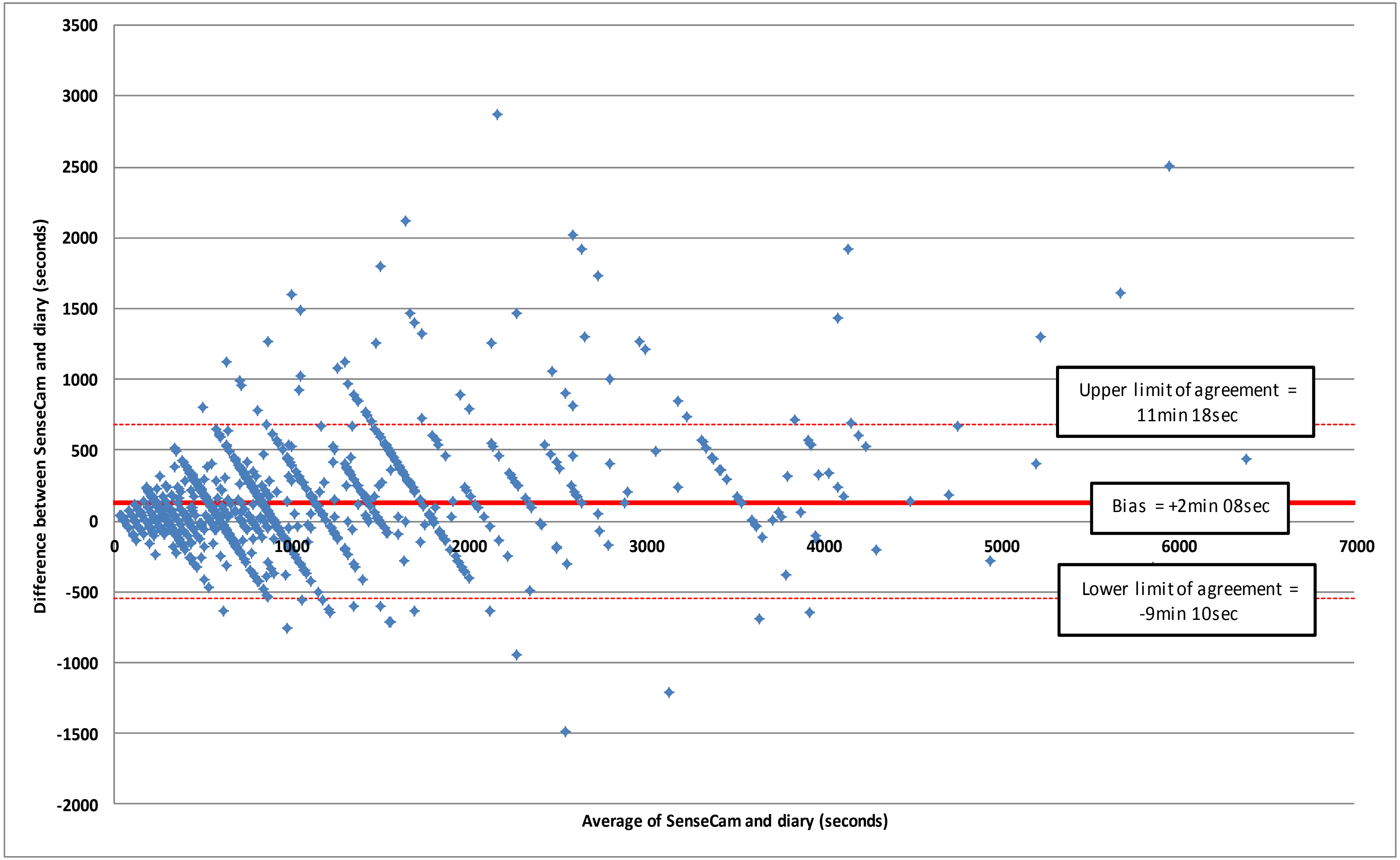
Table 8.10 Comparison of means for all matched-pairs of journeys

	Mean time (minutes)	**SE of mean (minutes)**	**SD of mean (minutes)**		
SenseCam	15:31	0:27	15:07		
Diary	17:39	0:29	16:31		
	Difference between means (minutes)	**SE of difference (minutes)**	**95% confidence interval of difference (minutes)**	**SD of difference (minutes)**	**95% limits of agreement of difference (minutes)**
Overall	2:08	0:10	1:48 to 2:28	5:46	-9:10 to 13:26

Table 8.10 shows that at the individual journey level the mean difference between SenseCam and the travel diary was 2:08 minutes. The low standard error gives a narrow 95% confidence interval around this estimate (1:48 to 2:28 minutes). The relatively large standard deviation gives wide 95% limits of agreement around this estimate.

The Bland-Altman plot illustrates the 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 level. The confidence intervals on the limits of agreement were calculated as described in Bland and Altman [161, 194] and were found to be +/- 35.0 seconds. These and the confidence intervals on the difference between means (+/-19.6 seconds) were not plotted for visual clarity. The apparent tendency for data points to lie on diagonal lines on the plot is another demonstration of the preference to report durations to the nearest five minutes.

Figure 8.4 Bland-Altman plot for all matched pairs of journeys (n = 1,127)



Agreement on mode

Considering the 1,127 matched journeys, the travel diary reported a different mode to that shown by the images on 14 occasions (1.2%). Put another way, the travel diary correctly identified mode on 98.8% of journeys. The most common mis-reporting was to record a Car Passenger journey as Car Driver (n=5).

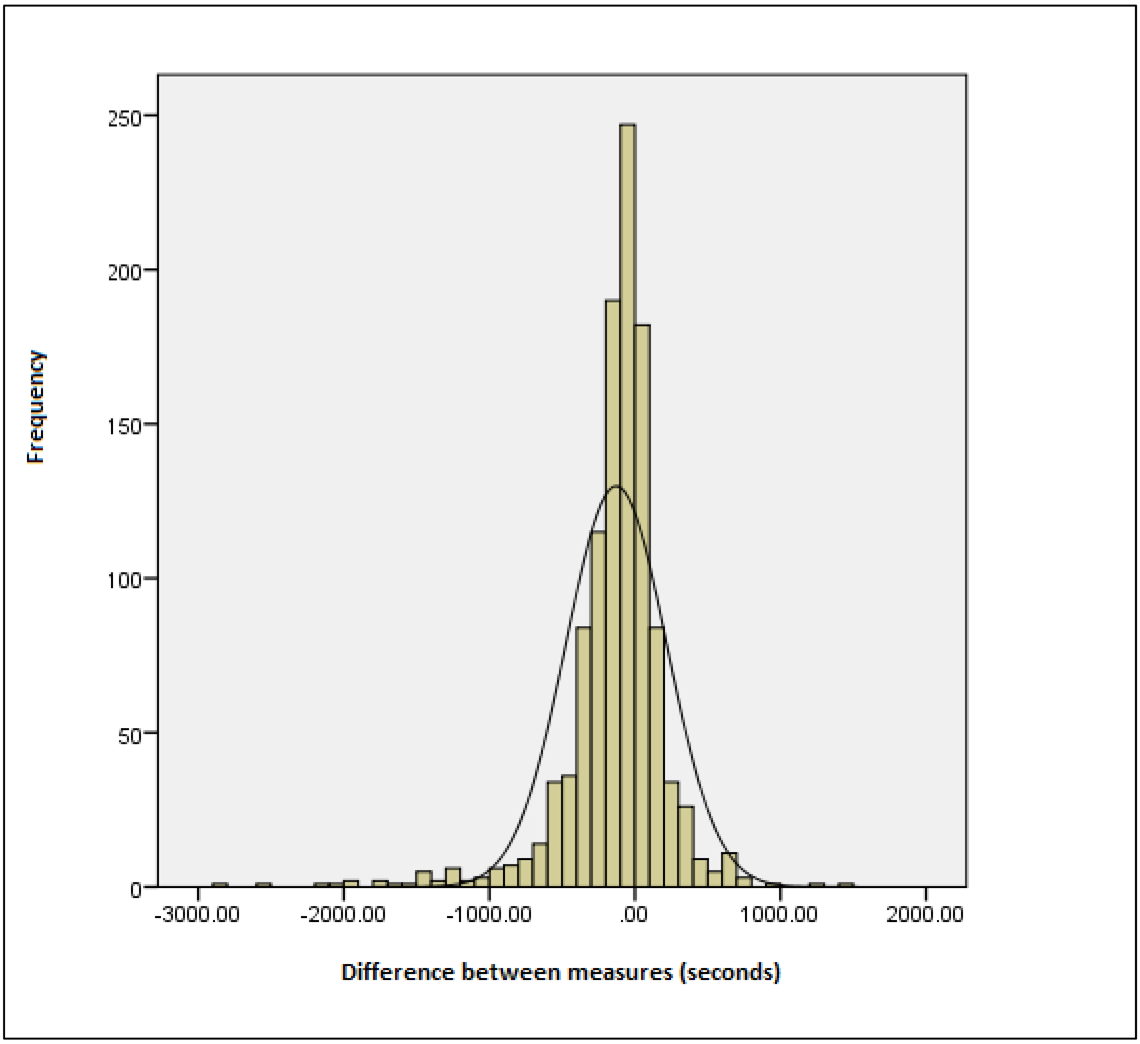
Using the Kappa statistic agreement on mode was found to be 0.983 (95%CI = 0.973 to 0.993). Landis and Koch suggest any value over 0.80 should be considered outstanding agreement between observations [210].

Assessment of normality

The paired-samples t-test used to investigate differences between means is a parametric test that assumes normally distributed differences [211]. To assess normality, I first plotted a histogram of the differences (between the 1,127 matched pairs) which is shown in Figure 8.5. The x-axis is in 60 second increments, and a normal distribution curve has been plotted using SPSS.

Initial visual inspection of the histogram shows the data fit the expected shape for normally distributed data. The Shapiro-Wilk test confirms this; statistic = 0.834; significance = 0.0005. Plotting a normal curve on the data shows the data display some mild kurtosis (peaking) and skewness (shifted to the right of the curve). I attempted some standard data transformations (log and square-root) but these did not improve the normality (results reported at end of chapter). I decided the data fit the assumptions for normal distribution and should remain untransformed.

Figure 8.5 Histogram showing frequency distribution of the differences between SenseCam and the travel diary for eligible matched pairs of journeys (n=1127)



Assessment of heteroscedasticity

If the differences between SenseCam and the travel diary increased with journey length (proportional bias) the data would be described as heteroscedastic. This would complicate the statistical tests being used as it suggests non-uniform variance [211]. The first test is visual inspection of the Bland-Altman plot in Figure 8.4. A definite "funnel shape" emanating from the origin would indicate heteroscedasticity [212]. Across the whole plot, there is little visual evidence of a funnel shape. However, it should be noted that between 0-1000 seconds on the x-axis, the data do appear to demonstrate a funnel shape before levelling off.

Presence of heteroscedasticity was then tested by calculating the Kendall's tau (τ) correlation between the absolute differences and the corresponding means. A positive τ greater than 0.1 would indicate the data were heteroscedastic; a τ less than or equal to 0.1 or a negative τ would indicate the data were non-heteroscedastic (homoscedastic). These cut-offs have been taken from a previously published study on heteroscedasticity and data transformation, though the authors acknowledge they have been chosen arbitrarily [212]. For my 1,127 matched-pairs of measurements Kendall's Tau was found to be 0.100 (significant at 0.01 level) and therefore deemed to be non-heteroscedastic. However, it should be noted that this value sits right on the boundary level of 0.1.

Between- and within- subject differences

The correlation coefficient was calculated for the relationship between the journey durations estimated from the two methods using both within-subject [195] and between-subject methods [213]. The between-subject correlation coefficient was found to be r = 0.953. The within-subject correlation coefficient was found to be r = 0.924.

This tells us that greater reported journey durations are strongly associated with greater SenseCam recorded journey durations (and vice versa). These are both considered high correlations in physical activity measurement; a value >0.80 is said to demonstrate acceptable validity [214]. 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.

Principal finding 2

There is a significant fixed bias of 2:08 minutes on reported journey durations. This represents over-reporting of journey duration and systematic error on self-reported durations. The wide limits-of-agreement (-9:10 to 13:26 minutes) demonstrate the large range (both positive and negative) for journey mis-reporting at an individual journey level. This represents large random error.

There is very high agreement between measures for mode (98.8% of journeys).

Research question 3 – is the bias on self-reported journey duration different according to certain variables?

As defined in the Chapter 7 (Study design), there are certain variables that are thought to influence error on self-reported travel behaviour. I classified these as journey, participant and study characteristics. I conducted a series of further paired-samples t-tests to investigate if there were any variables that might predict greater of lesser error on the travel diary. I tabularise the results by variable and show the mean difference graphically. I chose not to present Bland-Altman plots for each variable as differences were difficult to interpret visually.

Journey characteristics

Mean differences by journey mode

Table 8.11 shows the mean differences by mode and these are also displayed in Figure 8.6. The greatest mean difference (representing over-reporting and ignoring the single boat journey) was for cycling journeys; train journeys were the only under-reported mode though this difference was not significant, probably due to only having 13 matched pairs.

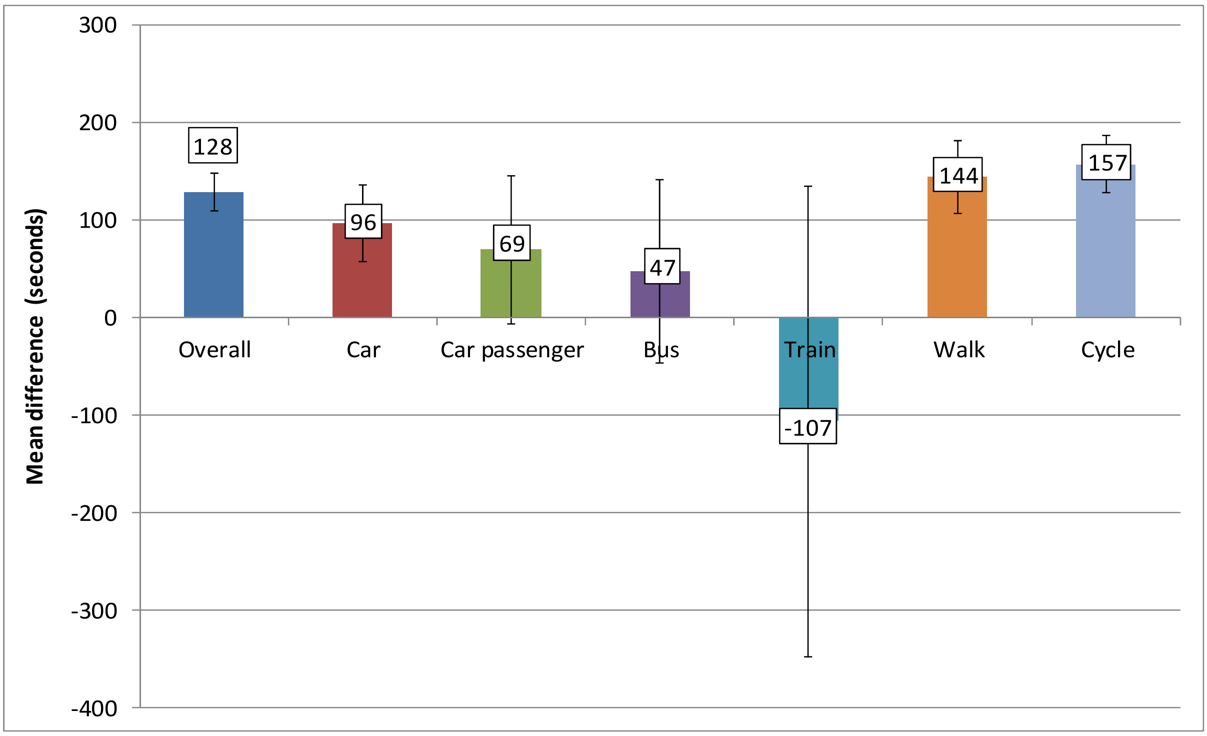
While active modes seem to be over-reported to a greater extent, comparing confidence intervals[27] suggests the only significant difference was between car and cycle journeys (Figure 8.6). It is worth noting that self-reported bus durations from the matched-pairs are much lower than in the crude data. This is because many of the journeys missed by SenseCam were longer (60-120 minutes) bus journeys that are excluded from matched pair analysis. They mainly came from one participant and there is no obvious reason why SenseCam was particularly poor at recording these journeys but this may introduce bias. Therefore, it is unclear if the small non-significant difference between means for bus journeys is unrepresentative. The single boat journey is not displayed.

[27] This is the approach used for comparisons between means for all descriptive variables in RQ3.I acknowledge this is just an indication of significance and formal tests are warranted. A post-hoc ANOVA has been suggested as a way to remove the possibility of chance findings.

Table 8.11 Comparison of difference between means for matched-pairs of journeys by journey mode

	Mean SenseCam time (minutes)	Mean travel diary time (minutes)	Difference between means (minutes)	SE of difference (minutes)	95% confidence interval of difference (minutes)	SD of diff (min-utes)	95% limits of agreement (minutes)
Overall	15:31	17:39	2:08	0:10	1:48 to 2:28	5:46	-9:10 to 13:26
Car (n=312)	14:42	16:19	1:36	0:20	0:57 to 2:16	5:55	-10:00 to 13:12
Car passenger (n=84)	17:08	18:18	1:09	0:39	-0:08 to 2:27	5:58	-10:33 to 12:51
Bus (n=34)	19:31	20:18	0:47	0:48	-0:51 to 2:26	4:44	-8:30 to 10:04
Train (n=13)	20:51	19:04	-1:47	2:03	-6:17 to 2:42	7:26	-12:47 to 16:21
Boat (n=1)	52:43	60:00	7:17	-	-	-	-
Walk (n=358)	11:27	13:52	2:24	0:19	1:45 to 3:04	6:16	-9:53 to 14:41
Cycle (n=325)	19:36	22:28	2:52	0:15	2:21 to 3:24	4:48	-6:32 to 12:17

Figure 8.6 Difference between means for SenseCam and travel diary by journey mode



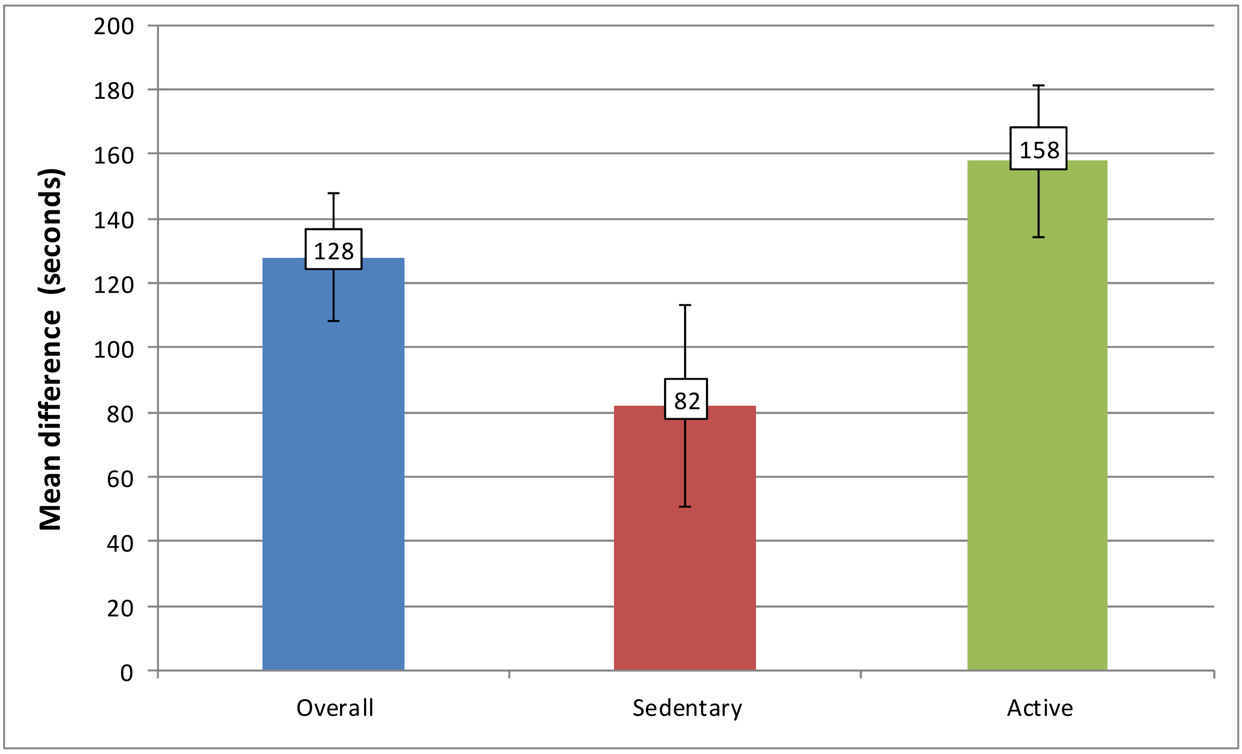
Mean differences by sedentary and active travel

Table 8.12 shows the mean differences by sedentary and active modes and these are also displayed in Figure 8.7. While both categories are over-reported, the active modes are by a greater extent and this difference appears to be significant.

Table 8.12 Comparison of difference between means for sedentary and active travel modes

	Mean SenseCam time (minutes)	Mean travel diary time (minutes)	Difference between means (minutes)	SE of difference (minutes)	95% confidence interval of difference (minutes)	SD of difference (minutes)	95% limits of agreement (minutes)
Overall	15:31	17:39	2:08	0:10	1:48 to 2:28	5:46	-9:10 to 13:26
Sedentary (n=444)	15:48	17:10	1:22	0:16	0:49 to 1:55	5:54	-10:12 to 12:56
Active (n=683)	15:19	17:57	2:38	0:12	2:12 to 3:03	5:37	8:23 to 13:39

Figure 8.7 Difference between means for SenseCam and travel diary by sedentary and active travel



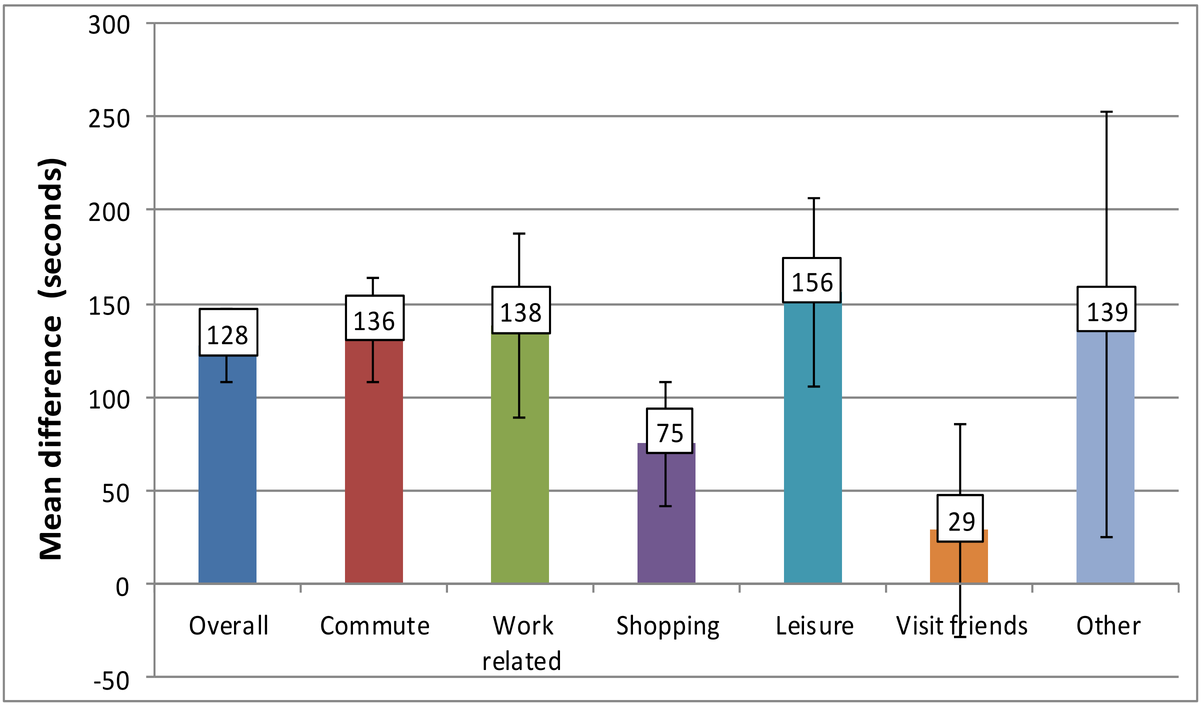
Mean differences by journey purpose

The differences varied by travel purpose as shown in Table 8.13 and Figure 8.8. Shopping trips appear to have a lower bias than other purposes, except trips to visit friends which have no significant bias at all.

Table 8.13 Comparison of means for matched-pairs of journeys by journey purpose

	Mean SenseCam time (minutes)	Mean travel diary time (minutes)	Difference between means (minutes)	SE of difference (minutes)	95% confidence interval of difference (minutes)	SD of diff (min-utes)	95% limits of agreement (minutes)
Overall	15:31	17:39	2:08	0:10	1:48 to 2:28	5:46	-9:10 to 13:26
Commute (n=380)	18:40	20:56	2:16	0:14	1:46 to 2:45	4:50	-7:12 to 11:44
Work related (n=192)	9:34	11:33	2:18	0:25	1:28 to 3:09	5:53	-9:14 to 13:50
Shopping (n=181)	11:52	13:08	1:15	0:17	0:42 to 1:49	3:51	-6:18 to 8:48
Leisure (n=283)	17:48	20:24	2:36	0:26	1:43 to 3:28	7:30	-12:06 to 17:18
Visit friends (n=37)	14:35	15:04	0:29	0:29	-0:30 to 1:29	2:59	-5:22 to 6:20
Other (n=54)	16:28	18:47	2:19	0:58	0:21 to 4:17	7:12	-11:48 to 16:26

Figure 8.8 Difference between means for SenseCam and travel diary by journey purpose



Mean differences by journey duration

The differences varied by travel duration as shown in Table 8.14 and Figure 8.9. These data show a clear increase in difference between means as journey duration increases. This finding seems to contrast the heteroscedasticity testing presented earlier. For this reason I decided to look at difference between means as proportional differences. Table 8.15 shows proportional differences by the average and the midpoint of the duration categories. These data show that there is a small but noticeable increase in proportional difference. However this increase appears to stop after the 11-15 minute category. Returning to the Bland-Altman plot in Figure 8.4 there is a funnel shape to about this point (660-900 seconds).

This would suggest that there is a small proportional bias to approximately 15 minutes and after that the bias is constant. This may explain why a log transform did not improve the results for heteroscedasticity (a similar finding to Pilot study 2 [157]).

Table 8.14 Comparison of means for matched-pairs of journeys by journey duration

	Mean SenseCam time (minutes)	Mean travel diary time (minutes)	Difference between means (minutes)	SE of difference (minutes)	95% confidence interval of difference (minutes)	SD of diff (min-utes)	95% limits of agreement (minutes)
Overall	15:31	17:39	2:08	0:10	1:48 to 2:28	5:46	-9:10 to 13:26
0-5 min (n=240)	4:07	4:16	0:08	0:07	-0:07 to 0:24	2:03	-3:53 to 4:09
6-10 min (n=295)	8:22	9:12	0:50	0:10	0:29 to 1:11	3:02	-5:07 to 6:47
11-15 min (n=228)	12:44	14:15	1:31	0:15	0:59 to 2:02	3:57	-6:14 to 9:16
16-30 min (n=233)	21:08	24:36	3:27	0:24	2:38 to 4:16	6:21	-9:00 to 15:54
31-120 min (n=131)	47:16	54:46	7:29	0:56	5:37 to 9:21	10:46	-13:37 to 28:35

Figure 8.9 Difference between means for SenseCam and travel diary by journey duration category

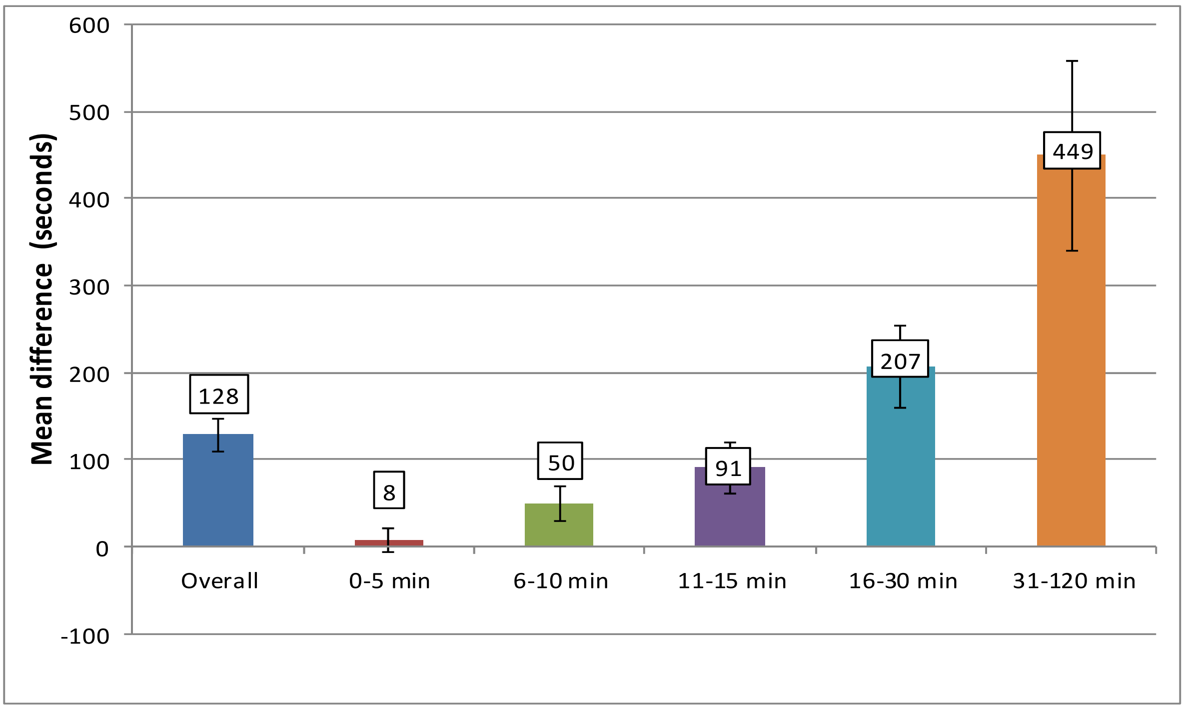


Table 8.15 Proportional differences by the average and the midpoint of the duration categories

	Difference between means (minutes)	Difference as % of average reported time of range	Difference as % of mid-point of reported range
Overall	2:08	12.1	3.5
0-5 min (n=240)	0:08	3.1	5.3
6-10 min (n=295)	0:50	9.1	10.4
11-15 min (n=228)	1:31	10.6	11.7
16-30 min (n=233)	3:27	14.0	15.0
31-120 min (n=131)	7:29	13.7	10.0

Mean differences by rounded to five minutes or not rounded

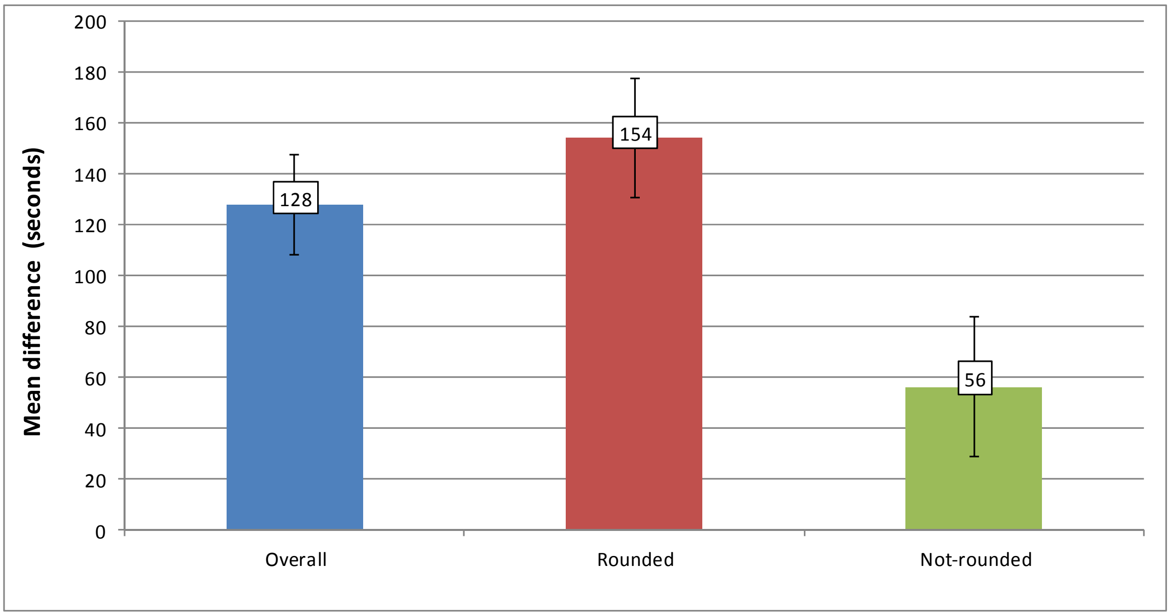
The differences varied by whether or not the journey was reported to the nearest five minutes. These data are shown in Table 8.16 and Figure 8.10. It is assumed this digit preference introduces error as the actual times are rounded from the true value.

Table 8.16 Comparison of means by rounding to five minutes

	Mean SenseCam time (minutes)	Mean travel diary time (minutes)	Difference between means (minutes)	SE of difference (minutes)	95% confidence interval of difference (minutes)	SD of difference (minutes)	95% limits of agreement (minutes)
Overall	15:31	17:39	2:08	0:10	1:48 to 2:28	5:46	-9:10 to 13:26
Rounded to five minutes (n=831)	16:46	19:20	2:34	0:12	2:08 to 2 59	6:11	-9:33 to 14:41
Not rounded (n=296)	12:56	12:00	0:56	0:14	0:28 to 1:24	4:07	-7:08 to 9:00

That the bias is greater for rounded journeys suggests that rounding is generally up rather than down, and introduces systematic error. The difference appears to be significant. It is also worth noting that the rounded journeys were approximately 25% greater in duration than the non-rounded journeys, suggesting longer journeys are more likely to be rounded to five (or 10) minutes.

Figure 8.10 Difference between means for SenseCam and travel diary by rounded to five minutes or not rounded



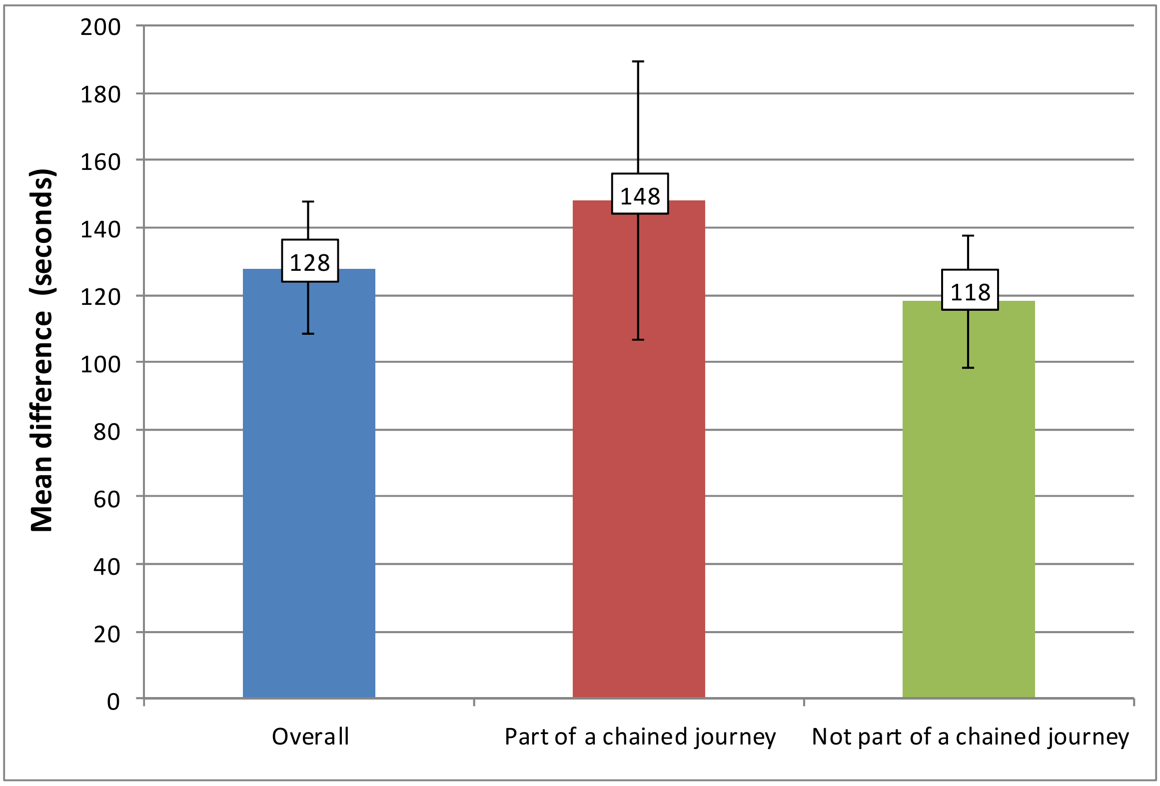
Mean differences by journey chaining

Journey chaining is defined in my protocol as a series of linked trips (e.g. walk to bus stop, bus to town, and walk to shops). Table 8.17 and Figure 8.11 show the comparison of journeys that were part of a series (chained) and those which were independent (not chained). The data suggest there were no significant differences and journeys in both categories displayed similar positive bias.

Table 8.17 Comparison of means by journey chaining

	Mean SenseCam time (minutes)	Mean travel diary time (minutes)	Difference between means (minutes)	SE of difference (minutes)	95% confidence interval of difference (minutes)	SD of difference (minutes)	95% limits of agreement (minutes)
Overall	15:31	17:39	2:08	0:10	1:48 to 2:28	5:46	-9:10 to 13:26
Chained (n=398)	15:32	17:58	2:28	0:21	1:43 to 3:08	7:11	-11:37 to 16:33
Not chained (n=729)	15:30	17:29	1:58	0:10	1:37 to 2:19	4:49	-7:28 to 11:24

Figure 8.11 Difference between means for SenseCam and travel diary by journey chaining



Principal finding 3.1

The following journey characteristics appeared to have a significant effect on the differences between bias:

- Journey duration (greater duration displayed greater positive bias)
- Journeys reported to 5 minute intervals (rounded journeys display greater positive bias)
- Whether journey is active or sedentary (active journeys display greater positive bias)

The following journey characteristics appeared to have no significant effect on the differences between bias:

- Journey mode (except car versus cycle)
- Journey purpose (except visiting friends which has significantly lower bias than most other purposes)
- Whether journey is a chain of linked stages

Participant characteristics

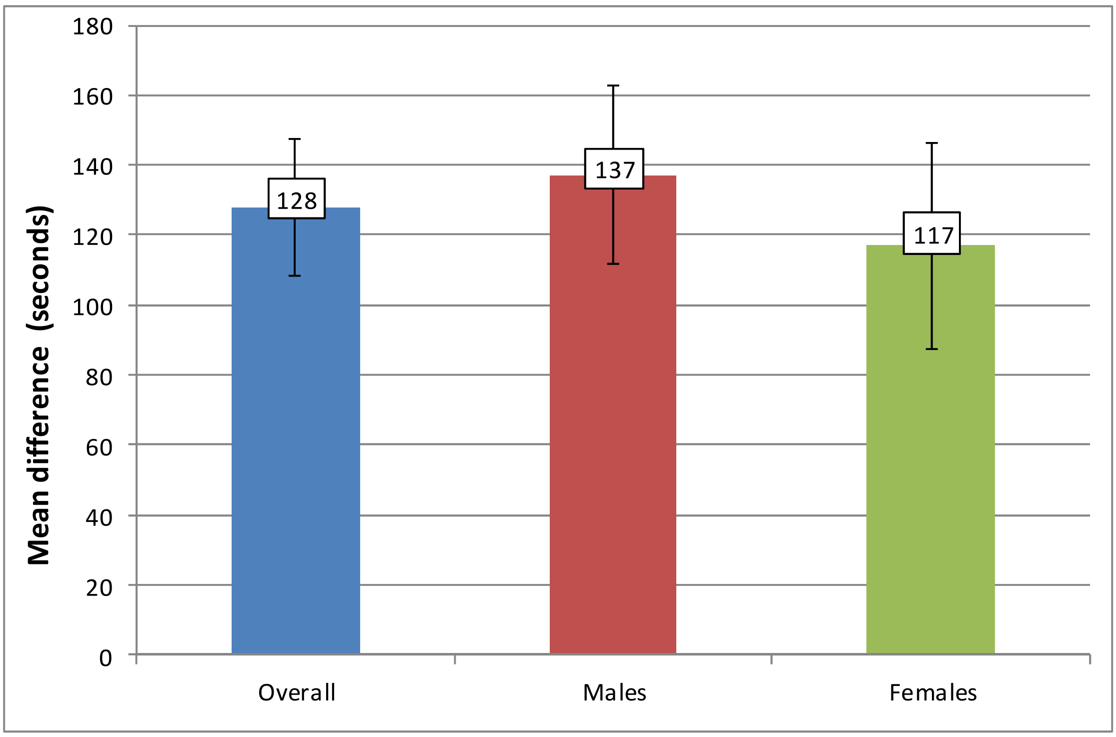
Mean differences by participant gender

The differences between biases by gender are shown in Table 8.18 and Figure 8.12. The data show that both males and females have a positive bias, and gender does not significantly influence the magnitude of this over-reporting bias.

Table 8.18 Comparison of means by gender

	Mean SenseCam time (minutes)	Mean travel diary time (minutes)	Difference between means (minutes)	SE of difference (minutes)	95% confidence interval of difference (minutes)	SD of difference (minutes)	95% limits of agreement (minutes)
Overall	15:31	17:39	2:08	0:10	1:48 to 2:28	5:46	-9:10 to 13:26
Male (n=626)	15:20	17:38	2:17	0:13	1:50 to 2:45	5:50	-9:09 to 13:43
Female (n=501)	15:44	17:41	1:57	0:15	1:27 to 2:27	5:40	-9:09 to 13:03

Figure 8.12 Difference between means for SenseCam and travel diary by participant gender



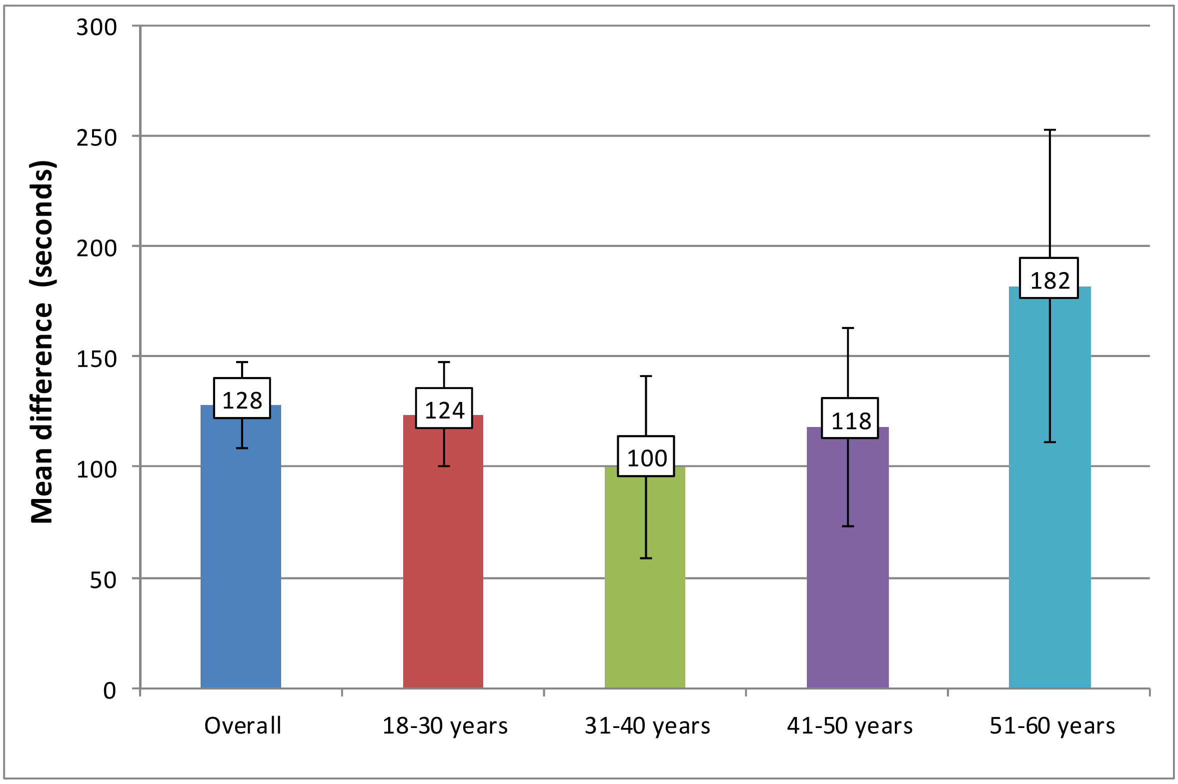
Mean differences by participant age group

I categorized participants into 10 year age groups. These data are displayed in Table 8.19 and Figure 8.13. They show that there is very little difference between the 18-30, 31-40 and 41-50 age groups; all have a similar positive bias. The 51-60 years age group appears to have a greater positive bias, but the difference to the other groups is non-significant. I tested 18-24 and 25-30 years age categories, but the results were highly similar so I grouped these categories.

Table 8.19 Comparison of means by age category

	Mean SenseCam time (minutes)	**Mean travel diary time (minutes)**	**Difference between means (minutes)**	**SE of difference (minutes)**	**95% confidence interval of difference (minutes)**	**SD of difference (minutes)**	**95% limits of agreement (minutes)**
Overall	15:31	17:39	2:08	0:10	1:48 to 2:28	5:46	-9:10 to 13:26
18-30 years (n=523)	14:53	16:58	2:04	0:12	1:40 to 2:29	4:49	-7:22 to 11:30
31-40 years (n=240)	14:42	16:23	1:40	0:21	0:58 to 2:21	5:25	-8:57 to 12:17
41-50 years (n=170)	14:34	16:32	1:58	0:23	1:11 to 2:44	5:07	-8:04 to 12:00
51-60 years (n=194)	19:01	22:04	3:02	0:36	1:51 to 4:13	8:24	-13:26 to 19:30

Figure 8.13 Difference between means for SenseCam and travel diary by age category



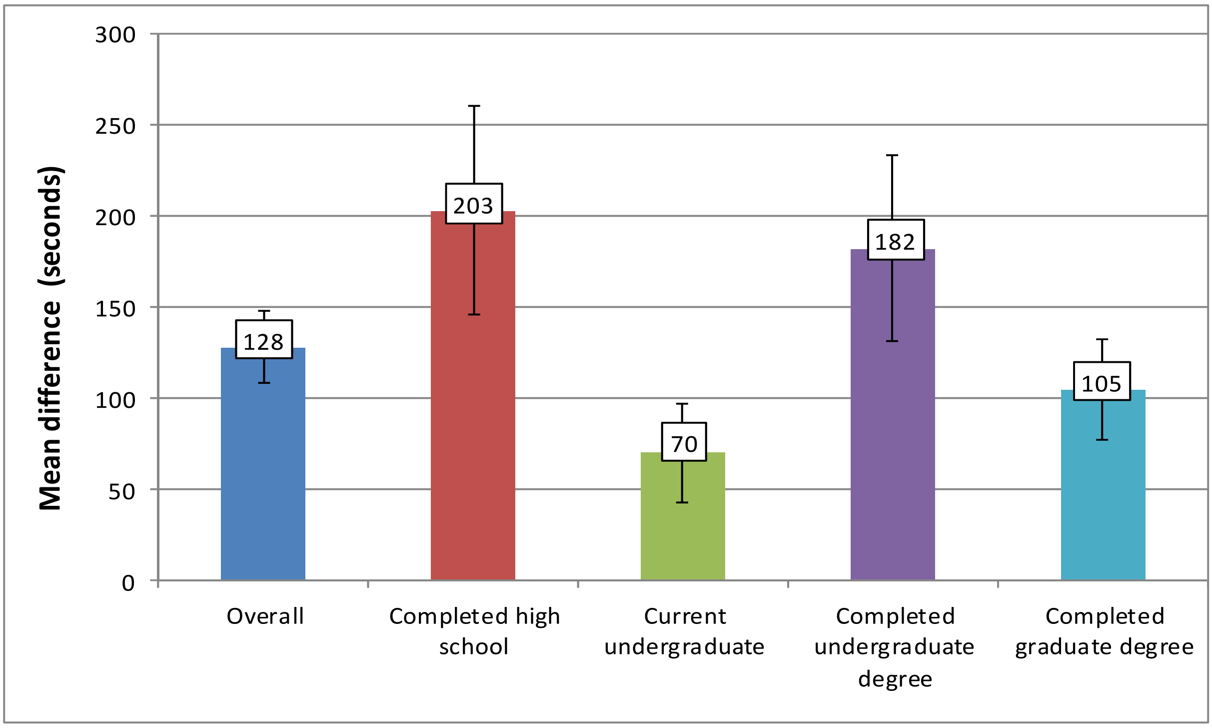
Mean differences by education status

The differences between biases by educational status are shown in Table 8.20 and Figure 8.14. There is no clear pattern here. Ignoring current undergraduates, there seems to be a trend for less bias as educational status increases, with a significant difference of 1:38 minutes per journey between participants who completed high school and participants who have completed a graduate degree. However, current undergraduates did not follow the trend and have a lower bias than any other group. This may be related to the fact they have the shortest journeys in terms of duration.

Table 8.20 Comparison of means by education status

	Mean SenseCam time (minutes)	Mean travel diary time (minutes)	Difference between means (minutes)	SE of difference (minutes)	95% confidence interval of difference (minutes)	SD of difference (minutes)	95% limits of agreement (minutes)
Overall	15:31	17:39	2:08	0:10	1:48 to 2:28	5:46	-9:10 to 13:26
Completed high school (n=133)	15:45	19:08	3:23	0:29	2:24 to 4:22	5:44	-7:51 to 14:37
Current undergraduate (n=226)	13:05	14:16	1:10	0:14	0:42 to 1:39	3:38	-5:57 to 8:17
Completed undergraduate degree (n=269)	18:14	21:16	3:02	0:26	2:10 to 3:54	7:13	-10:49 to 17:29
Completed graduate degree (n=499)	15:05	16:50	1:45	0:14	1:16 to 2:15	5:33	-9:08 to 12:38

Figure 8.14 Difference between means for SenseCam and travel diary by educational status



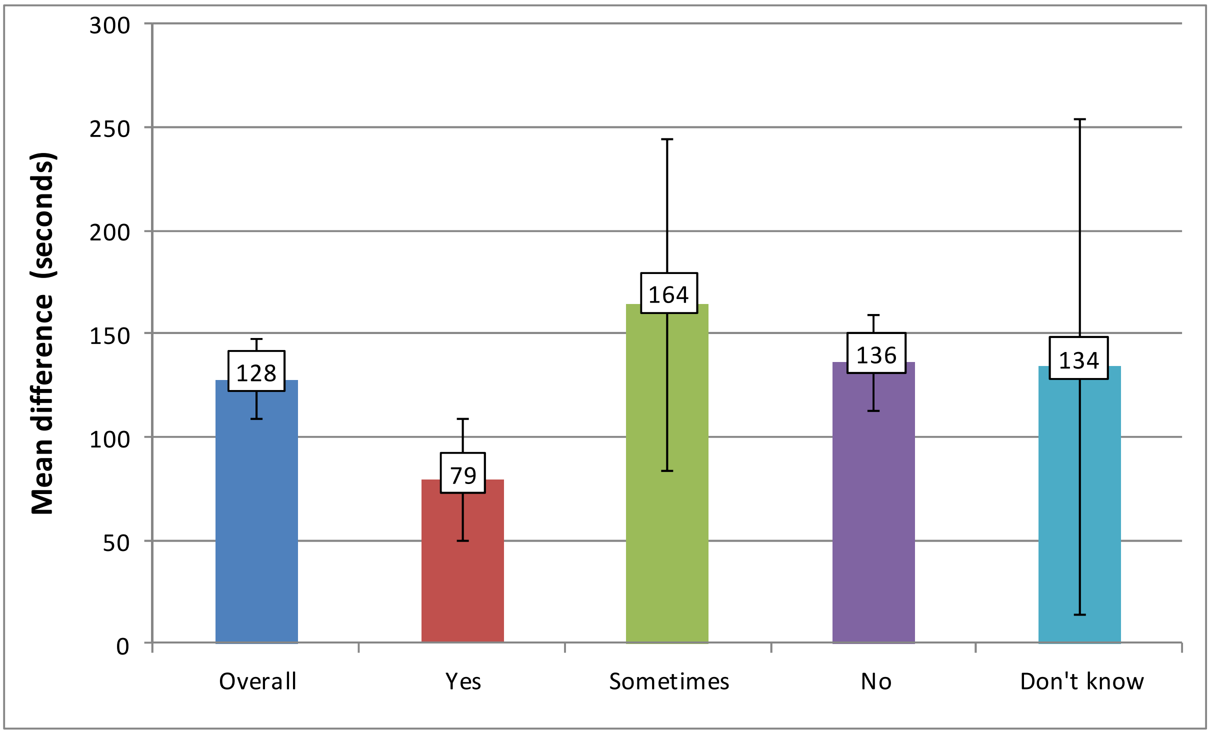
Mean differences by use of memory jogger

As previously mentioned, the participants were provided with a pocket sized memory jogger (as in the real NTS). The idea was for participants to note their travel during the day and transfer it to the diary at the end of the day. At the satisfaction interview, participants were asked if they had used it. The differences between biases by jogger use are shown in Table 8.21 and Figure 8.15. There is a greater positive bias (over-report) when the jogger is not used to when it is. The difference appears to be significant. There is no difference between sometimes using the jogger and never using the jogger.

Table 8.21 Comparison of means by memory jogger use

	Mean SenseCam time (minutes)	Mean travel diary time (minutes)	Difference between means (minutes)	SE of difference (minutes)	95% confidence interval of difference (minutes)	SD of difference (minutes)	95% limits of agreement (minutes)
Overall	15:31	17:39	2:08	0:10	1:48 to 2:28	5:46	-9:10 to 13:26
Yes (n=233)	15:40	17:00	1:19	0:15	0:48 to 1:50	3:59	-6:29 to 9:07
Sometimes (n=142)	17:35	20:20	2:44	0:41	1:22 to 4:07	8:14	-13:44 to 18:32
No (n=724)	14:55	17:12	2:16	0:12	1:52 to 2:41	5:39	-8:48 to 13:20
Don't know (n=28)	19:06	21:21	2:14	1:01	0:09 to 4:20	5:24	-8:21 to 12:49

Figure 8.15 Difference between means for SenseCam and travel diary by memory jogger use



Principal finding 3.2

The following participant characteristics appeared to have a significant effect on the differences between bias:

- Whether the jogger is used (greater positive bias when jogger not used)
- Educational status (no obvious pattern)

The following participant characteristics appeared to have no significant effect on the differences between bias:

- Gender
- Age category

Study characteristics

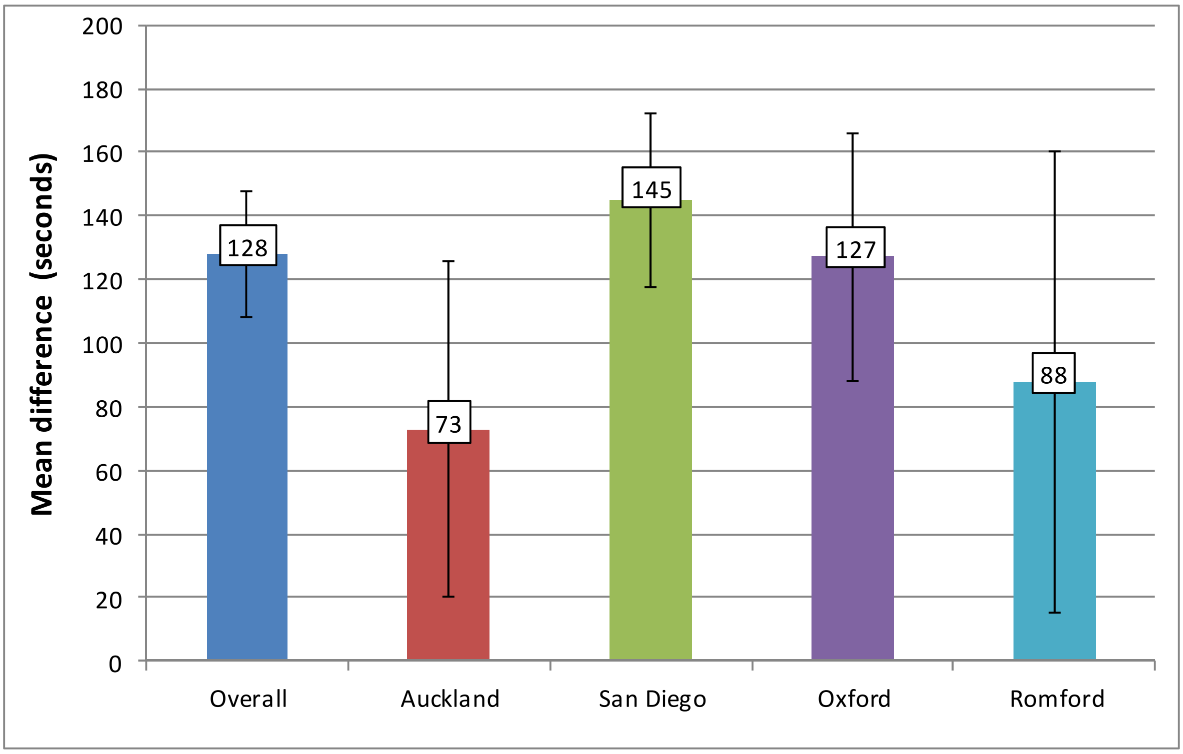
Mean differences by study location

The differences between biases by study location are shown in Table 8.22 and Figure 8.16. There were no significant differences, although Auckland and Romford had a lower bias than San Diego and Oxford. In all locations mean journey durations were over-reported.

Table 8.22 Comparison of means by study location

	Mean SenseCam time (minutes)	**Mean travel diary time (minutes)**	**Difference between means (minutes)**	**SE of difference (minutes)**	**95% confidence interval of difference (minutes)**	**SD of difference (minutes)**	**95% limits of agreement (minutes)**
Overall	15:31	17:39	2:08	0:10	1:48 to 2:28	5:46	-9:10 to 13:26
Auckland (n=151)	15:11	16:24	1:13	0:27	0:19 to 2:06	5:34	-9:42 to 12:08
San Diego (n=648)	16:28	18:53	2:25	0:14	1:57 to 2:52	5:58	-9:17 to 14:07
Oxford (n=277)	14:03	16:11	2:07	0:20	1:28 to 2:47	5:33	-8:46 to 13:00
Romford (n=51)	12:20	13:48	1:28	0:37	0:13 to 2:42	5:17	-8:53 to 11:49

Figure 8.16 Difference between means for SenseCam and travel diary by study location



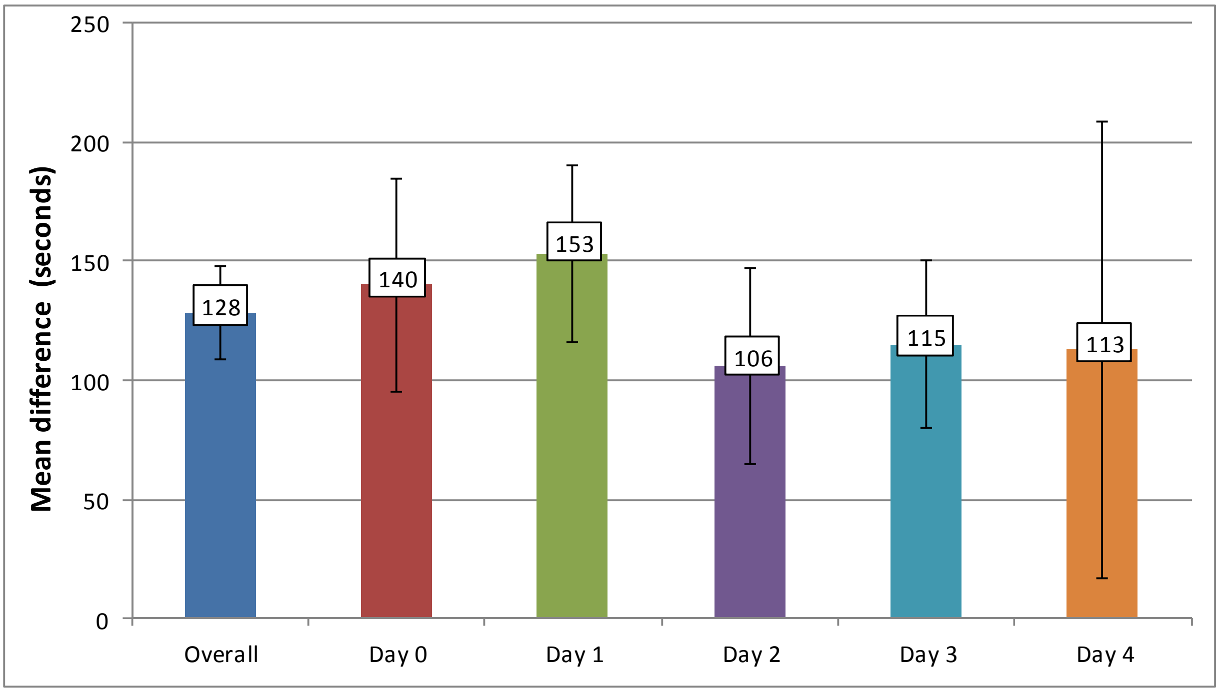
Mean differences by day of study

The differences between biases by day of study are shown in Table 8.23 and Figure 8.17. On all days the bias was positive indicating over-reporting, and the bias was greater on the first two days than the third, fourth and fifth days. None of the differences appeared to be significant.

Table 8.23 Comparison of means by study day

	Mean SenseCam time (minutes)	Mean travel diary time (minutes)	Difference between means (minutes)	SE of difference (minutes)	95% confidence interval of difference (minutes)	SD of difference (minutes)	95% limits of agreement (minutes)
Overall	15:31	17:39	2:08	0:10	1:48 to 2:28	5:46	-9:10 to 13:26
Day 0 (n=142)	15:04	17:24	2:20	0:23	1:33 to 3:07	4:42	-6:53 to 11:33
Day 1 (n=368)	14:39	17:13	2:33	0:19	1:55 to 3:12	6:16	-9:44 to 14:50
Day 2 (n=305)	17:26	19:12	1:46	0:21	1:04 to 2:27	6:07	-10:13 to 13:45
Day 3 (n=285)	14:28	16:23	1:55	0:18	1:18 to 2:32	5:16	-8:24 to 12:14
Day 4 (n=27)	18:45	20:38	1:53	0:49	0:11 to 3:36	4:19	-6:35 to 10:21

Figure 8.17 Difference between means for SenseCam and travel diary by study day



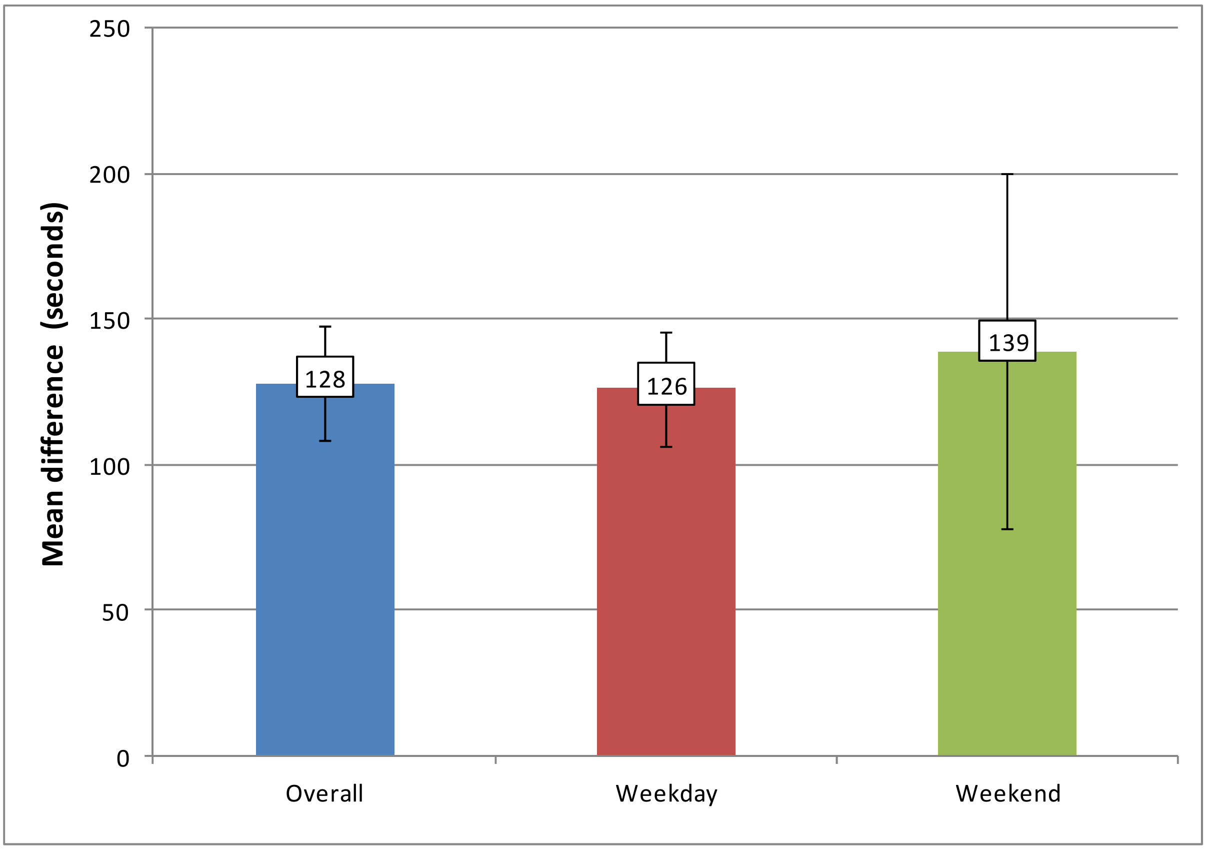
Mean differences by weekend or weekday

The differences between biases by weekday or weekend day are shown in Table 8.24 and Figure 8.18. There is no obvious difference by this variable.

Table 8.24 Comparison of means by weekend or weekday

	Mean SenseCam time (minutes)	Mean travel diary time (minutes)	Difference between means (minutes)	SE of difference (minutes)	95% confidence interval of difference (minutes)	SD of difference (minutes)	95% limits of agreement (minutes)
Overall	15:31	17:39	2:08	0:10	1:48 to 2:28	5:46	-9:10 to 13:26
Weekday (n=954)	15:08	17:14	2:06	0:10	1:45 to 2:27	5:31	-8:43 to 12:55
Weekend (n=173)	17:37	19:37	2:19	0:31	1:16 to 3:22	7:00	-11:24 to 16:02

Figure 8.18 Difference between means for SenseCam and travel diary by weekend and weekday



Principal finding 3.3

The following study characteristics appeared to have no significant effect on the differences between bias:

- Study location (though large substantial differences were apparent)
- Study day (though greater bias on first two days)
- Week day or weekend day

Research question 4 - How well does the travel diary measure daily summary travel behaviour compared to SenseCam (criterion validity)

This final research question looked at the daily summary travel behaviour. This is the type of data many epidemiological studies would use to classify activity status [53], and could be a relevant question to a public health audience.

To answer this research question I returned to the raw data and identified all days when the camera missed nothing. This was defined by days when there were no journeys in the travel diary missed by SenseCam, and there were no partial SenseCam journeys (start or end of journey missed due to device failure). This is based on an assumption that there were no journeys missed by both measures.

With these days identified I selected the journeys that took place on these days and calculated the summary travel duration. I consider the SenseCam summary duration the criterion measure for the purposes of this RQ[28]. I then compared these summaries to (1) the summaries on the corresponding travel diary days of data and (2) the summaries on the crude (total) travel diary data. As the daily summary duration is a function of journey duration and number of trips, both the reporting of duration and frequency will influence the bias. The method for selecting SenseCam data is shown in Figure 8.19 - Study flow chart 2.

Table 8.25 shows the average daily summary travel duration for (1) crude (total) SenseCam data, (2) the criterion SenseCam days, (3) the corresponding travel diary days and (4) the crude (total) travel diary data. SenseCam was a criterion measure of daily summary travel behaviour on 202 days (which represents 73.7% of the 274 days any SC data was collected, and 65.8% of the 307 days the device was expected to collect data).

[28] This has only been tested at the individual journey level rather than at the daily summary level. It is assumed the validity and reliability of SenseCam shown at an individual level translates to the daily summary level with the condition that the participant provides a full set of images. In the future this assumption should be tested, but is beyond the scope of this thesis.

Figure 8.19 Study flow chart 2 for Research question 4

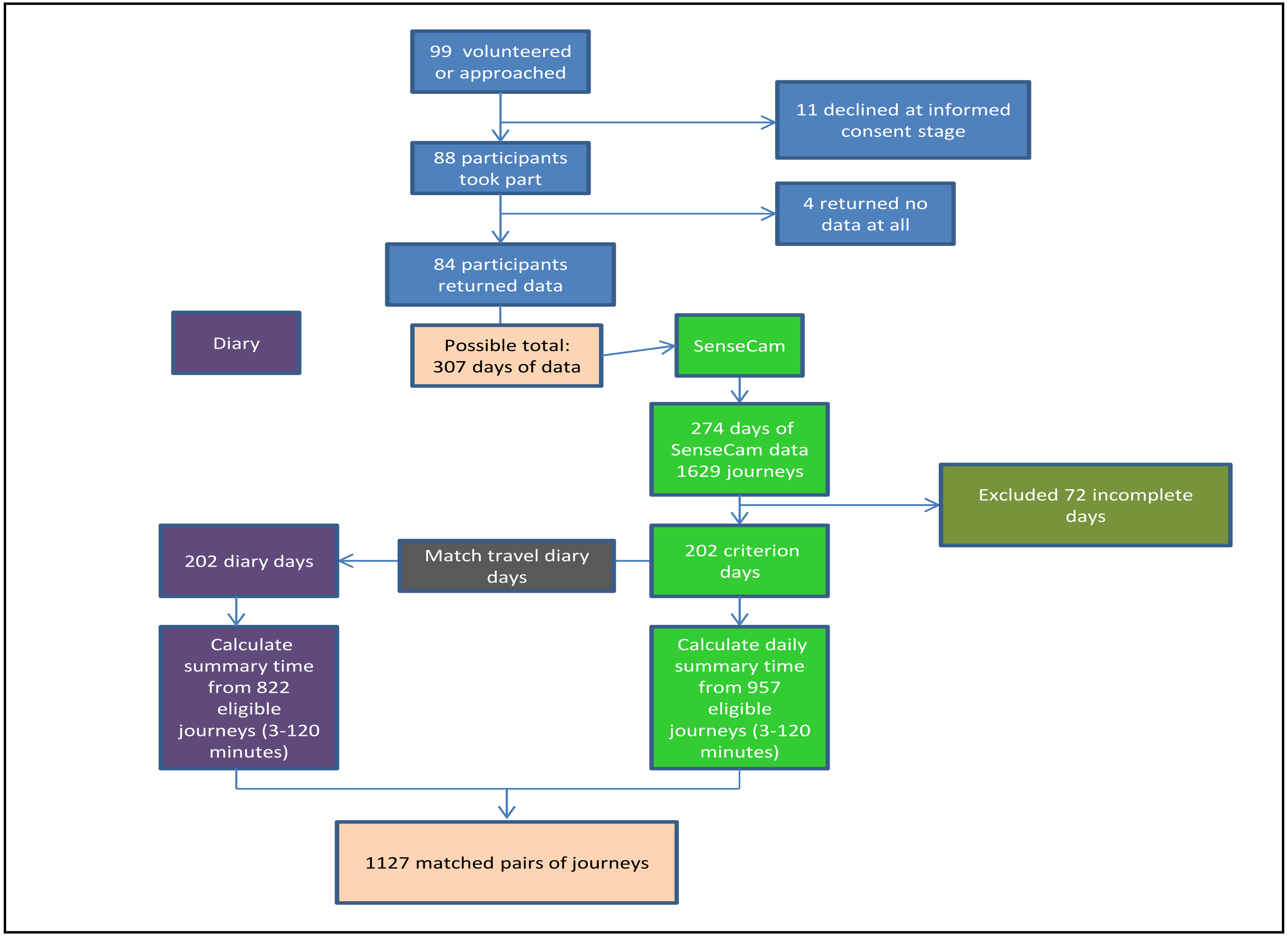


Table 8.25 Daily summary travel time analyses for all journeys 3-120 minutes

	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	202	278	202
Journeys recorded (n)	1270	957	1328	822
Journeys per day (n)	4.64 (SE = 0.15; SD = 2.48)	4.74 (SE = 0.17; SD = 2.44)	4.78 (SE= 0.15 ; SD = 2.55)	4.07 (SE = 0.15; SD = 2.12)
Mean journey duration (minutes)	15:39 (SE = 0:25; SD = 15:03)	16:03 (SE = 0:30; SD = 15:36)	17:46 (SE = 0:27; SD = 16:44)	18:49 (SE = 0:36; SD = 17:11)
Mean daily summary travel time (minutes)	**72:34 (SE = 3:09; SD = 52:02)**	**76:05 (SE = 3:42; SD = 52:41)***	**84:55 SE = 3:22; SD = 56:10)**	**76:09 (SE = 3:42; SD = 52:41)***

* I have re-checked and confirmed the standard error and standard deviation for these values are the same to the nearest second

Only taking the 202 criterion days of SenseCam (non-significantly) increases journeys per day and mean summary travel time per day compared to the crude SenseCam data. This would be expected with incomplete days excluded. There is also a small non-significant increase in mean journey duration. Looking at the travel diary for the corresponding 202 days, journeys per day and mean summary travel time are significantly lower. Mean journey duration also increases by a small amount.

Comparing the criterion SenseCam result for mean summary travel time, this is 8:50 minutes lower than the crude travel data and 0:04 min lower than the corresponding diary data.

Conducting a paired-samples t-test for the difference between criterion SenseCam days and corresponding diary days reveals the standard error to be 1:50 minutes and the standard deviation to be 25:58 minutes. This leads to a 95% confidence interval of -3:32 to 3:40 minutes and 95% limits-of-agreement of -50:50 to 50:58 minutes.

It is worth noting that the inclusion of 1-3 min journeys detected by SenseCam makes a non-significant 30 second difference (results not displayed), suggesting these short journeys do not play an important role in total summary travel times. However, that is not to say they are unimportant to certain areas of research.

Principal finding 4

These data show that SenseCam performed as a criterion measure of summary journey behaviour (as per my definition) on two-thirds of the days participants were asked to wear it. Selecting these criterion SenseCam days increased apparent journey frequency, mean journey duration and daily summary duration.

Selecting the corresponding days of diary data reduced apparent journey frequency, increased journey duration and decreased overall daily summary duration.

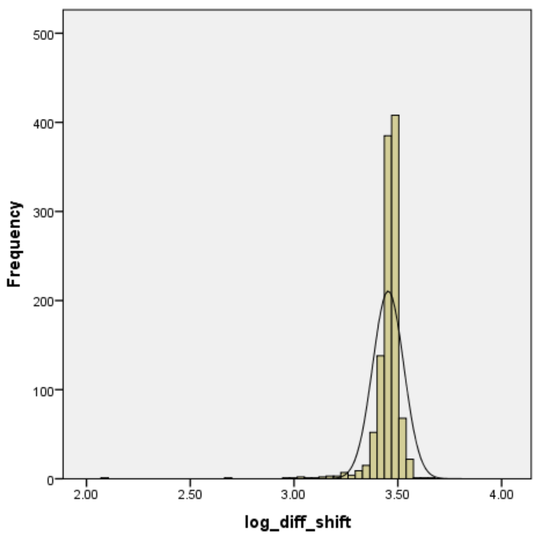
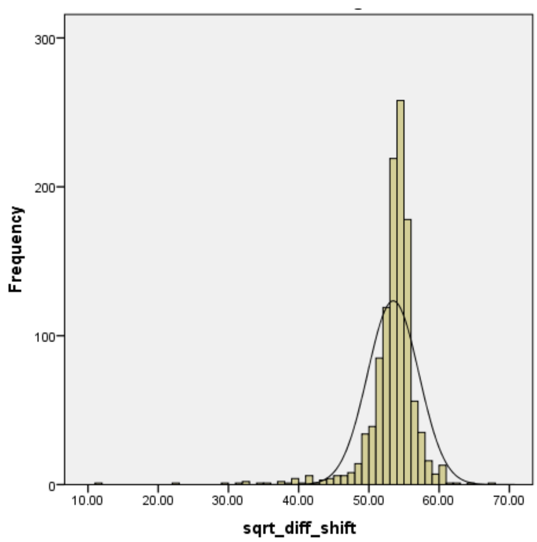
Comparing criterion SenseCam data and corresponding diary days of data suggests journey durations are over-reported, journey frequencies are under-reported and overall daily summary durations agree. Including journeys of 1-3 minutes detected by SenseCam did not affect this result.

There was a very wide 95% limits-of-agreement on the difference between daily summary durations from criterion SenseCam days and corresponding diary days.

Data transforms for normality tests

Table 8.26 shows the data transforms that were tested to improve normality of data distribution. As they did not improve normality, they were rejected.

Table 8.26 Transformations of data to improve normality

<table>
<tr><th>Log transform of differences</th><th>Square root transform of differences</th></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>Shapiro-Wilk statistic = 0.556
Significance = 0.0005</td><td>Shapiro-Wilk statistic = 0.737
Significance = 0.0005</td></tr>
</table>

Discussion of principal findings

The following section will be a discussion of the principal findings from this experiment.

Principal finding 1

SenseCam is a feasible tool for investigating travel behaviour in a sample of 80-90 participants. Data loss from both measures was approximately 10% in terms of days when no data was returned, and 15% in terms of missed journeys (as revealed by the other measure).

SenseCam and the travel diary give comparable results in terms of travel behaviour. Journey frequencies are very close, but the travel diary gives slightly longer journey durations, and as a result greater daily estimates of travel time (overall and by sedentary or active).

The first research question asked whether SenseCam could be used in a sample of this size for the first time (we already know diaries can be used in much larger samples of many thousands). It also asked how the crude data from the two measures compared. While the SenseCam was demonstrated to be feasible, 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. Whether this was 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. I would suggest it was both. Foster et al., have previously shown that the recruiting technique employed can influence the study sample [215].

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. Whether this will continue to be an issue when working with groups less familiar with research protocols remains to be seen. 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%) [120].

The choice of the four study locations was highly opportunistic. They were based on an opportunity to capitalise on my existing connections and research network. While not systematically chosen, they presented an advantage of increased diversity in terms of participants and travel environments (and in turn travel behaviours). This will reduce some threats to external validity. Likewise, the selection of participants was influenced by setting, and cannot necessarily be considered population representative. However, I did achieve a reasonable gender split and range of ages to suggest SenseCam has applications in fair proportion of groups and populations.

I considered how the data I collected compared to "real" NTS data as an indication of any bias from my methods. I conducted descriptive analysis of 2,139,541 journeys from the publicly available 2008 NTS dataset [202]. I chose to analyse the 6 most popular modes (Walk, Cycle, Drive, Vehicle passenger, Bus and Train) which comprised 2,052,864 journeys (96.0 % of the original sample).

Table 8.27 Comparison of data from the 2008 NTS and my travel diary for six most common modes

Mode or travel	**Number in 2008 NTS (n)**	**Percentage (%) of 2008 NTS total**	**Mean NTS journey duration (mins)**	**Number in my travel diary (n)**	**Percentage (%) of travel diary total**	**Mean travel diary journey duration (mins)**
Walk	241,943	11.8	24.2	400	30.1	14.4
Cycle	36,060	1.8	19.1	391	29.4	20.9
Car	1,016,402	49.5	19.3	378	28.5	16.0
Bus	150,532	7.3	32.9	52	3.9	29.7
Train	46,259	2.3	57.6	15	1.1	24.6
Car passenger	561,668	27.4	22.1	91	6.9	18.0

Inspection of Table 8.27 shows my data had a higher proportion of walks and car journeys, and a far higher proportion of cycle journeys (probably a result of the sample recruited). It had a lower proportion of bus and car passenger journeys, while train journeys were similar. Cycle, car, bus and car passenger journeys were similar in terms of mean duration; walk and train journeys were substantially shorter in my travel diary data. This suggests my travel diary data is close to "real" data, but there may be some sample bias on my results influencing generalisability particularly relating to cycling journeys.

The NTS acknowledge how sample differences can influence data, especially with relation to cycle journeys. For this reason they recommend "un-weighted samples under 100 should not be used while samples under 300 should be used cautiously" [60]. As my sample was 84, there was always going to be a limit to how representative the travel data would be.

While the compliance was high considering survey based research [53], SenseCam is not ubiquitously accepted, as demonstrated by the people who refused to take part. Their reasons were primarily to do with invasion of privacy and regulations at work (see Table 8.2). This will be an issue for any future research employing automated camera methods. In terms of administering SenseCam and providing participants with instruction, the group sessions in San Diego and Auckland proved more time-efficient than one-to-one sessions. However, it should be acknowledged that such an environment may have made some participants feel less comfortable asking questions or clarifying the protocol. Future research should balance this threat to autonomy with the efficiency benefits.

It took a substantial time (12 weeks) to code the SenseCam data. This suggests that currently the device is suited to small-medium scale studies, for example validating or calibrating measures that can be used on larger scales as in my thesis. In its current format it would not be suitable for large scale study (above 250 participants, depending on size of research team). For future studies of this size I would recommend a team of coders (appropriately trained with inter-rater reliability assessed). After my initial coding pass I decided to make a second pass. While I corrected less than 5% of coded journeys (and almost always by between one to five images, or between 10 and 50 seconds) this added a further substantial period of time to the data preparation and analysis. However it also added confidence in the coded data; it was a quality control similar to a test-retest assessment.

This experiment is an investigation of the way people report their travel behaviour. It is an inescapable fact of such research that the objective measure (SenseCam) may influence the behaviour of interest (self-report). This was discussed in the Chapter 7 – Study design. The test I devised for reactivity was based on percentage of journeys from the sample rounded to the nearest five minutes. My analysis of NTS data showed that 84.7% (95% CI 84.65-84.75) of journeys (from over 2 million) had been rounded. This number is consistent with the 85-90% reported in the Dutch Travel Survey [95]. Rounding occurred in 76.1%. (95% CI = 73.8-78.4%) of journeys in my experiment. Therefore 8.6% fewer journeys were rounded in my study than in "real" NTS self-report data and suggests there may have been mild reactivity from wearing SenseCam. This is also a larger effect than observed in Pilot study 1where 80.3% (95% CI 75.9-90.8) were rounded (though this was only from 88 journeys). It may be that the extra devices (accelerometer and GPS) used in this experiment further influenced reactivity. However, overall any effect is small, and a large proportion of journeys in my data were still rounded. For this reason I am confident that there is minimal bias influencing the internal validity of my results.

The second part of RQ1 asked how similar the data were from each measure. This is important as it tells researchers whether the measures can be used interchangeably, and how confident they can be comparing future studies that use one or other measure. Both devices performed similarly in terms of compliance and data captured. As expected the SenseCam was able to capture a substantial number of journeys <3 minutes duration which are thought to be difficult to measure by self-report [61].

In terms of journey totals, journey frequencies and daily travel durations, the measures broadly agreed. This suggests they were measuring the same thing; travel behaviour; and is an indicator of face validity for each measure. However, the travel diary data gave greater travel durations, and slightly greater journey frequencies, which combined to give over 12 minutes per day more travel per participant than the SenseCam (almost 6 minutes of which were active travel). Whether this finding is significant and has epidemiological or behavioural relevance will be discussed after deeper investigation in the next section.

Both measures missed similar numbers of eligible journeys the other recorded. As a percentage of recorded journeys, missed journeys were low. This gives further confidence that both SenseCam and the travel diary are acceptable ways to measure the travel behaviour of a study population of this size. The reasons for travel diary misses were participant failure to report. The majority of SenseCam misses were from failure to wear, or failure to charge. This suggests that when operated correctly SenseCam is a good tool to record journeys; the challenge is to elicit correct operation from participants in free-living. Future studies with such devices should look to further refine protocol and instruction, or reduce reliance on participant action, to reduce data lost for these reasons.

Implications

These findings suggest SenseCam is a feasible measure of travel behaviour in a sample of 80-90 participants. It may therefore be suitable for other similar validity studies for other measures and other health behaviours (e.g. sedentary behaviour). Data from the SenseCam and the travel diary look similar, but the choice of one or other measure in a population would likely lead to different results at a population level – whether this difference is acceptable or epidemiologically relevant depends on the research question being investigated.

Principal finding 2

There is a significant fixed bias of 2:08 minutes on reported journey durations. This represents over-reporting of journey duration and systematic error on self-reported durations. The wide limits of agreement (-9:10 to 13:26 minutes) demonstrate the large range (both positive and negative) for duration mis-reporting at an individual journey level. This represents large random error. There is very high agreement between measures for mode (98.8% of journeys).

This finding of consistent positive bias confirms Pilot 1 [189] and the GPS review presented in Chapter 2. The magnitude of this bias was found to be 2:08 minutes with a narrow 95% CI of 1:48 to 2:28 minutes. This is slightly lower than my Pilot 1 and half the magnitude of the aggregated result from my systematic review of GPS results. In particular it suggests the NTS is over-reporting by far less than the 2011 GPS pilot found [99] and less than the comparison to Map Point software conducted by Cronberg and Bonsall [121]. As a percentage of mean SenseCam journey duration this bias is 13.7% which is also less than the aggregate percentage bias calculated in the GPS review. This found 28.6% over all 12 studies, and 46.2% and 20.3% for unmatched and matched analysis (see Chapter 2).

Interestingly the matching of journey pairs did not influence the bias from the crude data. Comparison of crude diary and SenseCam data found a bias of 2:07 minutes per journey (see Table 8.4) while comparison of matched pairs found a bias of 2:08 minutes. This in contrast to the findings of my GPS review which found a larger differences by analysis type. It may suggest that the differences in GPS and reported durations by analysis type are more to do with GPS detection of journeys than participant interpretation or error.

Due to the criterion validity and reliability of my SenseCam method demonstrated in Chapter 4, I believe this is the best estimate yet of error or bias on self-reported journey durations in this type of self-report method. I will not propose any possible explanations for this until I have examined the potential influencing variables in RQ3. However, it should be noted this experiment is subject to some bias (particularly sample bias and possibly some reactivity). Assessment of the data for normality, heteroscedasticity and within- and between- subjects effects suggest that the parametric tests used, without data transformation, were appropriate.

The agreement on mode was exceptionally high (98.8%). This is higher than reported in GPS studies [61, 146]. This high agreement may be a result of only assessing matched trips. The 29 phantom journeys could be considered as mis-reported mode. Without a comparison measure (such as SenseCam) they would wrongly be included in analysis of self-report data. However, even when included there would still be agreement on over 95% of journeys. Therefore, these data suggest self-report of mode is a valid measure and researchers can have confidence in using self-reported modes.

Implications

The crude data from both measures suggest on average participants report 4.8 journeys per day (see Table 8.6). Therefore, using the bias found from criterion validity of journey time (2:08 minutes), on average the travel diary is overestimating travel time from these journeys by approximately 10 minutes per day. Considering the 2.8 active journeys per day, the travel diary is overestimating daily active travel by 6 minutes per day.

The next question is whether this is an acceptable bias (or measurement error). Assuming seven days of travel this is an overestimate of 42 minutes per week of active travel. Woodcock et al., [12] recently modelled health and environmental gains of weekly increases in walking and cycling of 51, 91 and 185 minutes, (finding large DALY gains and mortality reduction) which suggests 42 minutes is epidemiologically relevant. From these data and results I would therefore recommend adjusting or correcting population level self-reported travel data by 2:08 minutes per journey. In the next section I will investigate if any differential adjustments (e.g. by active or sedentary, gender, age, etc.) are warranted. In the final section I will examine another way of assessing daily travel time.

The wide limits-of-agreement (-9:10 to 13:26 minutes) as shown in Table 8.10 and Figure 8.4 show that for 95% of the journeys, the error on reported duration can be between a 9:10 minute under-report and a 13:26 minute over-report. This means that while the travel diary is a relatively good measure of population level journey duration, it is a particularly poor measure of individual journey duration. In other words, once enough measurements have been taken in a population, the overall average will be valid. However, for any given individual, who will have provided a much smaller number of measurements, the assessment of travel exposure could be substantially invalid.

This has implications for public health research. Self-report data is often used for categorising individuals by active travel status, detecting behaviour change in individuals or assessing eligibility for intervention which may be invalid if these errors are common to all self-report measures. This agrees with recommendations made by Dollman et al., for monitoring physical activity in adolescents which state self-report may be more suitable to population monitoring, descriptive and needs assessment based research than for intervention and individual based research [216]. Matthews et al., state 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 [53]. This suggests the large random error (95% limits of agreement) do not preclude self-reported travel from such studies. However, the authors go on to say that this random error is detrimental to the internal validity of association studies and in particular dose response studies. Large error at the individual level "lead to misclassification of exposure, loss of statistical power and attenuation of effect sizes" [53].

The results showing very good agreement for mode suggest self-report is a valid test and therefore may be suitable for detecting outcomes such as modal shift. However, it should be remembered that these results do not reveal anything about frequency. For example, if a travel diary reports three walks and two car journeys, we can be confident these modes are correctly reported, but we do not know what may have been left off (or erroneously included in) the diary.

Principal finding 3

The following journey characteristics were investigated:

Appeared to have a significant effect on bias	Appeared to have no significant effect on bias or no evidence for effect
Journey duration	Journey mode (car vs. cycle significant)
Rounding to five minutes	Journey purpose (visiting friends significantly different to some purposes)
Active or sedentary mode	Whether or not journey part of a chain

The following participant characteristics were investigated:

Appeared to have a significant effect on bias	Appeared to have no significant effect on bias or no evidence for effect
Whether jogger used	Gender
Educational status	Age category

The following study characteristics were investigated:

Appeared to have a significant effect on bias	Appeared to have no significant effect on bias or no evidence for effect
N/a	Study location
	Day of study
	Weekend or weekday

Investigation of the journey characteristics showed that journey length, reporting or rounding to five minutes and whether the journey was active or sedentary appeared to have a significant effect on the magnitude of the bias. These are discussed below.

An important aspect of the finding for rounding or digit preference is that it appears to be up (rather than down) increasing the systematic bias. It has previously been suggested that digit preference may skew results in this way [217]. It is worth noting non-rounded journeys are also over-reported, though to a lesser extent. As suggested from my previous work, I believe rounding-up happens because participants include other journey related activities in their reported time [154]. For example; putting on outdoor clothing before leaving the house or packing the car before setting off on a drive can make a 7.04 minute journey seem "about 10 minutes". These have been termed transitory activities in the travel literature [95]. I believe this to be an important explanation for the positive bias on reported journey duration.

It is unclear why active and sedentary journeys display a difference in the magnitude of over-reporting. For example, there is no obvious reason why the "transitory activities" mentioned above would be more important for active journeys. I speculate that participant perception of active and sedentary activities is different, but further research is warranted. The difference by journey length is likely to be due to some proportional bias, particularly important for 0-15 minute journeys, as discussed.

I did not find any evidence that independent journey mode, journey purpose and being part of a chain of journeys (stages) significantly influenced the bias. This is in contrast to some studies using GPS methods [61, 89, 196, 197] and also my own pilot work [154]. However, in my results there were some non-significant differences, so it may be a question of adequate power (as my study was designed for RQ2 at the whole group level).

Investigation of participant characteristics showed that use of the pocket jogger and educational status had a significant effect on the bias. The finding for jogger use is logical, but educational status revealed no clear trend and will require further investigation in future study.

Participant gender and age category had no significant effect. The age finding is in contrast to previous GPS work [89, 103, 124]. However, this could either be a power issue, or a result of not having many older adults in my sample and none beyond retirement age. This could be due to the recruitment setting or the requirement to use an electronic device (SenseCam). However, some previous studies have successfully used SenseCam in older populations [137].

Investigation of study characteristics showed study locations, day of study and weekend or weekday had no significant effect on bias. This suggests that these study variables at least did not bias the study and indicates good experimental validity. The non-significant difference between Auckland and the other locations could be due to the educational status of participants in that location, rather than the travel environment. It is also interesting that day of study did not influence bias, as reporting fatigue has been identified in a previous study [88]. It is possible wearing a SenseCam reminded participants they were in a travel study and increased diary compliance, or the sample I used were more familiar with and compliant to diary methods. Again, this is a potential source of experimental bias.

Implications

If researchers choose to adjust travel exposure data, differential adjustment by certain variables may be appropriate according to research question. In terms of public health it may be of particular interest that active journeys display a greater over-reporting bias than sedentary journeys.

Proportional or percentage adjustment could be considered. It would seem that journeys under five minutes (and to a lesser extent under 10) do not require as much adjustment as those between 10-120 minutes. Researchers who chose to adjust their data may decide to use the mean differences shown by journey length category, or the overall percentage bias of 13.7%, if appropriate to their research question. Researchers may also choose to adjust self-reported journey data differentially for journeys reported to the nearest five minute and for active and sedentary journeys. Using the bias of 2:38 minutes on active journeys, estimated weekly bias for active travel increases from 42 minutes (estimated in previous section) to 52 minutes.

Jogger use and educational status were also found to influence bias and could be considered if adjusting data. Researchers should find ways to increase jogger use in such studies, perhaps with clearer instructions or better materials. This also suggests techniques using real time reporting (e.g. smart phone applications) might be able to collect more valid and reliable data.

Principal finding 4

These data show that SenseCam performed as a criterion measure of summary journey behaviour (as per my definition) on two-thirds of the days participants were asked to wear it. Selecting these criterion SenseCam days increased apparent journey frequency, mean journey duration and daily summary duration.

Selecting the corresponding days of diary data reduced apparent journey frequency, increased journey duration and decreased overall daily summary duration in the travel diary.

Comparing criterion SenseCam data and corresponding diary days of data suggests journey durations are over-reported, journey frequencies are under-reported and overall daily summary durations agree. Including journeys of 1-3 minutes detected by SenseCam did not affect this result.

There was a very large 95% limits-of-agreement on the difference between daily summary durations from criterion SenseCam days and corresponding diary days.

On criterion SenseCam and corresponding travel diary days there was no meaningful difference between daily summary travel duration (4 seconds). This finding has raised three immediate possibilities. (1) This is a valid finding and is a result of considering journey frequency as well as duration. That is, participants over-report duration of individual journeys, but in effect cancel this out by missing journeys (at a rate of 0.67 journeys per day; see Table 8.25). A previous GPS study suggested forgotten trips may be a greater contributor to self-report error than "inaccuracy" [61]. Certainly, this finding is in contrast to my estimated summary over-report based on bias at an individual journey level (see RQ2); (2) Respondents who were more compliant with SenseCam and therefore provided more criterion days, were also the respondents more likely to be "good" diary completers; or (3) It may be that by filtering travel diary days by criterion SenseCam performance I introduced an artificial bias that has brought the travel diary daily summary down from 84:55 minutes per day to 76:09 minutes per day. The 29 phantom journeys should also be considered here. They made up 2.1% of the total diary journeys recorded and would have inflated the summary estimates from the diary but may have been filtered out by my SenseCam based criteria.

It is also important to note that if this had been a study of travel behaviour without the presence of SenseCam, the summary estimate based on 202 days of diary data would not exist as there would be no way of selecting these days. A researcher would have concluded summary travel behaviour based on the crude data. It may therefore be most relevant to compare criterion SenseCam data to crude travel diary data.

The very wide limits-of-agreement (~100 minutes) are more consistent with the findings from RQ2 that self-report of journey exposure has a large random error and is unsuitable for individual participant exposure assessment. The summary at the group level assessed by the diary appears to be valid by comparison to SenseCam. However, the uncertainty on an individual's exposure is very large.

Implications

I am not sure how to interpret these findings beyond the questions they have raised. It has certainly alerted me to the importance and effect of filtering data by compliance based criteria. This would also be important in accelerometer research with the analogy to wear-time criteria. Further, it suggests that considering two dimensions of a health behaviour (in this case duration and bout frequency) increases the complexity of the research, reduces the available data and can change the assumptions we might draw about exposure if we looked at single dimensions. Additionally, it calls into question the validity of only looking at duration and not journey frequency in my GPS systematic review.

Referring to Case-control studies, Altman and Bland said it is unclear how to analyse data from unmatched individuals and no really satisfactory solution exists for missing data [203]. Omitting people without complete data is known as complete case analysis and is most common, but results may be biased unless data missing completely at random [203]. This is analogous to my study having some days that were not complete. These days could not be included for summary analysis but may have influenced my results. Imputation is another option in Case-control studies, but would have been inappropriate for my comparison of measures experiment.

Finally, the very large random error is a finding consistent with RQ2. The implications in terms of population and individual measurement are the same.

Strengths and limitations of this experiment

This experiment has a number of strengths and limitations. This is the first study of this size to use SenseCam to validate a travel diary. There have only been two previous studies of this nature (Pilot 1 and Pilot 2 [154, 157]). The primary strength of this experiment is the use of a valid, reliable and objective measure (SenseCam) to compare to the travel diary data. Previous studies have used accelerometer or GPS but I believe SenseCam gives a more valid measure of travel mode, frequency and duration.

I demonstrated the repeated criterion validity of SenseCam compared to direct observation in Chapter 4 and have assumed it is applicable to the free-living in this experiment. I have assumed the reliability demonstrated between devices (inter-instrument) and between coding (intra-coder and inter-coder) applies to this experiment. However, as discussed in that chapter, the Criterion experiment had some limitations that may challenge these assumptions. Sensitivity to rule changes in the coding protocol are also untested.

A strength of the coding protocol is its ability to assess time spent travelling and exclude the previously discussed transitory activities. It may be that in the respondents' perceptions the journeys start before or during the transitory activities and they are reporting a valid assessment of the journey duration as they understand it. However, researchers wish to know the exposure to the activity itself, and using SenseCam and my protocol appears to allow this.

I believe the analysis approaches presented are a strength of this experiment. Many validation studies present only correlation coefficients, but by using paired-sample t-tests and Bland-Altman approaches I am able to investigate systematic and random bias separately [161, 218]. A correlation between the measured values measures the strength of the relation between them and will not be appropriate for assessing agreement or bias as two measures can be highly related without being very agreeable [161]. Perfect correlation can occur along any straight line, while perfect agreement would be a straight line on the line of equality and a test of significance is irrelevant to the question of agreement or bias [161]. Hopkins et al., challenge the Bland-Altman approach, especially when units of measurement differ, suggesting linear regression is superior. However they acknowledge that Bland-Altman may be suitable for method comparison when the units are the same, as is the case in my experiment [172].

The sample for this experiment was relatively young, highly educated and completed a large number of cycle journeys compared to real NTS data, and this may have biased my results. However, there was an even gender split. There is also a question as to whether the sort of people who volunteer to wear SenseCam are representative (and have representative reporting behaviour) of wider populations. The important point here is that the population they need to be representative of is the population of people who would take part in a travel survey, so this may introduce less bias than it initially appears. It should also 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 is not large. There were low numbers of public transport journeys due to my lower recruitment followed by the lower compliance in Romford where I anticipated high bus and train use. The bus journeys missed by SenseCam may also have under-represented this mode and biased the results.

Offering the $50 incentive in San Diego and not in the other locations was in line with institutional policy, but may have biased the recruitment and subsequent results. It is worth noting the NTS incentivises with £5 gift voucher, and consistency in future studies across locations may reduce recruitment and compliance bias. Previous monitoring studies (in high school students) have found monetary compensation the most effective for enhancing compliance [204].

The educational status variable was poorly categorised due to the mixed data I collected in the different sites and the coding of purpose variable was also confused by multipurpose journeys. I recently discovered the NTS codes “purpose from” and “purpose to” to alleviate this [120], and I would do this in future. All participants in my study are able bodied and therefore I do not have information on feasibility of use with disabled individuals. In Chapter 1 I reported the NTS classifies “readers and skimmers”, however I did not investigate this as a possible variable influencing bias. Neither did I collect any data on ethnicity of participants.

There are certain limitations of SenseCam as a device for measurement; (1) it may place burden upon the participants. There are particular settings where participants are not comfortable to wear the device and they have to remember to wear and charge it for a three day period[29]; (2) in certain situations such as very low light the images do not always show the journey clearly enough for identification. Images can also be lost when the lens is obscured by clothing or when participants forget to put it on; (3) it may introduce reactivity to the behaviour of interest; and (4) the approximate 10 second epoch between image capture introduces a small error on the proposed criterion assessment of journey duration.

Unlike most accelerometer based studies, I do not have wear-time criteria to identify valid criterion days. These are identified by travel diary journeys missed by SenseCam, or by incomplete SenseCam journeys. A participant could remove the device at work or at home for substantial periods (as directed by the ethical framework) and conduct journeys but these days could still be criterion days according to the above criteria. Triangulation against, for example GPS could be recommended for future analysis.

I based my self-report measure on the NTS. I have presented some public health uses of NTS data in Chapter 1. However, the NTS does not have the option of data linkage to health outcomes, as for example the Health Survey for England does, and future work with a measure more widely used in public health may be more useful. Further, I have primarily discussed the validity of journey duration estimates over three days using the NTS. However, as noted by Ogilvie and Panter, outcome measures for walking and cycling interventions can also include changes in usual or main mode or changes in mode share [219].

Finally, I accept there may have been some personal researcher bias. While I always aimed to achieve a detached evaluative stance, the way I constructed my protocol and procedures could have biased the

[29] Participant burden was assessed at the return study visit through a satisfaction survey. These data are not presented in this thesis but initial inspection suggests the burden is largely perceived; participants anticipate feeling awkward to drawing attention by wearing the device at work or in public though this rarely happens in practice. This agrees with the findings from both pilot studies used to inform the ethical framework.

outcomes. Repeated studies by other researchers are needed before the findings can be considered solid evidence.

Future study

I have made some recommendations for future research, specifically raised by this main experiment;

- Repeat the experiment with a more representative sample to investigate the feasibility of using in wider population groups, and the bias on reporting in these groups[30]. This might include low- and middle- income countries (LMICs)[31], though the cost of the device and resources available may prohibit this. There are certain LMICs where active travel is a particularly important intervention e.g., Ciclovia in Bogota, Columbia and measurement could be particularly instructive [220, 221];
- Repeat the study with other self-report measures of travel and active travel commonly used in public health research, for example the International Physical Activity Questionnaire (IPAQ) [222] or the EPIC Physical Activity Questionnaire (EPAQ-2) [223];
- Repeat with a greater focus on detecting journey frequency to better assess the daily summary question. Perhaps incorporating more advanced analysis techniques such as receiver operator characteristic (ROC) curves which has been used as part of a range of measures of agreement in pedometer studies [224];
- 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;
- Develop explanatory models that go beyond my descriptive variables for travel duration over-reporting. This could allow comparison of the relative importance of different variables. In other health behaviours, specifically food intake, researchers have developed regression calibration methods [225].

[30] I have had meetings with both Sustrans and NatCen about the possibility of such work. I plan to resume these discussions after completing my DPhil studies.

[31] I assisted an MSc Student who recently piloted SenseCam measurement of travel in Kingston Jamaica

Chapter conclusion

The main experiment presented in this chapter showed that SenseCam is a feasible tool for travel behaviour exposure assessment in a sample of 80-90 participants. It also showed that SenseCam and the NTS-based travel diary broadly agree on travel behaviour assessment.

Assessment of agreement, taking SenseCam to be a criterion measure, showed travel diaries over-report journey duration by 2:08 minutes (95% CI = 1:48 to 2:28) at a group level. Investigation of journey, participant, and study characteristics suggests certain variables that may influence the magnitude of this over-reporting. Specific variables that appear to influence over-reporting include journey duration, rounding to nearest five minutes, whether it is active or sedentary, use of a pocket memory jogger and educational status. This over-reporting may lead to epidemiologically relevant errors on assessment of exposure to active or sedentary travel over the course of a day, week, month or year. The agreement for mode was very high (Kappa = 0.983).

The 95% limits-of-agreement were wide (-9:10 to 13:26 minutes). This raises questions about the suitability of self-report methods to assess individual participant travel behaviour.

Inspection of daily summary data raised questions about considering journey frequency and the effect of filtering data by certain criteria.

Part 4 – Thesis discussion

CHAPTER 9 – THESIS DISCUSSION

Summary

In this final chapter I reflect on the contributions made by my thesis. I restate my thesis objectives and summarise the main findings and discussion points, including strengths and weaknesses. I then suggest the implications of these findings for policy, practice and research. This is followed by my suggestions for future research based on my thesis.

I briefly outline the on-going SenseCam work I am involved in and my plans for two papers based on this thesis. I also list my published papers from this work.

Finally, I place SenseCam in my table of travel measures and reflect on the strengths and weaknesses of this method. I conclude that I achieved my thesis aim of assessing the efficacy of SenseCam to help understand the error in a widely used travel diary.

Discussion of thesis

In Chapter 1, I stated my thesis aim as:

"…to assess the efficacy of utilising SenseCam to improve our understanding of the error on self-reported travel duration using a modified version of the National Travel Survey"

To achieve this aim, I formulated three thesis objectives:

1. Systematically review the existing literature comparing GPS measured and self-reported journey durations
2. Develop a new method for travel measurement with SenseCam, with a view to valid and reliable journey assessment
3. Test my developed SenseCam method in a suitably powered sample, and investigate the error on self-reported journey durations (and compare to results from GPS review)

This final chapter is a discussion and reflection on the extent to which I achieved these aims and objectives.

Objective 1

Having discussed the importance of travel and active travel as public health behaviours I identified self-report as the pervading technique, despite the emergence of objective technologies. To date, there have been a number of studies using one of these technologies (GPS) to assess the error in self-reported behaviour of travel. I conducted the first (to my knowledge) systematic review of studies that compared GPS measured journey durations to self-reported journey durations.

The review demonstrates there is a disagreement between self-reported and GPS measured journey time; it is consistent in direction (over-reporting), but variable in magnitude which means that the two measures should not be used interchangeably. I found an aggregated over-report of 4.4 mins (28.6%) per journey. It suggests (based on the assumption GPS is the more objective measure), studies using self-reported journey duration over multiple journeys and days are likely to be over-estimating the travel time or exposure to active and sedentary travel.

The review had certain limitations common to many systematic reviews (few studies, heterogeneous study characteristics, etc.). I also realised through the course of this thesis I should have assessed journey frequency as well as journey duration. If I had the opportunity to repeat this review, I would assess this dimension of travel behaviour as well.

This review was set up to assess the best available evidence for the error or bias on self-report. This led into my second thesis objective; to try and develop a better method with wearable cameras.

Objective 2

In field testing prior to my thesis, I had identified that SenseCam (a wearable camera) had the potential to be a near objective measure of travel behaviour (mode, duration and frequency). My second thesis objective was to develop a feasible, valid and reliable method to assess travel behaviour with this device. Through two pilot studies I developed and refined this method. Through a series of tests according to a methodological framework developed from the measurement literature I established the high validity and reliability of this method. I also developed an ethical framework for SenseCam research in health behaviour measurement research.

There were certain limitations to this approach. In particular, my testing of criterion validity was subject to a small number of simulated journeys, of limited modes and durations. I did not investigate public transport, or the performance of SenseCam and my methods in low light levels. However, the method performed very well on the tests I did conduct. This set up my third and final thesis objective; to test this method in a full validity experiment.

Objective 3

To complete my thesis aim, I designed an experiment to assess the error on the NTS using my new method. The NTS collects data on travel behaviour in the UK and the data (particularly on active travel) is often used in public health. In four locations, I collected data (SenseCam and self-report) from 84 participants over 3-4 days.

I compared the data from both measures through four research questions. Across 1,127 matched journeys, I found a positive bias of 2:08 minutes per journey in the self-report (taking SenseCam as the criterion measure). Using this bias and measured journey frequencies I estimated a daily over-estimation of active travel behaviour of +6 minutes (or 42 minutes per week). I found this bias to differ by certain variables and presented descriptive data for these. As active journeys appear to be subject to greater bias than sedentary journeys, this estimate of over-reported active travel could be considered conservative.

Next, I selected criterion days to allow assessment of journey frequency and calculation of daily summary travel duration. I found there was no bias when comparing matched days of data. Whether this was because under-reporting of frequency cancelled over-reporting of duration, or because filtering the data introduced experimental bias is unclear. However, analysis at both journey level, and daily summary level, revealed large random error (wide limits of agreement).

This final experiment has certain limitations. In particular, the study locations, while providing a range of travel environments and travel behaviours, were chosen opportunistically rather than with sound scientific reasoning. Further, there is uncertainty over how representative my sample was of wider populations using travel diaries. However, this study is the first of its size using automated cameras in health behaviour measurement. The proven feasibility of this method offers opportunities to assess other self-report measures and other health related behaviours.

Implications for policy, practice and research

I believe there are certain implications from this body of work. Primarily, this thesis suggests that anyone using self-reported journey durations should consider them an over-estimation of true travel time. I have suggested correction factors researchers using such data may wish to consider. Unfortunately it is not clear if the same is true for daily summary durations; this should be assessed in future research. My work is based on a widely used measure of travel behaviour already used by policy makers, practitioners and researchers; the findings presented in this thesis should help them to understand the confidence and potential bias in their own calculations.

Perhaps of greater impact is the finding that, at journey and daily summary level, the random error on this type of self-report is very high. This is of significance to any researcher or practitioner using self-reported travel data to calculate dose-response relationships or detect changes in behaviour and has been discussed in previous literature [216]. Particularly considering behaviour change, the size of the limits of agreement suggests the random error is in excess of plausible changes in active travel behaviour; any behaviour change will be undetectable in the noise and effective interventions could be rejected. I would recommend the use of objective measures (such as GPS or accelerometer) in such studies. Future research should assess this in other commonly used self-report travel measures such as the walking questions in IPAQ or EPAQ-2 [222, 223].

I was interested by the effect of simultaneously introducing a second domain (frequency) and filtering data to obtain criterion days. Roth and Mindell have previously reported the importance of assessing who provides criterion data in relation to accelerometer based research [226]. Loprinzi et al., refer to excluding participants with invalid accelerometer data as introducing “cut-off bias” [227]. I think the analysis shown in Research question 4 demonstrates the effects of this and is of interest to researchers who conduct similar investigations with other devices and measures.

Finally, the demonstration of a feasible method to assess health behaviours with an associated ethical framework should be of interest to any researcher looking to improve the way we can assess health behaviours. Beyond active travel, physical activity, sedentary behaviour, diet, alcohol intake (and to a lesser extent smoking) have their own issues with measurement [163]. While exact methods may differ, perhaps focussing on prompted recall or validation of other devices, the possibility to investigate and improve measurement has been shown. As previously reported, the first studies in other domains are already being published.

Recommendations for future research

To summarise my recommendations for future research, based on this thesis I suggest the following:

- Repeat my main study with a focus on a more representative sample, and equal attention given to assessment of frequency;

- Repeat this study with other common self-report measures of travel;
- Use my method to investigate other travel measures such as GPS or accelerometer;
- Assess the utility of SenseCam in the measurement of other important health behaviours;
- Assess the utility of SenseCam images as a prompt or recall tool. For example, in qualitative interviews about behavioural choice, or in formative evaluation of participant experience of an intervention;
- I believe there may be a role for SenseCam in intervention research. Firstly, the "better" measurement provided by SenseCam could give greater sensitivity to behaviour change. Perhaps more importantly, the images could become part of the intervention; personalised or reflective feedback from a nutritionist or physical activity councillor using the images of the participant's actual behaviour may prove more effective than general advice. The same may be true for smoking and alcohol behaviours.

Publications from my thesis

During the course of my thesis I have published the following studies (as first author) from my work;

1. **P Kelly**, M Murphy, P Oja, E Murtagh and C Foster; Estimates of the number of people in England who attain or exceed vigorous intensity exercise by walking at 3mph; Journal of Sports Sciences; 2011 29(15):1629;

2. **P Kelly**, A Doherty, E Berry, S Hodges, AM Batterham and C Foster; Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study; International Journal of Behavioural Nutrition and Physical Activity; 2011 8(1):44;

3. **P Kelly,** A Doherty, A Hamilton, AM Batterham, A Matthews, C Foster and G Cowburn; Evaluating the Feasibility of Measuring Travel to School Using a Wearable Camera; American Journal of Preventive Medicine; 2012, (43): 546-550;

4. **P Kelly**, S Marshall, A Doherty, J Kerr, H Badland, M Oliver and C Foster; An ethical framework for the use of automated digital image capture in health behaviour research; American Journal of Preventive Medicine; 2013.

I highlight number 2 as my biggest contribution to the literature. As of 1st May 2013 it has been accessed 2928 times (IJBNPA online statistics) and cited 16 times (Web of science). Other published SenseCam studies I have contributed to as co-author are listed in Appendix 12. These include three publications from the accelerometer and GPS data collected at the same time as my thesis data, as described in Chapter 8; Doherty et al., [228] Kerr et al., [140] and Oliver et al., [228].

Publication plan

From my thesis I plan to try and publish the following papers;

- The systematic review of GPS and self-report comparisons described in Chapter 2. A manuscript for this paper is currently under consideration at Transport Reviews;
- A paper reporting the criterion validity and reliability of a modified NTS diary using my main data set. This paper is not yet written. It will report the small systematic bias and large random bias as principal findings. I am considering IJBNPA where my first SenseCam pilot was published as a target journal.

On-going SenseCam work

As a result of my published ethical framework for observational research I have been asked to produce a second framework for the use of automated wearable cameras in eliciting behaviour change. This project is run by the London School for Economics and funded by the Economic and Social Research Council.

Dr Aiden Doherty and I have also been invited to run a pre-conference research workshop at the ISBNPA annual meeting in Ghent in May 2013. We will be delivering a practical session to 18 researchers and inviting them to consider how images from automated wearable cameras might contribute to their own research. I have also been invited to the Programme Committee for the SenseCam 2013 annual conference in San Diego.

I have had a series of meetings with Professor J Gershuny (Nuffield College, University of Oxford). His research is focussed on time-use diaries, behavioural trends, and newly, health outcomes. We are proposing to use my SenseCam method to investigate the bias or otherwise on the self-reported behaviour he uses in his own work. We plan to submit a funding proposal over the summer.

SenseCam in travel research

SenseCam provides near objective assessment of travel behaviour. To revisit Table 1.3 from the end of Chapter 1, I have added SenseCam to my own summary of the capabilities of this method. This is displayed in Table 8.1.

Limitations of SenseCam as a method

Beyond the strengths and limitations discussed with regard to the main experiment, SenseCam as a method has some general limitations. It is for these reasons I believe SenseCam (in its current form) is unlikely to replace many established techniques. SenseCam provides no information on location or intensity of activity, and is therefore only relevant to certain research questions. It would have to be combined with GPS, accelerometer or heart rate monitoring to be more widely relevant. While the feasibility of this has been demonstrated it increases the complexity of the data collection and the participant burden. Despite calling SenseCam an objective measure, the images still require subjective interpretation, and data can be lost for a variety of reasons.

My thesis, and other work, has shown some people are willing to wear a camera all day long and provide images of their lives. However, not everyone accepts this invasion of privacy, and it is likely that SenseCam will only be feasible in certain groups and sections of society. Finally, using SenseCam is resource intensive. The devices cost £300 each as of April 2013 and the images take significant researcher time to code and analyse. I think the primary role is in supplementing and validating measures that can be used on wider scales.

Table 8.1 My summary of the capabilities of common travel measurement techniques with SenseCam included

	Measure						
Dimension of travel behaviour	**Self-report**	**GPS**	**Accelerometer**	**Pedometer**	**Heart rate**	**Direct observation**	**SenseCam**
Mode	Subject to participant memory	Subject to detection algorithm	Subject to behavioural detection	Only walking	No	Yes	Near direct observation
Frequency	Subject to participant memory	Subject to detection algorithm	Subject to behavioural detection	No	No	Yes	Near direct observation subject to coding protocol and criterion wear-time
Duration	Subject to participant memory	Subject to detection algorithm	Subject to behavioural detection	Step counts per day	No	Yes	Near direct observation subject to coding protocol
Intensity	Subject to participant memory	Speed information subject to detection algorithm	Estimated energy expenditure subject to behavioural detection	No	Heart rate response	Some researcher estimation	No
Context or purpose (work, leisure)	Subject to participant memory	Some inference from land use	No	No	No	Yes	Some information. Would perform better as recall tool.
Travel Environment	Subject to participant memory	If combined with GPS	No	No	No	Yes	Some information; benefits are temporal data such as pedestrian levels or obstructions

Personal reflections

I have developed (or at least been shown the importance of) some critical research skills during this thesis. I attempt to summarise below;

- I have conducted my first systematic literature review. I learned the importance of a focussed RQ and defined eligibility criteria (especially after an aborted first review into technologies in activity assessment);
- I had not realised the differences within, and the importance of, the terms validity and reliability. I attempted to formulate my own framework around these concepts for establishing a new measure in health behaviour assessment;
- I have had my first attempt at taking a new measure from the idea stage, through testing and development to demonstrating efficacy in real people;
- I have learned about ethical research and existing ethical frameworks;
- I have run one pilot study, attached my measures to another, and designed and led an 84 participant data collection. This has given me experience of recruitment, compliance, data collection, coding and analysis;
- I have seen how adding more dimensions to a health exposure and filtering to reveal criterion data can increase complexity and change the outcomes and results from analysis.

Final remarks

The results presented in this thesis support previous claims for the suitability of self-report diaries for surveillance and monitoring of population trends. I believe self-report will continue in this function for many years. However, these results add evidence to the unsuitability of self-report diaries to detect behaviour change or exposure levels of active travel in individuals. Devices and techniques will continue to improve and I believe all trials and interventions will be utilising devices in the next 10-15 years. I hope the price makes this feasible in LMICs.

In this thesis I have stated my opinion that SenseCam is the closest measure to direct observation available to travel researchers. I have also presented some supporting evidence for this. However, SenseCam is not the pinnacle of device measurement and will be surpassed in the very near future. Better, smaller devices and multi-device platforms (including images, location, accelerometry and physiological response) will provide greater resolution to health assessment very soon. Ultimately, we may one day live in a world of continuous, ubiquitous health monitoring of our entire lives through a micro-chip in the body. It is unclear if this is a good thing.

Conclusions

I feel I have achieved my thesis aim, namely to assess the efficacy of a new methodology or device in the measurement of travel behaviour. I conclude it is a near objective measure of journey mode, frequency and duration, but is still subject to limitations, assumptions and subjectivity.

There is no single objective physical activity assessment instrument that is appropriate for all situations, populations, and research questions [229]. Selecting a physical activity monitoring system for a particular project depends on the objectives of the study and the resources available to purchase and use the instruments [53]. There are certain situations and studies when SenseCam may be the appropriate choice.

REFERENCES

1. World Health Organisation: **A guide for population based approaches to increasing levels of physical activity: Implementation of the WHO global strategy on diet, physical activity and health.** 2007.
2. Department of Health - Physical Activity Health Improvement and Prevention: **At least five a week: Evidence on the impact of physical activity and its relationship to health. A report from the Chief Medical Officer.** 2004.
3. Bull FC, Gauvin L, Bauman A, Shilton T, Kohl HW, 3rd, Salmon A: **The Toronto Charter for Physical Activity: a global call for action.** *J Phys Act Health* 2010, **7:**421-422.
4. World Health Organisation: **Global health risks: mortality and burden of disease attributable to selected major risks.**; 2009.
5. World Health Organisation: **Global recommendations on physical activity for health.**: World Health Organisation; 2010.
6. Institute for Health Metrics and Evaluation: **The Global Burden of Disease: Generating Evidence, Guiding Policy.** 2013.
7. National Institute for Health and Care Excellence: **Walking and cycling: local measures to promote walking and cycling as forms of travel or recreation [PH41].** 2012.
8. Ogilvie D, Foster CE, Rothnie H, Cavill N, Hamilton V, Fitzsimons CF, Mutrie N: **Interventions to promote walking: systematic review.** *BMJ* 2007, **334:**1204.
9. Smith G, Gidlow C, Davey R, Foster C: **What is my walking neighbourhood? A pilot study of English adults' definitions of their local walking neighbourhoods.** *Int J Behav Nutr Phys Act* 2010, **7:**34.
10. Ogilvie D, Bull F, Cooper A, Rutter H, Adams E, Brand C, Ghali K, Jones T, Mutrie N, Powell J, et al: **Evaluating the travel, physical activity and carbon impacts of a 'natural experiment' in the provision of new walking and cycling infrastructure: methods for the core module of the iConnect study.** *BMJ Open* 2012, **2**.
11. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, Macera CA, Heath GW, Thompson PD, Bauman A: **Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association.** *Med Sci Sports Exerc* 2007, **39:**1423-1434.
12. Woodcock J, Givoni M, Morgan AS: **Health impact modelling of active travel visions for england and wales using anIntegrated Transport and Health Impact Modelling Tool (ITHIM).** *PLoS One* 2013, **8:**e51462.

13. Oliver M, Badland H, Mavoa S, Duncan MJ, Duncan S: **Combining GPS, GIS, and accelerometry: methodological issues in the assessment of location and intensity of travel behaviors.** *J Phys Act Health* 2010, **7:**102-108.

14. Kate Fox, Leicha Rickards: **Sport and leisure: Results from the sport and leisure module of the 2002 General Household Survey.** Office for National Statistics; 2004.

15. Wilkinson R, Marmot M: **Social determinants of health – The solid facts.** *World Health Organization (Copenhagen Regional Office for Europe),* 2003.

16. Butland B, Jebb S, Kopelman P, McPherson K, Thomas S, Mardell J, Parry V: **Foresight. Tackling obesities: future choices. Project report.** *Foresight Tackling obesities: future choices Project report* 2007.

17. Hayashi T, Tsumura K, Suematsu C, Okada K, Fujii S, Endo G: **Walking to Work and the Risk for Hypertension in Men: The Osaka Health Survey.** *Annals of internal medicine* 1999, **131:**21-26.

18. Hu G: *Physical activity during commuting and recreation associated with cardiovascular risk factors in China, with reference to Finland.* Kuopio: Kuopion Yliopisto; 2001.

19. Hu G, Pekkarinen H, Hanninen O, Yu Z, Guo Z, Tian H: **Commuting, leisure-time physical activity, and cardiovascular risk factors in China.** *Med Sci Sports Exerc* 2002, **34:**234-238.

20. Lubans DR, Boreham CA, Kelly P, Foster CE: **The relationship between active travel to school and health-related fitness in children and adolescents: a systematic review.** *Int J Behav Nutr Phys Act* 2011, **8:**5.

21. Bricka S: *What Is Different About Non-Response in GPS-Aided Surveys?* Bingley: Emerald Group Publishing Limited; 2009.

22. Hamer M, Chida Y: **Walking and primary prevention: a meta-analysis of prospective cohort studies.** *Br J Sports Med* 2008, **42:**238-243.

23. Woodcock J, Franco OH, Orsini N, Roberts I: **Non-vigorous physical activity and all-cause mortality: systematic review and meta-analysis of cohort studies.** *Int J Epidemiol* 2011, **40:**121-138.

24. Hamer M, Steptoe A: **Walking, vigorous physical activity, and markers of hemostasis and inflammation in healthy men and women.** *Scand J Med Sci Sports* 2008, **18:**736-741.

25. American College of Sports Medicine: **Position Stand: The recommended quantity and quality of exercise for developing and maintaining cardiorespiratory and muscular fitness, and flexibility in healthy adults.** *Med Sci Sports Exerc* 1998, **30:**975-991.

26. Morris JN, Hardman AE: **Walking to health.** *Sports Medicine* 1997, **23:**306-332.

27. Lee IM, Buchner DM: **The importance of walking to public health.** *Med Sci Sports Exerc* 2008, **40:**S512-518.

28. Hillsdon M, Thorogood M: **A systematic review of physical activity promotion strategies.** *Br J Sports Med* 1996, **30:**84-89.

29. Hillsdon M, Foster C, Thorogood M: **Interventions for promoting physical activity.** *Cochrane Database Syst Rev* 2005**:**CD003180.

30. Stamatakis E, Ekelund U, Wareham NJ: **Temporal trends in physical activity in England: The Health Survey for England 1991 to 2004.** *Preventive Medicine* 2007, **45:**416-423.

31. Belanger M, Townsend N, Foster C: **Age-related differences in physical activity profiles of English adults.** *Prev Med* 2011, **52:**247-249.

32. Sports Council and Health Education Authority: **Allied Dunbar National Fitness Survey: main findings.** 1992.

33. Kelly P, Murphy M, Oja P, Murtaugh EM, Foster C: **Estimates of the number of people in England who attain or exceed vigorous intensity exercise by walking at 3 mph.** *J Sports Sci* 2011**:**1-6.

34. Murtagh EM, Boreham CA, Murphy MH: **Speed and exercise intensity of recreational walkers.** *Prev Med* 2002, **35:**397-400.

35. Yang L, Sahlqvist S, McMinn A, Griffin SJ, Ogilvie D: **Interventions to promote cycling: systematic review.** *BMJ* 2010, **341:**c5293.

36. Shephard R: **Is Active Commuting the Answer to Population Health?** *Sports Medicine* 2008, **38:**751-758.

37. Kahlmeier S, Racioppi F, Cavill N, Rutter H, Oja P: **"Health in all policies" in practice: guidance and tools to quantifying the health effects of cycling and walking.** *J Phys Act Health* 2010, **7 Suppl 1:**S120-125.

38. Oja P, Titze S, Bauman A, de Geus B, Krenn P, Reger-Nash B, Kohlberger T: **Health benefits of cycling: a systematic review.** *Scand J Med Sci Sports* 2011, **21:**496-509.

39. Andersen LB, Wedderkopp N, Kristensen P, Moller NC, Froberg K, Cooper AR: **Cycling to school and cardiovascular risk factors: a longitudinal study.** *J Phys Act Health* 2011, **8:**1025-1033.

40. **Evaluation of the Cycling City and Towns programme interim report** [https://www.gov.uk/government/publications/evaluation-of-the-cycling-city-and-towns-programme-interim-report] Retrieved 11/3/13.

41. de Hartog J, Boogaard H, Nijland H, Hoek G: **Do the health benefits of cycling outweigh the risks?** *Environ Health Perspect* 2010, **118:**1109-1116.

42. **Cyclists can help Britain's economy get back on its bike suggests research** [http://www2.lse.ac.uk/newsAndMedia/news/archives/2011/08/cycling.aspx] 2/1/13.

43. Ogilvie F, Goodman A: **Inequalities in usage of a public bicycle sharing scheme: socio-demographic predictors of uptake and usage of the London (UK) cycle hire scheme.** *Prev Med* 2012, **55:**40-45.

44. Goodman A, Mackett RL, Paskins J: **Activity compensation and activity synergy in British 8-13 year olds.** *Preventive Medicine* 2011, **53:**293-298.

45. Florez Pregonero A, Gomez LF, Parra DC, Cohen DD, Arango Paternina CM, Lobelo F: **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* 2012, **54:**402-404.

46. Parra DC, Lobelo F, Gomez LF, Rutt C, Schmid T, Brownson RC, Pratt M: **Household motor vehicle use and weight status among Colombian adults: are we driving our way towards obesity?** *Prev Med* 2009, **49:**179-183.

47. Fraser SD, Lock K: **Cycling for transport and public health: a systematic review of the effect of the environment on cycling.** *Eur J Public Health* 2010**:**738-743.

48. Litman T: **Transportation and public health.** *Annu Rev Public Health* 2013, **34:**217-233.

49. Litman T: **Integrating public health objectives in transportation decision-making.** *Am J Health Promot* 2003, **18:**103-108.

50. **Measuring effectiveness and cost effectiveness: the QALY** [http://www.nice.org.uk/newsroom/features/measuringeffectivenessandcosteffectivenesstheqaly.jsp] Retrieved 26/04/13.

51. Shephard RJ: **Limits to the measurement of habitual physical activity by questionnaires.** *Br J Sports Med* 2003, **37:**197-206;.

52. Sallis JF, Owen N, Fotheringham MJ: **Behavioral epidemiology: A systematic framework to classify phases of research on health promotion and disease prevention.** *Annals of Behavioral Medicine* 2000, **22:**294-298.

53. Matthews CE, Hagstromer M, Pober DM, Bowles HR: **Best practices for using physical activity monitors in population-based research.** *Med Sci Sports Exerc* 2012, **44:**S68-76.

54. Bauman A, Phongsavan P, Schoeppe S, Owen N: **Physical activity measurement- a primer for health promotion.** *Promotion & Education* 2006, **13:**92-103.

55. Wolf J, Guensler R, Bachman W: **Elimination of the travel diary - Experiment to derive trip purpose from global positioning system travel data.** *Transport Res Rec* 2001**:**125-134.

56. Janz KF: **Physical activity in epidemiology: moving from questionnaire to objective measurement.** *Br J Sports Med* 2006, **40:**191-192.

57. Staplin L, Gish KW, Joyce J: **'Low mileage bias' and related policy implications--a cautionary note.** *Accid Anal Prev* 2008, **40:**1249-1252.

58. Stopher P, Zhang Y: **The Repetitiveness of Daily Travel.** In *90th Annual Meeting Transportation Research Board; Washington.*; 2011

59. Kerr J, Duncan S, Schipperjin J: **Using Global Positioning Systems in Health Research: A Practical Approach to Data Collection and Processing.** *Am J Prev Med* 2011, **41:**532-540.

60. Department for Transport: **National Travel Survey 2009: Statistical release.** 2010.

61. Stopher P, FitzGerald C, Xu M: **Assessing the accuracy of the sydney household travel survey with GPS.** *Transportation* 2007, **34:**723-741.

62. Krizek KJ, Handy SL, Forsyth A: **Explaining changes in walking and bicycling behavior: challenges for transportation research.** *Environment and Planning B: Planning and Design* 2009, **36:**725-740.

63. Milton K, Kelly P, Bull F, Foster C: **A formative evaluation of a family-based walking intervention-Furness Families Walk4Life.** *BMC Public Health* 2011, **11:**614.

64. Stopher P, FitzGerald C, Zhang J: **Search for a global positioning system device to measure person travel.** *Transportation Research Part C: Emerging Technologies* 2008, **16:**350-369.

65. National Center for Social Research: **NTS Travel Record Review Stage 1.** Prepared for Department for Transport; 2006.

66. Merom D, van der Ploeg HP, Corpuz G, Bauman AE: **Public Health Perspectives on Household Travel Surveys: Active Travel Between 1997 and 2007.** *Am J Prev Med* 2010, **39:**113-121.

67. Baranowski T, Masse LC, Ragan B, Welk G: **How many days was that? We're still not sure, but we're asking the question better!** *Med Sci Sports Exerc* 2008, **40:**S544-549.

68. Sallis J, Owen N: *Physical activity & behavioral medicine.* Thousand Oaks, Calif.: Sage Publications; 1999.

69. McLeroy KR, Bibeau D, Steckler A, Glanz K: **An ecological perspective on health promotion programs.** *Health Educ Q* 1988, **15:**351-377.

70. Trost S: **Physical Acitvity Measurement - Lecture at Physical Activity and Public Health Course (13/9/11).** 2011.

71. Welk G: *Physical Activity Assessments for Health-Related Research.* Champaign, IL: Human Kinetics; 2002.

72. Grengs J, Wang XG, Kostyniuk L: **Using GPS data to understand driving behavior.** *J Urban Technol* 2008, **15:**33-53.

73. Armstrong N, Welsman JR: **The physical activity patterns of European youth with reference to methods of assessment.** *Sports Medicine* 2006, **36:**1067-1086.

74. Maddocks M, Petrou A, Skipper L, Wilcock A: **Validity of three accelerometers during treadmill walking and motor vehicle travel.** *Br J Sports Med* 2010, **44:**606-608.

75. Jones AP, Coombes EG, Griffin SJ, van Sluijs EM: **Environmental supportiveness for physical activity in English schoolchildren: a study using Global Positioning Systems.** *Int J Behav Nutr Phys Act* 2009, **6:**42.

76. Duncan JS, Badland HM, Schofield G: **Combining GPS with heart rate monitoring to measure physical activity in children: A feasibility study.** *J Sci Med Sport* 2009, **12:**583-585.

77. Jago R, Baranowski T, Baranowski JC: **Observed, GIS, and self-reported environmental features and adolescent physical activity.** *Am J Health Promot* 2006, **20:**422-428.

78. Kocur P, Deskur-Smielecka E, Wilk M, Dylewicz P: **Estimation of energy expenditure during various forms of exercise training in early cardiac rehabilitation. [Polish, English].** *Fizjoterapia* 2009, **17 (2):**3-11.

79. Duncan MJ, Badland HM, Mummery WK: **Applying GPS to enhance understanding of transport-related physical activity.** *J Sci Med Sport* 2009, **12:**549-556.

80. Troped PJ, Oliveira MS, Matthews CE, Cromley EK, Melly SJ, Craig BA: **Prediction of activity mode with global positioning system and accelerometer data.** *Med Sci Sports Exerc* 2008, **40:**972-978.

81. Mavoa S, Oliver M, Witten K, Badland HM: **Linking GPS and travel diary data using sequence alignment in a study of children's independent mobility.** *International Journal of Health Geographics* 2011, **10:**64.

82. Nakamya J, Moons E, Koelet S, Wets G: **Impact of data integration an some important travel behavior indicators.** *Transportation Research Record* 2007**:**89-94.

83. Hodges S, Williams L, Berry E, Izadi S, Srinivasan J, Butler A, Smyth G, Kapur N, Wood K: **SenseCam: A retrospective memory aid.** *Ubicomp 2006: Ubiquitous Computing, Proceedings* 2006, **4206:**177-193.

84. Rubin D: *Remembering Our Past: Studies in Autobiographical Memory.* Cambridge University Press; 1999.

85. Sallis JF, Saelens BE: **Assessment of physical activity by self-report: status, limitations, and future directions.** *Res Q Exerc Sport* 2000, **71:**S1-14.

86. Bricka SG, Sen S, Paleti R, Bhat CR: **An analysis of the factors influencing differences in survey-reported and GPS-recorded trips.** *Transportation Research Part C-Emerging Technologies* 2012, **21:**67-88.

87. Janz KF: **Physical activity in epidemiology: moving from questionnaire to objective measurement.** *British journal of sports medicine* 2006, **40:**191-192.

88. Murakami E, Wagner DP: **Can using global positioning system (GPS) improve trip reporting?** *Transportation Research Part C-Emerging Technologies* 1999, **7:**149-165.

89. Blanchard RA, Myers AM, Porter MM: **Correspondence between self-reported and objective measures of driving exposure and patterns in older drivers.** *Accid Anal Prev* 2010, **42:**523-529.

90. Sahlqvist S, Song Y, Bull F, Adams E, Preston J, Ogilvie D: **Effect of questionnaire length, personalisation and reminder type on response rate to a complex postal survey: randomised controlled trial.** *BMC Med Res Methodol* 2011, **11:**62.

91. Du J, Aultman-Hall L: **Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues.** *Transportation Research Part A: Policy and Practice* 2007, **41:**220-232.

92. Marshall SC, Wilson KG, Molnar FJ, Man-Son-Hing M, Stiell I, Porter MM: **Measurement of driving patterns of older adults using data logging devices with and without global positioning system capability.** *Traffic Injury Prevention* 2007, **8:**260-266.

93. Broughton P: **Handbook of transport geography and spatial systems.** *Tourism Manage* 2006, **27:**1086-1087.

94. Tourangeau R, Rips LJ, Rasinski KA: *The psychology of survey response.* Cambridge Universitty Press; 2000.

95. Rietveld P: **Rounding of Arrival and Departure Times in Travel Surveys: An Interpretation in Terms of Scheduled Activities.** Tinbergen Institute; 2001.

96. Duncan MJ, Mummery WK: **GIS or GPS? A comparison of two methods for assessing route taken during active transport.** *Am J Prev Med* 2007, **33:**51-53.

97. Spittaels H, Foster C, Oppert JM, Rutter H, Oja P, Sjostrom M, De Bourdeaudhuij I: **Assessment of environmental correlates of physical activity: development of a European questionnaire.** *Int J Behav Nutr Phys Act* 2009, **6:**39.

98. Spittaels H, Verloigne M, Gidlow C, Gloanec J, Titze S, Foster C, Oppert JM, Rutter H, Oja P, Sjostrom M, De Bourdeaudhuij I: **Measuring physical activity-related environmental factors: reliability and predictive validity of the European environmental questionnaire ALPHA.** *Int J Behav Nutr Phys Act* 2010, **7:**48.

99. Department for Transport: **National Travel Survey 2011 GPS pilot: summary analysis.** 2011.

100. Department for Transport: **National Travel Survey GPS Feasibility Study Final Report.** 2009.

101. Stopher P, Clifford P, Zhang E, Fitzgerald F: **Deducing mode and purpose from GPS data.** *Institute of Transport and Logistics Studies Working Paper* 2008.

102. Bricka S, Bhat CR: **Comparative analysis of global positioning system-based and travel survey-based data.** In *Travel Survey Methods, Information Technology, and Geospatial Data,.* 2006: 9-20: *Transportation Research Record-Series*].

103. Wolf J, Oliveira M, Thompson M: **Impact of underreporting an mileage and travel time estimates - Results from Global Positioning System-enhanced Household Travel Survey.** In *Traveler Behavior and Values 2003: Planning and Administration.* 2003: 189-198: *Transportation Research Record*].

104. Rodriguez DA, Brown AL, Troped PJ: **Portable global positioning units to complement accelerometry-based physical activity monitors.** *Med Sci Sports Exerc* 2005, **37:**S572-581.

105. Krenn PJ, Titze S, Oja P, Jones AP, Ogilvie D: **Use of Global Positioning Systems to Study Physical Activity and the Environment: A Systematic Review.** *Am J Prev Med* 2011, **41:**508-515.

106. Cho SJ, Ramachandran G, Grengs J, Ryan AD, Eberly LE, Adgate JL: **Longitudinal evaluation of allergen and culturable fungal concentrations in inner-city households.** *Journal of occupational and environmental hygiene* 2008, **5:**107-118.

107. Cooper AR, Page AS, Wheeler BW, Hillsdon M, Griew P, Jago R: **Patterns of GPS measured time outdoors after school and objective physical activity in English children: the PEACH project.** *Int J Behav Nutr Phys Act* 2010, **7:**31.

108. Gong H, Chen C, Bialostozky E, Lawson CT: **A GPS/GIS method for travel mode detection in New York City.** *Computers, Environment and Urban Systems* 2012, **36:**131-139.

109. Stopher PR, Stecher C: *Travel survey methods : quality and future directions.* Amsterdam ; Oxford: Elsevier; 2006.

110. Rowlands AV: **Accelerometer assessment of physical activity in children: an update.** *Pediatr Exerc Sci* 2007, **19:**252-266.

111. Welch WA, Bassett DR, Thompson DL, Freedson PS, Staudenmayer JW, John D, Steeves JA, Conger SA, Ceaser T, Howe CA, et al: **Classification Accuracy of the Wrist-Worn GENEA Accelerometer.** *Medicine and Science in Sports and Exercise* 2013.

112. Cooper AR, Page AS, Wheeler BW, Griew P, Davis L, Hillsdon M, Jago R: **Mapping the Walk to School Using Accelerometry Combined with a Global Positioning System.** *Am J Prev Med* 2010, **38:**178-183.

113. Rodriguez DA, Cho GH, Elder JP, Conway TL, Evenson KR, Ghosh-Dastidar B, Shay E, Cohen D, Veblen-Mortenson S, Pickrell J, Lytle L: **Identifying walking trips from GPS and accelerometer data in adolescent females.** *J Phys Act Health* 2012, **9:**421-431.

114. Louisa Ells, John Wilkinson, Stuart Simms, Gillian Bryant , Marianne Law, Heasman M: **Physical Activity in the North East.** North East Public Health Observatory; 2008.

115. Bajekal M, Osborne V, Yar M, Meltzer H: **Focus on health.** 2006.

116. Coronini-Cronberg S, Millett C, Laverty AA, Webb E: **The impact of a free older persons' bus pass on active travel and regular walking in England.** *Am J Public Health* 2012, **102:**2141-2148.

117. Mindell JS, Leslie D, Wardlaw M: **Exposure-based, 'like-for-like' assessment of road safety by travel mode using routine health data.** *PLoS One* 2012, **7:**e50606.

118. Brennan G, Blake L, Hill-McManus D, Payne N, Buckley Woods H, Blank L: **Health and economic modelling report.** School of Health and Related Research; 2012.

119. National Center for Social Research: **NTS Travel Record Review Stage 2.** Prepared for Department for Transport; 2006.

120. Josi Rofique, Alun Humphrey, Kevin Pickering, Tipping S: **National Travel Survey 2010 Technical Report.** Prepared for Department of Transport; 2011.

121. Alexandra Cronberg, Peter Bonsall: **Technical Annex: Analysis of the accuracy of NTS respondents' estimates of trip distance and duration.** Prepared for the Department for Transport; 2006.

122. Forrest TL, Pearson DF: **Comparison of trip determination methods in household travel surveys enhanced by a Global Positioning System.** *Data Initiatives* 2005**:**63-71.

123. Ogle J: **Quantitative assessment of driver speeding behavior using instrumented vehicles.** Georgia Institute of Technology, Department of Civil Engineering; 2005.

124. Wolf J, Loechl M, Thompson M, Arce C: **Trip rate analysis in GPS-enhanced personal travel surveys.** *Transport Survey Quality and Innovation* 2003**:**483-498.

125. Huebner KD, Porter MM, Marshall SC: **Validation of an electronic device for measuring driving exposure.** *Traffic Injury Prevention* 2006, **7:**76-80.

126. Tudor-Locke CE, Myers AM: **Challenges and opportunities for measuring physical activity in sedentary adults.** *Sports Medicine* 2001, **31:**91-100.

127. Tudor-Locke C, Neff LJ, Ainsworth BE, Addy CL, Popkin BM: **Omission of active commuting to school and the prevalence of children's health-related physical activity levels: the Russian Longitudinal Monitoring Study.** *Child Care Hlth Dev* 2002, **28:**507-512.

128. Stopher P, Fitzgerald P, Bretain C, Zhang P: **Variability in day-to-day travel: analysis of a 28-day GPS survey.** *Australasian Transport Research Forum* 2006, **29**.

129. Houston D, Ong P, Jaimes G, Winer A: **Traffic exposure near the Los Angeles-Long Beach port complex: using GPS-enhanced tracking to assess the implications of unreported travel and locations.** *Journal of Transport Geography* 2011, **19:**1399-1409.

130. Hodges S, Berry E, Wood K: **SenseCam: a wearable camera that stimulates and rehabilitates autobiographical memory.** *Memory* 2011, **19:**685-696.

131. Bush V: **As we may think (Published originally in English in the Atlantic Monthly, July 1945).** *Revista De Occidente* 2001**:**19.

132. Mann S: **Wearable computing: A first step toward personal imaging.** *Computer* 1997, **30:**25-31.

133. Berry E, Byrne D, Doherty A, Gurrin C, Smeaton A: **Proceedings of the second annual SenseCam symposium (SenseCam 2010).** In *SenseCAm 2010*; *16-17 September 2010; Dublin City University, Dublin, Ireland*. 2010

134. Prosser J, Loxley A: **ESRC National Centre for Research Methods Review Paper: Introducing visual methods.** National Centre for Research Methods; 2008.

135. Arab L, Estrin D, Kim DH, Burke J, Goldman J: **Feasibility testing of an automated image-capture method to aid dietary recall.** *Eur J Clin Nutr* 2011, **65:**1156-1162.

136. Zhang H, Li L, Jia WY, Fernstrom JD, Sclabassi RJ, Sun MG: **Recognizing Physical Activity from Ego-motion of A Camera.** *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Embc)* 2010**:**5569-5572.

137. Doherty AR, Hodges SE, King AC, Smeaton AF, Berry E, Moulin CJA, Lindley S, Kelly P, Foster C: **Wearable Cameras in Health: The State of the Art and Future Possibilities.** *Am J Prev Med* 2013, **44:**320-323.

138. Gurrin C, Qiu Z, Hughes M, Caprani N, Doherty AR, Hodges SE, Smeaton AF: **The Smartphone As a Platform for Wearable Cameras in Health Research.** *Am J Prev Med* 2013, **44:**308-313.

139. Kelly P, Marshall SJ, Badland H, Kerr J, Oliver M, Doherty AR, Foster C: **An Ethical Framework for Automated, Wearable Cameras in Health Behavior Research.** *Am J Prev Med* 2013, **44:**314-319.

140. Kerr J, Marshall S, Godbole S, Chen J, Legge A, Doherty A, Kelly P, Oliver M, Badland H, Foster C: **Using the SenseCam to Improve Classifications of Sedentary Behavior in Free-Living Settings.** *Am J Prev Med* 2013, **44:**290-296.

141. O'Loughlin G, Cullen SJ, McGoldrick A, O'Connor S, Blain R, O'Malley S, Warrington GD: **Using a Wearable Camera to Increase the Accuracy of Dietary Analysis.** *Am J Prev Med* 2013, **44:**297-301.

142. Silva AR, Pinho S, Macedo LM, Moulin CJ: **Benefits of SenseCam Review on Neuropsychological Test Performance.** *Am J Prev Med* 2013, **44:**302-307.

143. Stopher P, Kockelman K, Greaves S, Clifford E: **Reducing Burden and Sample Sizes in Multiday Household Travel Surveys.** *Transportation Research Record: Journal of the Transportation Research Board* 2008, **2064:**12-18.

144. Kirkwood BR, Sterne JAC: *Essential medical statistics.* 2nd ed. / Betty R. Kirkwood, Jonathan A.C. Sterne. edn. Malden, Mass. ; Oxford: Blackwell Science; 2003.

145. Bachu P, Dudala T, Kothuri S: **Prompted recall in global positioning system survey - Proof-of-concept study.** In *Transportation Data and Information Technology: Planning and Administration.* 2001: 106-113: *Transportation Research Record*].

146. Bricka S, Zmud J, Wolf J, Freedman J: **Household Travel Surveys with GPS An Experiment.** *Transportation Research Record* 2009**:**51-56.

147. Cho G, Rodriguez D, Evenson K: **Identifying Walking Trips using GPS Data.** In *89th Annual Meeting Transportation Research Board; Washington.* 2010

148. Stopher P, Bullock C, Horst P: **Exploring the use of passive GPS devices to measure travel.** *Working paper (University of Sydney Institute of Transport and Logistics Studies) ; ITS-WP-02-06* 2002.

149. Stopher P, Greaves S: **Missing and inaccurate information from travel surveys : pilot results.** Institute of Transport and Logistics Studies; 2010.

150. Schönfelder S, Samaga U: **Where do you want to go today? - More observations on daily mobility.** *3rd Swiss Transport Research Conference Proceedings* 2003.

151. Schuessler N, Axhausen KW: **Processing Raw Data from Global Positioning Systems Without Additional Information.** *Transportation Research Record* 2009**:**28-36.

152. Krygsman S, Nel J: **The use of global positioning devices in travel surveys - a developing country application.** *28th Annual Southern African Transport Conference 2009 Proceedings* 2009.

153. Auld J, Williams C, Mohammadian A, Nelson P: **An Automated GPS-Based Prompted Recall Survey with Learning Algorithms.** *Transportation Letters International Journal of Transportation Research* 2009, **1**.

154. Kelly P, Doherty A, Berry E, Hodges S, Batterham AM, Foster C: **Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study.** *Int J Behav Nutr Phy* 2011, **8**.

155. Clark, Mckimm S: **Using GPS to collect journey time information.** *Traffic Engineering & Control* 2003, **44**.

156. Li YY, McDonald M: **Link travel time estimation using single GPS equipped probe vehicle.** *Ieee 5th International Conference on Intelligent Transportation Systems, Proceedings* 2002**:**932-937.

157. Kelly P, Doherty AR, Hamilton A, Matthews A, Batterham AM, Nelson M, Foster C, Cowburn G: **Evaluating the Feasibility of Measuring Travel to School Using a Wearable Camera.** *Am J Prev Med* 2012, **43:**546-550.

158. Chen C, Gong HM, Lawson C, Bialostozky E: **Evaluating the feasibility of a passive travel survey collection in a complex urban environment: Lessons learned from the New York City case study.** *Transport Res a-Pol* 2010, **44:**830-840.

159. Garg AX, Hackam D, Tonelli M: **Systematic review and meta-analysis: when one study is just not enough.** *Clin J Am Soc Nephrol* 2008, **3:**253-260.

160. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC: **Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist.** *Qual Life Res* 2012, **21:**651-657.

161. Bland JM, Altman DG: **Statistical methods for assessing agreement between two methods of clinical measurement.** *Lancet* 1986, **1:**307-310.

162. Pearson D, Hard E, Farnsworth S, Forrest T, Spillane D, Ojah M, Womack K, Boxill S, Lewis C: **Improving the accuracy in household and external travel surveys.** Texas Transportation Institute The Texas A&M University System College Station, Texas 77843-3135; 2010.

163. **Diet and physical activity measurement toolkit (accessed 7/7/12)** [http://dapa-toolkit.mrc.ac.uk/] Retrieved 7/7/12.

164. Worthen BR, Borg WR, White K: *Measurement and evaluation in the schools.* Longman; 1993.

165. Mehrens WA, Lehmann IJ: *Using standardized tests in education.* Longman; 1987.

166. Rennie KL, Wareham NJ: **The validation of physical activity instruments for measuring energy expenditure: problems and pitfalls.** *Public Health Nutrition* 1998, **1:**265-271.

167. **Research Methods Knowledge Base (accessed 7/7/12),** [http://www.socialresearchmethods.net/kb/index.php] Retrieved 7/7/12.

168. Patterson P: **Reliability, validity, and methodological response to the assessment of physical activity via self-report.** *Research quarterly for exercise and sport* 2000, **71:**S15-20.

169. Anastasi A, Urbina S: *Psychological testing.* Prentice Hall; 1997.

170. Macfarlane DJ, Lee CCY, Ho EYK, Chan KL, Chan D: **Convergent validity of six methods to assess physical activity in daily life.** *Journal of Applied Physiology* 2006, **101:**1328-1334.

171. Hermida J, Nicholas DD, Blumenfeld SN: **Methodology matters. Comparative validity of three methods for assessment of the quality of primary health care.** *International Journal for Quality in Health Care* 1999, **11:**429-433.

172. Hopkins WG, Marshall SW, Batterham AM, Hanin J: **Progressive statistics for studies in sports medicine and exercise science.** *Med Sci Sports Exerc* 2009, **41:**3-13.

173. Gibson RS: *Principles Of Nutritional Assessment.* Oxford University Press, Incorporated; 2005.

174. Hanneman SK: **Design, analysis, and interpretation of method-comparison studies.** *AACN Adv Crit Care* 2008, **19:**223-234.

175. Mantha S, Roizen MF, Fleisher LA, Thisted R, Foss J: **Comparing methods of clinical measurement: reporting standards for bland and altman analysis.** *Anesth Analg* 2000, **90:**593-602.

176. Streiner DL, Norman GR: **“Precision” and “Accuracy”: Two Terms That Are Neither.** *Journal of Clinical Epidemiology* 2006, **59:**327-330.

177. Transport for NSW Bureau of Transport Statistics: **Household Travel Survey 2010/11 Technical Documentation.** 2012.

178. California Department of Transportation: **California Statewide Household Travel Survey Final Report.** 2002.

179. van Sluijs EMF, van Poppel MNM, Twisk JWR, Chin A Paw MJ, Calfas KJ, van Mechelen W: **Effect of a Tailored Physical Activity Intervention Delivered in General Practice Settings: Results of a Randomized Controlled Trial.** *American Journal of Public Health* 2005, **95:**1825-1831.

180. Doherty AR, Moulin CJ, Smeaton AF: **Automatically assisting human memory: a SenseCam browser.** *Memory* 2011, **19:**785-795.

181. Doherty AR, Smeaton AF: **Automatically Segmenting LifeLog Data into Events.** In *Image Analysis for Multimedia Interactive Services, 2008 WIAMIS '08 Ninth International Workshop on*; *7-9 May 2008*. 2008: 20-23.

182. O'Hara K, Tuffield M, Shadbolt N: **Lifelogging: Privacy and empowerment with memories for life.** *Identity in the Information Society* 2008, **1:**155-172.

183. David H. Nguyen, Gabriela Marcu, Gillian R. Hayes, Khai N. Truong, James Scott, Marc Langheinrich, Roduner C: **Encountering SenseCam: Personal Recording Technologies in Everyday Life.** In *UbiComp 2009; Orlando, FL*. 2009

184. British Sociological Association: **Statement of ethical practice for the British Sociological Assocoiation - Visual Sociology Group.** 2006.

185. Wiles R, Prosser J, Bagnoli A, Clark A, Davies K, Holland S, Renold E: **Visual Ethics: Ethical Issues in Visual Research.** National Centre for Research Methods; 2008.

186. Papademas D: **Editor's introduction: ethics in visual research.** *Visual Studies* 2004, **19:**122-126.

187. Sun M, Fernstrom JD, Jia W, Hackworth SA, Yao N, Li Y, Li C, Fernstrom MH, Sclabassi RJ: **A wearable electronic system for objective dietary assessment.** *J Am Diet Assoc* 2010, **110:**45-47.

188. Arab L, Estrin D, Kim DH, Burke J, Goldman J: **Feasibility testing of an automated image-capture method to aid dietary recall.** *Eur J Clin Nutr* 2011, **65:**1156-1162.

189. Kelly P, Doherty A, Berry E, Hodges S, Batterham AM, Foster C: **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* 2011, **8:**44.

190. Andrade N, Martin A, Monteleone S: **All the Better to See You with, My Dear": Facial Recognition and Privacy in Online Social Networks.** *Security & Privacy, IEEE* 2013, **PP:**1-1.

191. Allen AL: *Unpopular privacy : what must we hide?* New York ; Oxford: Oxford University Press; 2011.

192. P. Novitzky, B. Gordijn, T. Jacquemard, F. O'Brolcháin: **Dementia Ambient Care: Multi-Sensing Monitoring for Intelligent Remote Management and Decision Support: Ethical Literature Review.** 2012.

193. Hofmann B: **Ethical Challenges with Welfare Technology: A Review of the Literature.** *Sci Eng Ethics* 2012, **5:**5.

194. Bland JM, Altman DG: **Agreement between methods of measurement with multiple observations per individual.** *J Biopharm Stat* 2007, **17:**571-582.

195. Bland JM, Altman DG: **Calculating correlation coefficients with repeated observations: Part 1--Correlation within subjects.** *BMJ* 1995, **310:**446.

196. Stopher PR, Greaves SP: **Household travel surveys: Where are we going?** *Transport Res a-Pol* 2007, **41:**367-381.

197. Zmud J, Wolf J: **Identifying the correlates of trip misreporting—results from the California statewide household travel survey GPS study.** *International conference on travel behaviour research* 2003.

198. McNamara E, Hudson Z, Taylor SJ: **Measuring activity levels of young people: the validity of pedometers.** *Br Med Bull* 2010, **95:**121-137.

199. Ozdoba R, Corbin C, Le Masurier G: **Does reactivity exist in children when measuring activity levels with unsealed pedometers?** *Pediatr Exerc Sci* 2004, **16:**158-166.

200. Craig CL, Tudor-Locke C, Cragg S, Cameron C: **Process and Treatment of Pedometer Data Collection for Youth: The Canadian Physical Activity Levels among Youth Study.** *Medicine and Science in Sports and Exercise* 2010, **42:**430-435.

201. Rowe DA, Mahar MI, Raedeke TD, Lore J: **Measuring physical activity in children with pedometers: Reliability, reactivity, and replacement of missing data.** *Pediatr Exerc Sci* 2004, **16:**343-354.

202. **UK DATA SERVICE** [http://ukdataservice.ac.uk/] 2011.

203. Altman DG, Bland JM: **Missing data.** *BMJ* 2007, **334:**424.

204. Sirard JR, Slater ME: **Compliance with wearing physical activity accelerometers in high school students.** *J Phys Act Health* 2009, **6 Suppl 1:**S148-155.

205. Sport England: **Local area walking and cycling statistics: England 2011/12.** Department for Transport,; 2013.

206. **Romford - transport** [http://en.wikipedia.org/wiki/Romford] Retrieved 18/4/2013.

207. Tin S, Woodward A, Thornley S, Ameratunga S: **Regional variations in pedal cyclist injuries in New Zealand: safety in numbers or risk in scarcity?** *Australian and New Zealand Journal of Public Health* 2011, **35:**357-363.

208. Buckland L, Jones MG: **Estimating Bicycle and Pedestrian Demand in San Diego.** In *Transportation Research Board 87th Annual Meeting*. 2008

209. [http://www.sdmts.com/MTS/timeline.asp] Retrieved 18/4/2013.

210. Landis JR, Koch GG: **The measurement of observer agreement for categorical data.** *Biometrics* 1977, **33:**159-174.

211. Bland JM, Altman DG: **Transforming data.** *BMJ* 1996, **312:**770.

212. Brehm MA, Scholtes VA, Dallmeijer AJ, Twisk JW, Harlaar J: **The importance of addressing heteroscedasticity in the reliability analysis of ratio-scaled variables: an example based on walking energy-cost measurements.** *Dev Med Child Neurol* 2012, **54:**267-273.

213. Bland JM, Altman DG: **Calculating correlation coefficients with repeated observations: Part 2--Correlation between subjects.** *BMJ* 1995, **310:**633.

214. Welk G: *Physical activity assessments for health-related research.* Champaign, IL: Human Kinetics; 2002.

215. Foster CE, Brennan G, Matthews A, McAdam C, Fitzsimons C, Mutrie N: **Recruiting participants to walking intervention studies: a systematic review.** *Int J Behav Nutr Phys Act* 2011, **8:**137.

216. Dollman J, Okely AD, Hardy L, Timperio A, Salmon J, Hills AP: **A hitchhiker's guide to assessing young people's physical activity: Deciding what method to use.** *Journal of science and medicine in sport / Sports Medicine Australia* 2009, **12:**518-525.

217. Tarrant MA, Manfredo MJ: **Digit Preference, Recall Bias, and Nonresponse Bias in Self-Reports of Angling Participation.** *Leisure Sciences* 1993, **15:**231-238.

218. Schmidt ME, Steindorf K: **Statistical methods for the validation of questionnaires--discrepancy between theory and practice.** *Methods Inf Med* 2006, **45:**409-413.

219. Ogilvie D, Panter J: **Expert paper 5 'Submission to the NICE programme development group on walking and cycling'.** 2012.

220. Sarmiento O, Torres A, Jacoby E, Pratt M, Schmid TL, Stierling G: **The Ciclovia-Recreativa: A mass-recreational program with public health potential.** *J Phys Act Health* 2010, **7 Suppl 2:**S163-180.

221. Torres A, Sarmiento OL, Stauber C, Zarama R: **The Ciclovia and Cicloruta programs: promising interventions to promote physical activity and social capital in Bogota, Colombia.** *Am J Public Health* 2013, **103:**e23-30.

222. Craig CL, Marshall AL, Sjostrom M, Bauman AE, Booth ML, Ainsworth BE, Pratt M, Ekelund U, Yngve A, Sallis JF, Oja P: **International physical activity questionnaire: 12-country reliability and validity.** *Medicine and Science in Sports and Exercise* 2003, **35:**1381-1395.

223. Wareham NJ, Jakes RW, Rennie KL, Mitchell J, Hennings S, Day NE: **Validity and repeatability of the EPIC-Norfolk Physical Activity Questionnaire.** *International Journal of Epidemiology* 2002, **31:**168-174.

224. Adams MA, Johnson WD, Tudor-Locke C: **Steps/day translation of the moderate-to-vigorous physical activity guideline for children and adolescents.** *Int J Behav Nutr Phys Act* 2013, **10:**49.

225. Spiegelman D, Zhao B, Kim J: **Correlated errors in biased surrogates: study designs and methods for measurement error correction.** *Stat Med* 2005, **24:**1657-1682.

226. Roth MA, Mindell JS: **Who provides accelerometry data? Correlates of adherence to wearing an accelerometry motion sensor: the 2008 Health Survey for England.** *J Phys Act Health* 2013, **10:**70-78.

227. Rutter H, Cavill N, Racioppi F, Dinsdale H, Oja P, Kahlmeier S: **Economic impact of reduced mortality due to increased cycling.** *Am J Prev Med* 2013, **44:**89-92.

228. Oliver M, Doherty AR, Kelly P, Badland HM, Mavoa S, Shepherd J, Kerr J, Marshall S, Hamilton A, Foster C: **Utility of passive photography to objectively audit built environment features of active transport journeys: an observational study.** *Int J Health Geogr* 2013, **12:**20.

229. McClain JJ, Tudor-Locke C: **Objective monitoring of physical activity in children: considerations for instrument selection.** *Journal of science and medicine in sport / Sports Medicine Australia* 2009, **12:**526-533.

Appendix 1 - Estimates of the number of people in England who attain or exceed vigorous intensity exercise by walking at 3 mph

Journal of Sports Sciences, 2011; 1–6, i*First* article

Routledge
Taylor & Francis Group

Estimates of the number of people in England who attain or exceed vigorous intensity exercise by walking at 3 mph

PAUL KELLY[1], MARIE MURPHY[2], PEKKA OJA[3], ELAINE M. MURTAUGH[4], & CHARLIE FOSTER[1]

[1]*BHF Health Promotion Research Group, University of Oxford, Oxford, UK,* [2]*University of Ulster, Sport and Exercise Science Research Institute, Jordanstown, UK,* [3]*UKK Institute for Health Promotion Research, Tampare, Finland, and* [4]*Mary Immaculate College, Arts Education and Physical Education, South Circular Road, Limerick, Co. Limerick, Ireland*

(Accepted 29 July 2011)

Abstract

Walking is a safe, accessible and low cost activity, amenable to change and known to have great potential to increase physical activity levels in sedentary individuals. The objective of this study is to estimate the proportion of the 2009 adult population of England who would attain or exceed vigorous intensity activity ($>70\%$ maximum heart rate [HR_{max}]) by walking at 3 mph. We conducted predictive impact modelling using participants' ($n = 1741$, aged 25–64 years) cardiovascular fitness data from treadmill walking tests. We combined this data with English population estimates adjusted for age and sex to estimate the numbers of individuals that would exceed 70% HR_{max} (an intensity considered sufficient for fitness gains) when walking at 3 mph (4.8 km $\cdot$ h^{-1}). We estimate 1.5 million men (95% confidence interval [CI] 0.9–2.2 million) (from 13.4 million corresponding to 11.6% (95% CI 7.0–16.2%)) and 3.9 million women (95% CI 3.0–4.8 million) (from 13.6 million corresponding to 28.6% (95% CI 22.0–35.1%)) in England aged 25–64 years would benefit from regularly walking at 3 mph. In total, a projected 5.4 million individuals (95% CI 3.9–6.9 million) aged 25–64 (from 26.97 million corresponding to 20.1% (95% CI 14.6–25.7%)) could benefit from walking at 3 mph. Our estimates suggest a considerable number of individuals in the English population could receive fitness and health benefits by walking regularly at 3 mph. Physical activity messages that promote walking at this speed may therefore have the potential to significantly impact national fitness levels and health in England.

Keywords: *Walking, fitness, public health*

Introduction

Walking has been described as the safest, most convenient form of physical activity. It is low-impact, low cost and readily accessible, requiring no special skills or equipment (Albright, 2000; Ogilvie et al., 2007). It can be readily assimilated into daily life through active transportation and can be continued into old age (Ogilvie et al., 2007). It is an important form of activity to many individuals due to an unwillingness or inability of a large proportion of the population to participate in more vigorous activities (Hamer & Steptoe, 2008). Systematic review of exercise interventions has shown walking to have the greatest potential for increasing activity levels in sedentary individuals (Hillsdon, Foster, & Thorogood, 2005; Hillsdon & Thorogood, 1996). As many adults in England are not meeting physical activity recommendations (Department of Health, 2006, 2008), walking may be a suitable activity to address this issue.

Despite the potential benefits of walking current levels are relatively low; the 2008 Health Survey for England found that only 41% of men and 34% of women reported any walking in the previous four weeks (Department of Health, 2008). Other studies have shown that this walking is socially patterned and the pattern varies by walking purpose; walks over 30 minutes at a fast or brisk pace are more common among affluent groups, whereas walking for transport reveals an opposite trend (Department of Health, 2006, 2008; Scottish Executive Health Department, 2005; Transport Statistics, 2006).

Attaining moderate levels of fitness through physical activities such as walking is related to significantly lower death rates, and has protective

Correspondence: Paul Kelly, BHF Health Promotion Research Group, University of Oxford, Oxford, United Kingdom. E-mail: paul.kelly@dph.ox.ac.uk

ISSN 0264-0414 print/ISSN 1466-447X online © 2011 Taylor & Francis
http://dx.doi.org/10.1080/02640414.2011.609897

Appendix 2 – Quantifying the difference between self-reported and GPS measured journey durations: a systematic review

Transport Reviews, 2013
http://dx.doi.org/10.1080/01441647.2013.815288

Quantifying the Difference Between Self-Reported and Global Positioning Systems-Measured Journey Durations: A Systematic Review

PAUL KELLY*§, PATRICIA KRENN**, SYLVIA TITZE**, PETER STOPHER† AND CHARLIE FOSTER*

*Department of Public Health, British Heart Foundation Health Promotion Research Group, University of Oxford, Oxford, UK; **Institute of Sport Science, University of Graz, Graz, Austria; †Institute of Transport and Logistics Studies, University of Sydney, Sydney, Australia

(*Received 4 March 2013; revised 10 June 2013; accepted 11 June 2013*)

ABSTRACT *Accurate measurement of travel behaviour is vital for transport planning, modelling, public health epidemiology, and assessing the impact of travel interventions. Self-reported diaries and questionnaires are traditionally used as measurement tools; advances in Global Positioning Systems (GPS) technology allow for comparison. This review aimed to identify and report about studies comparing self-reported and GPS-measured journey durations. We systematically searched, appraised, and analysed published and unpublished articles from electronic databases, reference lists, bibliographies, and websites up to December 2012. Included studies used GPS and self-report to investigate trip duration. The average trip duration from each measure was compared and an aggregated, pooled estimate of the difference, weighted by number of trips, was calculated. We found 12 results from eight eligible studies. All studies showed self-reported journey times were greater than GPS-measured times. The difference between self-report and GPS times ranged from over-reporting of +2.2 to +13.5 minutes per journey. The aggregated, pooled estimate of the difference, weighted by number of trips, was over-report of +4.4 minutes (+28.6%). Studies comparing self-reported and GPS-measured journey duration have shown self-reported to be consistently over-reported across the study sample. Our findings suggest that when using self-reported journey behaviour, the journey durations should be treated as an over-estimation.*

Keywords: measurement; self-report; Global Positioning Systems; health; travel duration

Introduction

Active transport, primarily walking and cycling, is an increasingly important behaviour in the fields of public health, environmental sustainability, and transport planning (Cavill, Rutter, & Hill, 2007; Ogilvie et al., 2007; Smith, Gidlow, Davey, & Foster, 2010). It provides an opportunity to accumulate health-enhancing physical activity that can contribute to meeting physical activity recommendations (Bull et al., 2010; Department of Health, 2004; World Health Organisation, 2007).

§Corresponding author. Email: paul.kelly@dph.ox.ac.uk

Search terms by database searched

Embase (OvidSP) [1974 – present]
1. car driving/ or car driver/
2. walking/
3. bicycle/
4. "traffic and transport"/
5. (driving adj5 (pattern* or behavior* or behaviour* or habit* or frequen*)).ti,ab.
6. (travel* adj5 (pattern* or behavior* or behaviour* or habit* or frequen*)).ti,ab.
7. (transport* adj5 (pattern* or behavior* or behaviour* or habit* or frequen*)).ti,ab.
8. (driving adj5 (distance* or duration or time* or mileage*)).ti,ab.
9. (travel* adj5 (distance* or duration or time* or mileage* or data)).ti,ab.
10. (transport* adj5 (distance* or duration or time* or mileage* or data)).ti,ab.
11. (walk or walking).ti,ab.
12. (bicycling or cycling).ti,ab.
13. (active adj5 (travel* or transport* or commut*)).ti,ab.
14. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13
15. geographic information system/
16. global positioning.ti,ab.
17. (positioning adj3 (system or systems or technology or device*)).ti,ab.
18. (geographic information adj3 (system or systems or technology or device*)).ti,ab.
19. (gps or gis).ti,ab.
20. ((invehicle or in-vehicle) adj3 (system or systems or technology or device*)).ti,ab.
21. ((tracking or tracker) adj3 (system or systems or technology or device*)).ti,ab.
22. (electronic adj3 (system or systems or technology or device*)).ti,ab.
23. 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22
24. 14 and 23
25. data collection method/
26. questionnaire/
27. self report/
28. (self-report* or selfreport*).ti,ab.
29. (self-assess* or selfassess*).ti,ab.
30. (diary or diaries).ti,ab.
31. ((trip or trips or travel*) adj3 (log or logs or logging)).ti,ab.
32. survey*.ti,ab.
33. 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32
34. 14 and 23 and 33

GeoRef (OvidSP) [1966 – present]
1. transportation/
2. (driving adj5 (pattern* or behavior* or behaviour* or habit* or frequen*)).ti,ab.
3. (travel* adj5 (pattern* or behavior* or behaviour* or habit* or frequen*)).ti,ab.
4. (transport* adj5 (pattern* or behavior* or behaviour* or habit* or frequen*)).ti,ab.
5. (driving adj5 (distance* or duration or time* or mileage*)).ti,ab.
6. (travel* adj5 (distance* or duration or time* or mileage* or data)).ti,ab.
7. (transport* adj5 (distance* or duration or time* or mileage* or data)).ti,ab.
8. (walk or walking).ti,ab.
9. (bicycling or cycling).ti,ab.
10. (active adj5 (travel* or transport* or commut*)).ti,ab.
11. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
12. Geographic Information Systems/
13. global positioning system/
14. global positioning.ti,ab.
15. (positioning adj3 (system or systems or technology or device*)).ti,ab.
16. (geographic information adj3 (system or systems or technology or device*)).ti,ab.
17. (gps or gis).ti,ab.

18. ((invehicle or in-vehicle) adj3 (system or systems or technology or device*)).ti,ab.
19. ((tracking or tracker) adj3 (system or systems or technology or device*)).ti,ab.
20. (electronic adj3 (system or systems or technology or device*)).ti,ab.
21. 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20
22. (self-report* or selfreport*).ti,ab.
23. (self-assess* or selfassess*).ti,ab.
24. (diary or diaries).ti,ab.
25. ((trip or trips or travel*) adj3 (log or logs or logging)).ti,ab.
26. survey*.ti,ab.
27. data collection.ti,ab.
28. 22 or 23 or 24 or 25 or 26 or 27
29. 11 and 21 and 28

Medline (OvidSP) [1948 – present]
1. Automobile Driving/
2. exp Walking/
3. Bicycling/
4. Transportation/
5. (driving adj5 (pattern* or behavior* or behaviour* or habit* or frequen*)).ti,ab.
6. (travel* adj5 (pattern* or behavior* or behaviour* or habit* or frequen*)).ti,ab.
7. (transport* adj5 (pattern* or behavior* or behaviour* or habit* or frequen*)).ti,ab.
8. (driving adj5 (distance* or duration or time* or mileage*)).ti,ab.
9. (travel* adj5 (distance* or duration or time* or mileage* or data)).ti,ab.
10. (transport* adj5 (distance* or duration or time* or mileage* or data)).ti,ab.
11. (walk or walking).ti,ab.
12. (bicycling or cycling).ti,ab.
13. (active adj5 (travel* or transport* or commut*)).ti,ab.
14. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13
15. Geographic Information Systems/
16. global positioning.ti,ab.
17. (positioning adj3 (system or systems or technology or device*)).ti,ab.
18. (geographic information adj3 (system or systems or technology or device*)).ti,ab.
19. (gps or gis).ti,ab.
20. ((invehicle or in-vehicle) adj3 (system or systems or technology or device*)).ti,ab.
21. ((tracking or tracker) adj3 (system or systems or technology or device*)).ti,ab.
22. (electronic adj3 (system or systems or technology or device*)).ti,ab.
23. 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22
24. 14 and 23
25. Data Collection/
26. Questionnaires/
27. Self Report/
28. Self-Assessment/
29. (self-report* or selfreport*).ti,ab.
30. (self-assess* or selfassess*).ti,ab.
31. (diary or diaries).ti,ab.
32. ((trip or trips or travel*) adj3 (log or logs or logging)).ti,ab.
33. survey*.ti,ab.
34. 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33
35. 14 and 23 and 34

Transport Database (OvidSP) [1966 – present]
1. (driving adj5 (pattern* or behavior* or behaviour* or habit* or frequen*)).ti,ab.
2. (travel* adj5 (pattern* or behavior* or behaviour* or habit* or frequen*)).ti,ab.
3. (transport* adj5 (pattern* or behavior* or behaviour* or habit* or frequen*)).ti,ab.
4. (driving adj5 (distance* or duration or time* or mileage*)).ti,ab.
5. (travel* adj5 (distance* or duration or time* or mileage* or data)).ti,ab.
6. (transport* adj5 (distance* or duration or time* or mileage* or data)).ti,ab.
7. ((car or automobile*) adj3 (driver* or driving)).ti,ab.

8. (walk or walking).ti,ab.
9. (bicycling or cycling).ti,ab.
10. (active adj5 (travel* or transport* or commut*)).ti,ab.
11. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
12. global positioning.ti,ab.
13. (positioning adj3 (system or systems or technology or device*)).ti,ab.
14. (geographic information adj3 (system or systems or technology or device*)).ti,ab.
15. (gps or gis).ti,ab.
16. ((invehicle or in-vehicle) adj3 (system or systems or technology or device*)).ti,ab.
17. ((tracking or tracker) adj3 (system or systems or technology or device*)).ti,ab.
18. (electronic adj3 (system or systems or technology or device*)).ti,ab.
19. 12 or 13 or 14 or 15 or 16 or 17 or 18
20. (self-report* or selfreport*).ti,ab.
21. (self-assess* or selfassess*).ti,ab.
22. (diary or diaries).ti,ab.
23. ((trip or trips or travel*) adj3 (log or logs or logging)).ti,ab.
24. survey*.ti,ab.
25. data collection.ti,ab.
26. 20 or 21 or 22 or 23 or 24 or 25
27. 11 and 19 and 26

Web of Knowledge:

- Science Citation Index Expanded (SCI-EXPANDED) --1945-present
- Social Sciences Citation Index (SSCI) --1956-present
- Conference Proceedings Citation Index- Science (CPCI-S) --1990-present
- Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH) --1990-present

6 #5 AND #4 AND #3

5 Topic=(questionnaire*) OR Topic=("self report*" OR selfreport*) OR Topic=(diary OR diaries) OR Topic=("self assess* OR selfassess*") OR Topic=("data collection") OR Topic=(survey*)

4 Topic=("positioning system*" OR "positioning technolog*" OR "positioning device*") OR Topic=("global positioning") OR Topic=("geographic information system*" OR "geographic information technolog*" OR "geographic information device*") OR Topic=(gps OR gis) OR Topic=((invehicle OR in-vehicle OR tracker OR tracking) SAME (system* OR technolog* OR device*)) OR Topic=("electronic system*" OR "electronic technolog*" OR "electronic device*")

3 #2 OR #1

2 Topic=(walk OR walking) OR Topic=(bicycling OR cycling) OR Topic=(active SAME (travel* OR transport* OR commut*))

1 Topic=("car driving" OR "automobile driving" OR "car driver" OR "automobile driver") OR Topic=((driving OR travel OR transport*) SAME (pattern* OR behavior* OR behaviour* OR habit* OR frequen*)) OR Topic=((travel* OR transport*) SAME (distance* OR duration OR time* OR mileage* OR data))

Appendix 3 - Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study

INTERNATIONAL JOURNAL OF BEHAVIORAL NUTRITION AND PHYSICAL ACTIVITY

RESEARCH Open Access

Can we use digital life-log images to investigate active and sedentary travel behaviour? Results from a pilot study

Paul Kelly[1*], Aiden Doherty[1], Emma Berry[2], Steve Hodges[2], Alan M Batterham[3] and Charlie Foster[1]

Abstract

Background: Active travel such as walking and cycling has potential to increase physical activity levels in sedentary individuals. Motorised car travel is a sedentary behaviour that contributes to carbon emissions. There have been recent calls for technology that will improve our ability to measure these travel behaviours, and in particular evaluate modes and volumes of active versus sedentary travel. The purpose of this pilot study is to investigate the potential efficacy of a new electronic measurement device, a wearable digital camera called SenseCam, in travel research.

Methods: Participants (n = 20) were required to wear the SenseCam device for one full day of travel. The device automatically records approximately 3,600 time-stamped, first-person point-of-view images per day, without any action required by the wearer. Participants also completed a self-report travel diary over the same period for comparison, and were interviewed afterwards to assess user burden and experience.

Results: There were a total of 105 confirmed journeys in this pilot. The new SenseCam device recorded more journeys than the travel diary (99 vs. 94). Although the two measures demonstrated an acceptable correlation for journey duration ($r = 0.92$, $p < 0.001$) self-reported journey duration was over-reported (mean difference 154 s per journey; 95% CI = 89 to 218 s; 95% limits of agreement = 154 ± 598 s (-444 to 752 s)). The device also provided visual data that was used for directed interviews about sources of error.

Conclusions: Direct observation of travel behaviour from time-stamped images shows considerable potential in the field of travel research. Journey duration derived from direct observation of travel behaviour from time-stamped images appears to suggest over-reporting of self-reported journey duration.

Background

Active transportation, primarily walking and cycling, is an important behaviour in the fields of public health, environmental sustainability and transport planning [1-3]. From an environmental perspective, replacing carbon emitting motorised transport journeys with walking or cycling reduces pollutants and emissions, and can help to reduce traffic levels [2,4,5]. From a public health perspective, increasing an individual's walking and cycling contributes to meeting the international guideline amounts of five times thirty minutes per week of moderate to vigorous physical activity [6-10].

A recent meta analysis showed that regular walking is significantly associated with reduced risk for all cause mortality [11]. In terms of public health it is an important form of activity because of an unwillingness or inability of a large proportion of the population to participate in more vigorous activities [12]. It has been described as the safest, most convenient form of physical activity as it is low-impact, low cost and readily accessible, requiring no special skills or equipment [1,5,13]. Furthermore, walking can be easily assimilated into daily life and continued into old age [1]. Cycling, traditionally considered a more vigorous activity, has also been shown to be associated with reduced mortality and reduced cardiovascular risk [14,15].

Surveillance and monitoring of active and sedentary transport levels are therefore of interest. Much research

* Correspondence: paul.kelly@dph.ox.ac.uk
[1]British Heart Foundation Health Promotion Research Group, University of Oxford, UK
Full list of author information is available at the end of the article

Appendix 4 – Ethics approval from Oxford University

SOCIAL SCIENCES & HUMANITIES
INTER-DIVISIONAL RESEARCH ETHICS COMMITTEE

Hayes House, 75 George Street, Oxford. OX1 2BQ
Tel: +44(0)1865 614871 Fax: +44(0)1865 614855
ethics@socsci.ox.ac.uk www.socsci.ox.ac.uk

Co-ordinator of the IDREC
Social Sciences Divisional Office

UNIVERSITY OF OXFORD

Ref. SSD/2/3/IDREC

16 July 2010

Mr Paul Kelly,
Department of Public Health
DPHPC
Old Road Campus
Headington
Oxford
OX3 7LF

Dear Mr Paul Kelly,

Application Approval

Ref No.: SSD/CUREC1A/10-054

Title: Validation study between self-reported journey time and objectively measured journey time from digital images

The above application has been considered on behalf of the Social Sciences and Humanities Inter-divisional Research Ethics Committee (IDREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to the IDREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly approval has been granted.

Should there be any subsequent changes to the project, which raise ethical issues not covered in the original application, you should submit details to the IDREC for consideration.

Yours sincerely,

Chris. Ballinger

Dr Chris. Ballinger

cc. Dr Charlie Foster, Department of Public Health, Rosemary Rue Building, Old Road Campus
Carol Green, Division for Medical Sciences

HBP / CAJB

Appendix 5 - Evaluating the Feasibility of Measuring Travel to School Using a Wearable Camera

Evaluating the Feasibility of Measuring Travel to School Using a Wearable Camera

Paul Kelly, MSc, Aiden R. Doherty, PhD, Alex Hamilton, BA, Anne Matthews, PhD, Alan M. Batterham, PhD, Michael Nelson, PhD, Charlie Foster, PhD, Gill Cowburn, BSc

Background: The school journey is often studied in relation to health outcomes in children and adolescents. Self-report is the most common measurement tool.

Purpose: To investigate the error on self-reported journey duration in adolescents, using a wearable digital camera (Microsoft SenseCam).

Methods: During March–May 2011, participants (n=17; aged 13–15 years) from four schools wore wearable cameras to and from school for 1 week. The device automatically records time-stamped, first-person point-of-view images, without any action from the wearer. Participants also completed a researcher-administered self-report travel survey over the same period. Analysis took place in November 2011. Within- and between-subjects correlation coefficients and Bland-Altman 95% limits of agreement were derived, accounting for the multiple observations per individual.

Results: Self-report data were collected for 150 journey stages and SenseCam data for 135 (90%) of these. The within-subjects correlation coefficient for journey duration was 0.89 (95% CI=0.84, 0.93). The between-subjects correlation coefficient was 0.92 (95% CI=0.79, 0.97). The mean difference (bias) between methods at the whole sample level was small (10 seconds per journey, 95% CI= −33, 53). The wide limits of agreement (±501 seconds, 95% CI= −491, 511) reveal large random error.

Conclusions: Compared to direct observation from images, self-reported journey duration is accurate at the mean group level but imprecise at the level of the individual participant.
(Am J Prev Med 2012;43(5):546–550)

Background

Physical activity is associated with important health outcomes in children, including body composition, type 2 diabetes, and cardiovascular fitness.[1–6] Active travel to school, including walking and cycling, can be an important contributor to physical activity levels.[7,8] Conversely, travel in motor vehicles is a sedentary behavior representing a lost opportunity for physical activity.[9]

Research into school-related travel behavior faces methodologic challenges and valid, accurate measures are required.[10,11] Self-report is the most common tool, but its accuracy and precision are debated.[10,12–15] Better understanding of the accuracy and potential error in self-report will improve the calculation of health associations and the ability to detect changes in behavior.

Wearable digital cameras are novel devices that may help develop such an understanding. Microsoft's SenseCam is one such camera, worn on a lanyard around the neck that automatically records time-stamped, first-person point-of-view images (Figure 1), without any action required by the wearer.[12] It has been shown that a wearable camera can be used to estimate the bias and error on self-reported journey duration in adults.[16] The present study aims to see if the protocol can be repeated in a younger population. This study has two research questions: (1) Can a wearable camera be used to measure travel behavior in a sample of teenagers aged 13–15 years? (2) What do wearable camera–recorded journey durations reveal about self-reported journey durations?

From the British Heart Foundation Health Promotion Research Group (Kelly, Doherty, Hamilton, Matthews, Foster, Cowburn), University of Oxford, the Health and Social Care Institute (Batterham), Teesside University, and The School Food Trust (Nelson), United Kingdom

Address correspondence to: Paul Kelly, MSc, Department of Public Health, Old Road Campus, Headington, Oxford, OX3 7LF, United Kingdom. E-mail: paul.kelly@dph.ox.ac.uk.

0749-3797/$36.00

http://dx.doi.org/10.1016/j.amepre.2012.07.027

Methods

Participants

Volunteer participants (aged 13–15 years) were recruited from four secondary schools in England (Oxfordshire [three] and Yorkshire [one]) with a range of geographic locations (one city, two suburban, and one rural).

Appendix 6 – Ethics approval for School Food Journey study

SOCIAL SCIENCES & HUMANITIES
INTER-DIVISIONAL RESEARCH ETHICS COMMITTEE

Hayes House, 75 George Street, Oxford. OX1 2BQ
Tel: +44(0)1865 614871 Fax: +44(0)1865 614855
ethics@socsci.ox.ac.uk www.socsci.ox.ac.uk

Co-ordinator of the IDREC
Social Sciences Divisional Office

UNIVERSITY OF OXFORD

Ref. SSD/2/3/IDREC

8th November, 2010

Ms Gill Cowburn (Senior Researcher)
Department of Public Health
DPHPC, Old Road Campus
Headington
Oxford
OX3 7LF

Dear Gill,

Application Approval

Ref No.: SSD/CUREC1A/10-092

Title: The journey to school: Food access, purchases and consumption

The above application has been considered on behalf of the Social Sciences and Humanities Inter-divisional Research Ethics Committee (IDREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to the IDREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly approval has been granted.

Should there be any subsequent changes to the project, which raise ethical issues not covered in the original application, you should submit details to the IDREC for consideration.

Yours sincerely,

Chris. Ballinger

Dr Chris. Ballinger

Cc. Dr Anne Matthews, British Heart Foundation Health Promotion Research Group University of Oxford Department of Public Health Old Road Campus Headington Oxford
Dr Michael Nelson, Director of Research & Nutrition, School Food Trust, 6th Floor Sanctuary Buildings, Great Smith Street London SW1P 3BT

HBP / CAJB

Appendix 7 – Participant information sheet, ethical information and informed consent form

British Heart Foundation Health Promotion Research Group
Department of Public Health
Old Road Campus, Roosevelt Drive, Headington, Oxford OX3 7LF, UK
Tel: +44 (0)1865 289245, Fax: +44 (0) 1865 617789,

Information form for participants

Project title: **Assessing travel and sitting behaviour using self –reported travel diaries and digital photography**

May 2011

You are being invited to take part in a research project that will investigate the way we measure two important health behaviours; (1) how people travel, and (2) the type of sitting people do. This information sheet is designed to fully inform you about the project before you agree to take part.

People travel between their destinations by a variety of methods. These include walking, cycling, driving and using public transport. Research is needed to understand the health impacts of these behaviours, and to inform investment in local and national transport infrastructure. Sitting is also an important heath behaviour. There are many types; for example, working at a desk, watching TV, or reading. The most common way of gathering this information is by asking people to complete travel diaries that tell researchers what they have done. The purpose of this research project is to assess if we can get more detailed information from a wearable camera.

We have approached you to ask if you will take part in our project. Anyone between the age of 18-65 years is eligible to take part.

You are completely free to decide whether or not to take part in this study. If you decide to participate we will ask you to wear a digital camera during waking hours for three days. You are free to take off the camera while at home, work, or any other time you wish. We will also ask you to wear a small accelerometer (motion sensor) and GPS which can also be removed at any time.

In addition, we will ask you to complete a travel diary for the same three day period, and a sitting questionnaire after the 1st and 3rd day. These take approximately 5 minutes each. A researcher will visit you at work or home (depending on your preference) to deliver the camera and travel diary, and give instruction on how to use both. The camera works automatically so you just have to know how to turn it on and off. You may also collect the camera and diary yourself if you prefer. We will offer to send you optional email and/or SMS reminders to help you charge and wear the device.

After the third day, the researcher will return to collect the devices and diaries. You will be shown all of your images and given the option to delete any or all without giving any reason. All images can be deleted without the researcher seeing them. There will be an optional interview where you will be asked about your experiences of wearing the camera and completing the diary. You are free to opt out at anytime, without giving any reason.

We anticipate that participating in this project will have no significant effects on your lifestyle.

It is possible that while wearing the camera you may be asked about the device by members of the public. In this case we recommend saying the following:

> *"I am volunteering for a research project into journey behaviour. The device is called SenseCam and the images will be used to record my travel from place to place."*

Your confidentiality will be respected at all times. All your data and images will be anonymised and stored on a password protected computer. The results from this research project will be submitted for publication in an academic journal, and will be used to inform travel researchers on future work.

This project is supported by the British Heart Foundation and Microsoft Research Limited.

The project has been reviewed and approved by the Social Sciences & Humanities IDREC (Inter Divisional Research Ethics Committee) at the University of Oxford. If you have any questions or would like any more information please contact:

Mr Paul Kelly (paul.kelly@dph.ox.ac.uk; +44 (0)1865 617 782)

British Heart Foundation Health Promotion Research Group
Department of Public Health
Old Road Campus, Roosevelt Drive, Headington, Oxford OX3 7LF, UK
Tel: +44 (0)1865 289245, Fax: +44 (0) 1865 617789,

Ethics information for participants

This study has received ethics approval from the Social Sciences and Humanities Inter-divisional Research Ethics Committee (IDREC) in accordance with the procedures laid down by the University of Oxford (Ref No.: SSD/CUREC1A/10-054).

The data collected will be anonymised and stored on a password protected computer and any researchers using the data will have been trained in the ethical handling of such data.

In a typical day of wear, the SenseCam may take up to 3,000 images depicting where you go and the kinds of activities you do (e.g. walking to the shop, cycling, socialising). Some images are displayed below for you to get an idea of the content and quality of images that are recorded:

The devices are configured such that the data can only be accessed by members of the research team and it should be impossible for participants and/or third parties who find the devices to access the images. It will be stored according to the Data Protection Act (1998). Data of illegal activities may not be protected by confidentiality and may be passed to law enforcement.

It is important for you to understand that you are free to take off, or pause the recording of, the device at any time without giving a reason. You should not feel obliged to wear the device in situations where wearing the device may make you (or others) feel uncomfortable. If others around you feel uncomfortable with you wearing the device, you should offer to remove or temporarily switch off the device. Places where wearing the SenseCam may not be appropriate include banks, changing rooms/swimming pools, schools and cinemas.

We strongly recommend checking that friends, family, and co-workers understand the nature of the study and are happy for you to take part in advance. Their behaviour will not be reported as part of the research. They are welcome to contact the research team if they have any questions. If you are worried that the camera may have taken images of others that they would feel uncomfortable with, both you and the third party are free to request for those images to be deleted without giving any reason.

Over the course of the study you may forget you are wearing the camera and take images that are unwanted or unflattering. At the end of the period for which you are wearing the camera you will have the opportunity to view and, if necessary, delete any images that you do not wish to be included in the study. You will have the option of doing this either in private or with the help of a researcher. As the images are research data, you will not be able to keep any copies of the images.

Images may be used in scientific papers, presentations, and reports. We will not use any image that identifies people, participants' names or places of work without express written consent.

British Heart Foundation Health Promotion Research Group
Department of Public Health
Old Road Campus, Roosevelt Drive, Headington, Oxford OX3 7LF, UK
Tel: +44 (0)1865 289245, Fax: +44 (0) 1865 617789,

Informed consent form for participants

Project title: **Validation study between self-reported journey time and objectively measured journey time from digital images**

May 2011

Principal Investigator: **Mr Paul Kelly**

Please initial box

I confirm I have read and understood the participant information sheet dated May 2011 for the above study and have had the opportunity to ask questions which have been answered fully	
I understand that my participation is voluntary and I am free to withdraw at anytime without giving any reason, without my legal rights being affected	
I understand that I may request to have any or all of the digital images deleted at any time without giving any reason	
I agree to take part in the above study	

Name of participant……………………………………… Signature………………………………….. Date………………..

Participants date of birth…………………

Name of person taking consent……………..………… Signature………………………………….. Date………………..

Principal investigator………………………..…………… Signature………………………………….. Date………………..

1 copy for participant; 1copy for principal investigator

Appendix 8 – Participant instructions for using SenseCam

British Heart Foundation Health Promotion Research Group
Department of Public Health
Old Road Campus, Roosevelt Drive, Headington, Oxford OX3 7LF, UK
Tel: +44 (0)1865 289245, Fax: +44 (0) 1865 617789,

SenseCam Protocol

Thank you for agreeing to take part in this pilot study. If you have any further questions please contact paul.kelly@dph.ox.ac.uk

- Wear SenseCam around your neck so that it feels comfortable. The camera unit should be at chest height (as shown) with the lens facing horizontally forwards
- The height can be adjusted with the black chord.
- Wear outside of clothes
- Do not get wet
- Wear for all journeys outside of the home
- SenseCam can be used for cycling, walking, driving and public transport
- SenseCam is not suitable for prolonged running or jogging
- **All images will be treated with the strictest confidence and participants will be given the option to review and delete any or all images before analysis**

What if someone asks me what SenseCam is for?

It is unusual for somebody to be questioned about SenseCam. However we have found that explaining the nature of the device and of the study often successfully dispels any curiosity;

> *"I am volunteering for a research project into behaviour. The device is called SenseCam and the images record my travel and sitting time."*

If people are interested in this study and would like more information or would like to volunteer for this study they may contact Paul Kelly (paul.kelly@dph.ox.ac.uk).

Where should and shouldn't I wear SenseCam?

We are interested in recording your journeys and sitting time so we would like you to wear SenseCam as much as possible for the 3 day period.

We appreciate that in some places e.g. schools, hospitals or airports it may not be appropriate to wear the camera and you should feel comfortable to remove it at any time.

You do not need to wear SenseCam while exercising e.g. running, swimming or at the gym (though we would like you to record your journey there).

Personal safety

Remove the camera in any situation where you feel unsafe. For example, if you happen to be out on your own late at night, you may prefer to hide the camera to avoid unwanted attention. If someone tries to take the camera off you do not attempt to stop them.

If you are engaged in certain manual tasks or using machinery you SenseCam sometimes swing around. For example if you are using gardening machinery, we advise you remove the device until the activity is finished to avoid discomfort, or the possibility of the device getting caught dangerously.

British Heart Foundation Health Promotion Research Group
Department of Public Health
Old Road Campus, Roosevelt Drive, Headington, Oxford OX3 7LF, UK
Tel: +44 (0)1865 289245, Fax: +44 (0) 1865 617789,

Turning the camera on and off

The camera is turned on and off using the small round button on the top of the unit. A rising tones indicates that the camera is switching on and a falling tone indicates that the camera is switching off. The power button must be pressed for several seconds before the camera will respond and you will hear the tone.

The camera takes several seconds to power up, during which time the yellow light will be on. When the camera is ready for use, it will beep and the green power light will come on instead.

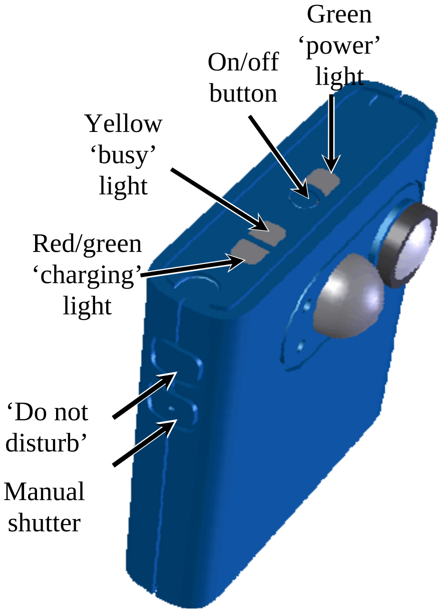
In use

The orange light will blink whenever the camera is busy taking photographs and when there is other internal activity, such as powering up. The green power light will be on continuously.

In order to explicitly take a picture, press the manual shutter button (the bottom of the two side buttons) at any time. Note that the camera takes several seconds to save the photograph, during which time further presses of the manual shutter will have no effect.

Privacy

The 'do not disturb' button (the top of the two side buttons) allows you to temporarily turn the camera off for a period of about 7 minutes. After pressing this button, the red light will come on. A beep will sound 15 seconds before normal activity is resumed, which happens automatically. To extend the privacy time, press the 'do not disturb' button again. If you press the bottom of the two side buttons before the 7 minutes is up, you can force normal activity to be resumed manually.



Appendix 9 – Modified 3 day travel diary and pocket memory jogger

ID:
Date:

TRAVEL DIARY

A travel diary is a daily record of the trips you take from place to place. A trip is defined as a journey lasting longer than 3 minutes that is to a destination (e.g., to work, school, store, etc).

We are interested in all types of trips: using the car, walking, taking the bus, train, taxi, plane, etc. Use a new line in the diary for each journey to take. For each trip, we would like you record information in each of the columns presented below. Specific notes about how to complete each column are presented on the next page. To make it easier to record your trips during the day we will also provide you with a portable "memory jogger." At the end of the day, you can simply transfer the details from your portable memory jogger into your travel diary. We will not collect your memory jogger diary; this is for your personal use only.

Here is an example of a person's travel diary entries for one day. As you can see, the person went to work by car, but came home by bus. The journey home involved walking to the bus stop (5 min), riding the bus (40 min), and walking from the bus stop home (15 min).

DAY 1: Wednesday		**DATE:** July 6th, 2011			**STAGES** These columns are for entering the details of each stage of your journey					
Journeys Please record each journey using a separate row and remember to tell us about short journeys										Only fill this column if you used a CAR or OTHER MOTOR VEHICLE
A What was the purpose of the journey?	**B** What time did you leave?	**C** What time did you arrive?	**D** Where did you start your journey? (Check home or give the name of the neighbourhood town, or area)	**E** Where did you go to? (Check "home" or give the name of the neighbourhood, town or area)		**F** What method of travel did you use for each stage of your journey?	**G** How far did you travel? (Miles)	**H** How long did you spend travelling? (Minutes)	**I** How many people travelled including you?	**J** Were you the driver (D) or passenger (P)
1. Go to work	Time: 7.25 ☒ am ☐ pm	Time: 7.54 ☒ am ☐ pm	☒ Home	☐ Home Central Oxford	**1**	Car	9.2	29 minutes	1	☐ D ☒ P
					2					☐ D ☐ P
					3					☐ D ☐ P
2. Go home from work	Time: 5.15 ☐ am ☒ pm	Time: 6.35 ☐ am ☒ pm	☐ Home Central Oxford	☒ Home	**1**	walk	0.1	5 minutes	1	☐ D ☐ P
					2	Bus	8	40 minutes	1	☐ D ☐ P
					3	walk	0.25	15 minutes	1	☐ D ☐ P

ID:
Date:

NOTES ABOUT COMPLETING THE TRAVEL DIARY

Column A
What was the purpose of your journey?
Please give a simple description such as ‘go to work’, ‘take children to school’ or ‘go home.’ If you went shopping please note whether it was ‘food shopping’ or ‘other shopping’.

Columns B and C
What time did you leave/arrive?
Write in hours and minutes (e.g. 9.15). Please check am or pm to show the time of day

Columns D and E
Where did you start/go to? (Check ‘Home’ or give the name of the village town or area)
Please note down the name of the place where your journey started and finished. If this was a large town or city give the name of the area. If you went to a shopping centre or visitor attraction please tell us its name. Please be as precise as possible. If your journey started or finished at home, you only need to check ‘Home’.

Column F
What method of travel did you use for each stage of your journey?
Use a different line for the method of travel you used at each stage of your journey (e.g. car, train, bus, bike).

Column G
How far did you travel? (Miles)
Please give us the distance you travelled in miles or yards (e.g. 3 miles, 0.5 miles, 200 yards)

Column H
How long did you spend travelling? (Minutes)
Please note the amount of time you spent travelling and do not include any time spent waiting for public transport.

Column I
How many people travelled including you
Please write in the number of people, including yourself, who set out together. Only include people who were with you for at least half the distance of your journey.

Column J
Were you the driver (D) or passenger (P)?

ID:
Date:

Please check 'D' if you were the driver or 'P' if you were the passenger of the vehicle.

DAY 1:	DATE:				STAGES These columns are for entering the details of each stage of your journey					
Journeys Please record each journey using a separate row and remember to tell us about short journeys. You do not need to record journeys lasting less than 3 minutes.										Only fill this column if you used a CAR or OTHER MOTOR VEHICLE
A What was the purpose of the journey?	**B** What time did you leave?	**C** What time did you arrive?	**D** Where did you start your journey? (Check home or give the name of the neighbourhood town, or area)	**E** Where did you go to? (Check "home" or give the name of the neighbourhood, town or area)		**F** What method of travel did you use for each stage of your journey?	**G** How far did you travel? (Miles)	**H** How long did you spend travelling? (Minutes)	**I** How many people travelled including you?	**J** Were you the driver (D) or passenger (P)
1.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
2.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
3.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
4.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
5.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
6.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
7.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
8.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
USE THIS SPACE FOR ANYTHING ELSE YOU WANT TO TELL US										

ID:
Date:

DAY 2:	DATE:				STAGES These columns are for entering the details of each stage of your journey					
Journeys Please record each journey using a separate row and remember to tell us about short journeys. You do not need to record journeys lasting less than 3 minutes.										Only fill this column if you used a CAR or OTHER MOTOR VEHICLE
A What was the purpose of the journey?	**B** What time did you leave?	**C** What time did you arrive?	**D** Where did you start your journey? (Check home or give the name of the neighbourhood town, or area)	**E** Where did you go to? (Check "home" or give the name of the neighbourhood, town or area)		**F** What method of travel did you use for each stage of your journey?	**G** How far did you travel? (Miles)	**H** How long did you spend travelling? (Minutes)	**I** How many people travelled including you?	**J** Were you the driver (D) or passenger (P)
1.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	1					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
2.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	1					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
3.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	1					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
4.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	1					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
5.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	1					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
6.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	1					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
7.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	1					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
8.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	1					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
USE THIS SPACE FOR ANYTHING ELSE YOU WANT TO TELL US										

ID:
Date:

DAY 3:		DATE:			STAGES These columns are for entering the details of each stage of your journey					
Journeys Please record each journey using a separate row and remember to tell us about short journeys. You do not need to record journeys lasting less than 3 minutes.										Only fill this column if you used a CAR or OTHER MOTOR VEHICLE
A What was the purpose of the journey?	**B** What time did you leave?	**C** What time did you arrive?	**D** Where did you start your journey? (Check home or give the name of the neighbourhood town, or area)	**E** Where did you go to? (Check "home" or give the name of the neighbourhood, town or area)		**F** What method of travel did you use for each stage of your journey?	**G** How far did you travel? (Miles)	**H** How long did you spend travelling? (Minutes)	**I** How many people travelled including you?	**J** Were you the driver (D) or passenger (P)
1.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
2.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
3.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
4.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
5.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
6.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
7.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
8.	Time: ☐ am ☐ pm	Time: ☐ am ☐ pm	☐ Home	☐ Home	**1**					☐ D ☐ P
					2					☐ D ☐ P
					3					☐ D ☐ P
USE THIS SPACE FOR ANYTHING ELSE YOU WANT TO TELL US										

P2520

Area

Address

Household

Per. No.

Thank you very much

Travel Survey

In confidence

MEMORY JOGGER OF

START Day

FINISH Day

Include all journeys by transport (bus, train, car, bike etc). Include walks of 1 mile or more.

Where did you go and purpose of your journey?	When did you leave?	When did you arrive?
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm

Include all journeys by transport (bus, train, car, bike etc). Include walks of 1 mile or more.

Where did you go and purpose of your journey?	When did you leave?	When did you arrive?
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm

Include all journeys by transport (bus, train, car, bike etc). Include walks of 1 mile or more.

Where did you go and purpose of your journey?	When did you leave?	When did you arrive?
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm

Include all journeys by transport (bus, train, car, bike etc). Include walks of 1 mile or more.

Where did you go and purpose of your journey?	When did you leave?	When did you arrive?
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm

Include all journeys by transport (bus, train, car, bike etc). Include walks of 1 mile or more.

Where did you go and purpose of your journey?	When did you leave?	When did you arrive?
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm

Include all journeys by transport (bus, train, car, bike etc). Include walks of 1 mile or more.

Where did you go and purpose of your journey?	When did you leave?	When did you arrive?
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm
	am pm	am pm

Appendix 10 – Participant demographic survey (Oxford) and satisfaction survey

British Heart Foundation Health Promotion Research Group
Department of Public Health
Old Road Campus, Roosevelt Drive, Headington, Oxford OX3 7LF, UK
Tel: +44 (0)1865 289245, Fax: +44 (0) 1865 617789,

Questionnaire for Participants

Project title: **Measurement of travel and sedentary behaviour**

May 2011

Principal Investigator: **Mr Paul Kelly**

Name: ______________________________

Age: ______________________________

Sex: ______________________________

Height: ______________________________

Weight: ______________________________

Occupation: ______________________________

Address: ______________________________

Work address: ______________________________

Telephone number*: ______________________________

Email Address*: ______________________________

*Please note, information will only be used for research purposes and will not be passed on to third parties.

ID:
Date:

SATISFACTION SURVEY

I. Vicon Revue

1a. Please place an **X** by your answer.

	Never a problem	Sometimes a problem	Often a problem	Always a problem
a. Did you have any problems wearing this device?				

1b. Please tell us any comments/concerns you had about wearing the Vicon Revue:

____

____

2. Please tell us about the following potential problems with wearing the Vicon Revue. Please place an **X** by your answer.

	Never a problem	Sometimes a problem	Often a problem	Always a problem
a. Trouble remembering to wear the Vicon Revue				
b. Discomfort from wearing the Vicon Revue				
d. Problem with Vicon Revue slipping with movement				
d. Trouble with the instructions for the Vicon Revue				
e. The amount of time I had to wear the Vicon Revue				

3. Please give us any additional feedback you wish on what your experience with the Vicon Revue was like.

____

____

ID:
Date:

II. General

1. Please tell us about the following potential problems. Please place an **X** by your answer.

	Never a problem	Sometimes a problem	Often a problem	Always a problem
a. Problems with the number of devices you were asked to wear				

	YES	**NO**	
b. Did you use the small pocket travel jogger?			
c. Were you satisfied with the duration of the study visits?			
d. Were you satisfied with the contacts you had with study staff?			
e. Were you comfortable with the information you had to provide as part of this study?			
f. Did you change your walking behavior at all because you were wearing these devices?			If yes, how?
g. Could we have done anything better in the study?			If yes, please explain:

2. If you removed the device at any point, other than to charge it or for privacy reasons, please outline why.
____
____
____

Appendix 11 – Ethical approvals for University of California, San Diego (UCSD) and Auckland University of Technology (AUT)

#071243

University of California, San Diego
Consent to Act as a Research Participant
Pilot Testing Phase IV

Physical Activity and Location Measurement System (PALMS)

Kevin Patrick, MD, MS and Jacqueline Kerr, PhD in the Dept of Family and Preventive Medicine are conducting this research study to develop a Physical Activity and Location Measurement System (PALMS) that collects information on an individual's physical activity behaviors and where they occur. This study is being conducted in collaboration with researchers from Oxford University, UK and San Diego State University.

You are being asked to participate in a research study that lasts for up to 3 days. Before you decide to participate, it is important that you read the following information and ask as many questions as necessary to be sure you understand what you will be asked to do.

The purpose of this study is to understand the types of physical activity you do, as well as where you do these activities. This study is also about the trips that people make by bicycle. You are being asked to participate in the study because you have indicated that you use a bicycle for transportation on a regular basis. There will be approximately 20 people from the San Diego area in this study.

To be eligible for this study you have to meet the following criteria:

- Aged between 18-70 years old
- Regularly commute by bicycle or use a bicycle for routine trips
- Are able to complete a short written survey and briefly review the data collected with UCSD staff, in English

Description of the Study

If you agree to be in this study, the following will happen to you:

1. Wear three small electronic devices for up to three days during waking hours. You will be instructed to wear the devices at all times except when showering, swimming or playing contact sports. The devices you will be asked to wear are:

    a. a small motion-detection device (called an 'accelerometer') worn at the waist that measures the motion of your hips. This device is about the size of a wrist watch.

    b. a location detecting device called a Global Positioning System (GPS). This is a device that records where you travel and is worn on a belt around your waist. The GPS device is about the size of a small cell phone.

   c. a small 'Vicon Revue' device worn on a lanyard around your neck. The Vicon Revue includes a small camera, and sensors that measure light, temperature and movement. The camera takes photographs at regular intervals which helps researchers understand the context in which your activity is occurring (e.g., busy streets, presence of bike lanes, sitting at a desk, etc.) To protect your privacy, you can turn the camera off at any time. Special procedures to protect your privacy when using the Vicon Revue are described in the section on "Confidentiality" below.

2. Complete a bicycling log. Each day you wear the devices you will be asked to write down the start and end times of your bicycling trips, as well the purpose of the trip.

3. Complete a short paper survey and review the images from the Vicon Revue. The survey will ask you questions about you r physical activity and your experience wearing the devices.

You will be asked to meet with our staff at main UCSD campus in La Jolla two times. During the first visit, you will receive instructions on how to wear and charge the three devices. You will be provided with a phone number to call if you have any questions about the study or the devices. During the second visit, you will be asked to complete a short paper survey and be given the opportunity to review and delete any images on the camera you do not wish us to see. Each visit will last approximately 30-60 minutes.

What is Experimental in this Study

Wearing the motion-detection device, the GPS and the Vicon Revue during your daily life and completing the surveys is experimental.

Risks

The risks involved in this study are minimal. Participation in this study may involve some physical discomfort to you from wearing the devices. You may also experience discomfort or a sense of loss of privacy as a result of revealing your location and daily activities for three continuous days to researchers or answering survey questions. You may feel discomfort wearing the Vicon Revue because it takes images of what you are doing throughout the day. You are free to remove the camera at any stage, and there is a privacy button you can press to stop the camera recording for 4 minutes. Also you will be able to review the images and delete any you do not wish researchers to see. You can discontinue participation at any time. There may also be concern for privacy related to divulging personal information or the loss of confidentiality if device data are accessed by an unauthorized person. Your data will be stored appropriately in locked cabinets and on a secure computer server that is password protected in order to minimize the risk of loss of confidentiality. There may also be some unknown risks that are currently unforeseeable. You will be informed of any significant new findings.

Benefits
Benefits of this study are principally to the scientific community to help better understand the accuracy of sensors for measuring human behavior. There may or may not be any direct benefit to you from participation in this study. Potential benefits may include an increased awareness of your physical activity level and pattern and/or your surrounding environment. However, we cannot guarantee that you will receive any benefit from participant in this study.

Payment for Participation
In compensation for your time and travel, you will receive $50 after wearing the devices for 3 days. You will be paid via a preprinted check that can be cashed at any Wells Fargo Bank. No form of ID is required to cash the check.

Care If Harmed
If you are injured as a result of participation in this research, the University of California will provide any medical care needed to treat those injuries. The University of California will not provide any other form of compensation if you are injured. You may call the UCSD Human Research Protections Program office at (858) 455-5050 for more information about this, to inquire about your rights as a research subject, or to report research-related problems.

Questions about the Study
Dr Jacqueline Kerr or ______________________, of the research team has explained this study to you and answered your questions. If you have other questions or research related problems, you may reach Dr. Kerr at (858) 534-9305.

Voluntary Nature of Participation
Participation in this study is entirely voluntary. You may refuse to participate or withdraw at any time. Your decision of whether or not to participate will not jeopardize your future relations with the University of California San Diego or benefits to which you are otherwise entitled.

You may be withdrawn from the study if the Principal Investigator believes that it is in your best interest. You may also be withdrawn from the study if you do not follow the instructions provided by the study personnel.

Alternatives to Participation
The alternative to participating in this research project is to choose not to participate.

Confidentiality
Research records will be kept confidential to the extent provided by law. All responses and data collected from the devices will be kept confidential within

the research team, and no material that could personally identify you will be used in any reports or publications from this study. Your data will be stored appropriately in locked cabinets and on a secure computer server in order to minimize the risk of loss of confidentiality. All information collected for this research that is stored on our secure computer will be identified by subject ID number only. Results of this study may be reported in scientific journals, meeting, and news media. None of these reports will use your name or use data that can point to any person who took part in the study. Strict security measures will be taken to insure that none of these procedures results in release of any information you provide us. You give your permission to the researchers to store your data for use in analyses that have not yet been specifically planned at this time. Your data will be kept for ten years. Researchers conducting these studies will not contact you for any additional information.

Our research team includes collaborators from Oxford University, UK and San Diego State University, who will receive a copy of your data via secure data transmission protocols. No personally identifiable information about you will be released. Your contact information will not be shared. All data (including survey answers and digital images) will be anonymous and stored on a password protected computer.

Several aspects of the procedure are designed to safeguard your rights and privacy as a participant. The Vicon Revue is equipped with two methods to deactivate the camera at any time, should you want your activities to remain private (i.e., certain places, times, or situations). The two methods are:

1) The Vicon Revue has an ON/OFF button for complete deactivation (off) or reactivation (on).
2) The Vicon Revue has a "DO NOT DISTURB" button which will put the camera into "DO NOT DISTURB" mode. In this mode, the camera will remain on but absolutely no images will be captured. The Vicon Revue will remain deactivated for 7 minutes, and will alert you with a beep 15 seconds before it reactivates. You may press the "DO NOT DISTURB" button again to resent the "DO NOT DISTURB" mode to last for another 7 minutes. You may also reactivate the Vicon Revue from "DO NOT DISTURB" mode at any time by pressing the manual shutter button.

In addition, should you want any images to be deleted that were inadvertently captured (e.g. during a time when you would have preferred for the Vicon Revue to be deactivated but forgot to deactivate it), you will be able to note the time period in a small notepad which we will provide to you. As soon as you return to the measurement office you will be given the opportunity to review and delete any images you wish prior to them being seen by research staff.

Furthermore, several aspects of the procedure are designed to ensure the rights and privacy of other people who may appear in the images captured by Vicon Revue. These include:

1) A reference card will be provided for you to carry around while wearing the device. This includes:
    a. a prepared statement to read to anyone with questions or concerns about the Vicon Revue; "I am participating in an experiment on physical activity and the environment. This is a digital camera that automatically captures low-resolution still images throughout the day, which will later be used to describe my behavior and environment. It does not record audio or full-motion video. Any images captured will not be made public in any fashion and will only be seen by the researchers. If you would prefer, I can turn off or temporarily deactivate the camera, and/or make a note and have the images just taken deleted without anyone seeing them. I can also provide contact information for the researchers."
    b. a list of places and situation in which we ask you to deactivate the Vicon Revue
        i. Place/Situations:
            1. Any restroom
            2. Any changing room, locker room, etc.
            3. Doctor's office
            4. ATM or bank
            5. Wherever/Whenever you would prefer for images not to be captured
            6. Wherever/Whenever anyone requests deactivating
    c. Contact information for the investigators (email, address and phone number)

2) A small notepad in which to note periods of time for image that you want deleted.

<u>Procedures for Termination</u>
If you decide to withdraw from this study, you will be required to notify a member of the research team and return the PALMS devices if they are still in your possession.

<u>Sponsor Inspection</u>
The sponsor of this study, the National Institutes of Health, National Cancer Institute, will have access to general research data but will not have the right to access patient medical records or clinical charts.

Research records may be reviewed by the UCSD Institutional Review Board.

<u>Agreement</u>

You have received a copy of this consent document to keep and the *"Experimental Subjects Bill of Rights."* You agree to participate.

____ _______________

Signature Date

____ _______________

Witness Date

UCSD Human Research Protections Program
UCSD
Approved
Initial Approval: 09/22/2007
Current Approval: 06/28/2011
Do not use after: 06/21/2012

MEMORANDUM

Auckland University of Technology Ethics Committee (AUTEC)

To: Melody Oliver
From: **Dr Rosemary Godbold and Madeline Banda** Executive Secretary, AUTEC
Date: 25 May 2011
Subject: Ethics Application Number 11/114 **Characterising environmental contexts for physical activity: An assessment of objective and self report methods.**

Dear Melody

We are pleased to advise that the Auckland University of Technology Ethics Committee (AUTEC) approved your ethics application at their meeting on 9 May 2011. Your application is now approved for a period of three years until 9 May 2014.

AUTEC commends you on the quality of the Information Sheet.

We advise that as part of the ethics approval process, you are required to submit to AUTEC the following:

- A brief annual progress report using form EA2, which is available online through http://www.aut.ac.nz/research/research-ethics/ethics. When necessary this form may also be used to request an extension of the approval at least one month prior to its expiry on 9 May 2014;
- A brief report on the status of the project using form EA3, which is available online through http://www.aut.ac.nz/research/research-ethics/ethics. This report is to be submitted either when the approval expires on 9 May 2014 or on completion of the project, whichever comes sooner;

It is a condition of approval that AUTEC is notified of any adverse events or if the research does not commence. AUTEC approval needs to be sought for any alteration to the research, including any alteration of or addition to any documents that are provided to participants. You are reminded that, as applicant, you are responsible for ensuring that research undertaken under this approval occurs within the parameters outlined in the approved application.

Please note that AUTEC grants ethical approval only. If you require management approval from an institution or organisation for your research, then you will need to make the arrangements necessary to obtain this. Also, if your research is undertaken within a jurisdiction outside New Zealand, you will need to make the arrangements necessary to meet the legal and ethical requirements that apply within that jurisdiction.

When communicating with us about this application, we ask that you use the application number and study title to enable us to provide you with prompt service. Should you have any further enquiries regarding this matter, you are welcome to contact Charles Grinter, Ethics Coordinator, by email at ethics@aut.ac.nz or by telephone on 921 9999 at extension 8860.

On behalf of the AUTEC and ourselves, we wish you success with your research and look forward to reading about it in your reports.

Yours sincerely

Dr Rosemary Godbold and Madeline Banda
Executive Secretary
Auckland University of Technology Ethics Committee

From the desk of …
Dr Rosemary Godbold and Madeline Banda
Executive Secretary
AUTEC

Private Bag 92006, Auckland 1142
New Zealand
E-mail: ethics@aut.ac.nz

Tel: 64 9 921 9999
ext 8860
Fax: 64 9 921 9925

Appendix 12 – Other SenseCam publications I have co-authored and conference presentations I have given

SenseCam publications I have co-authored

1. L Gemming, A Doherty, **P Kelly**, J Utter, C Ni Mhurchu; Self-reported dietary intake using passive image capture: a pilot study; International Journal of Behavioural Nutrition and Physical Activity; 2013 (Under submission);
2. Oliver M, Doherty AR, **Kelly P**, Badland HM, Mavoa S, Shepherd J, J Kerr, S Marshall, A Hamilton and C Foster. Utility of passive photography to objectively audit built environment features of active transport journeys: an observational study. International Journal of Health Geographics 2013;12(1):20;
3. Kerr J, Marshall SJ, Godbole S, Chen J, Legge A, Doherty AR, **P Kelly**, M Oliver, H Badland, C Foster. Using the SenseCam to Improve Classifications of Sedentary Behavior in Free-Living Settings. American Journal of Preventive Medicine 2013;44(3):290-296;
4. Doherty AR, **Kelly P**, Kerr J, Marshall S, Oliver M, Badland H, A Hamilton, C Foster. Using wearable cameras to categorise type and context of accelerometer-identified episodes of physical activity. International Journal of Behavioral Nutrition and Physical Activity 2013;10(1):22;
5. Doherty AR, Hodges SE, King AC, Smeaton AF, Berry E, Moulin CJ, S Lindley, **P Kelly**, C Foster. Wearable cameras in health: the state of the art and future possibilities. American Journal of Preventive Medicine 2013;44(3):320-323;
6. Doherty A, **Kelly P**, Foster C. Wearable Cameras: Identifying Healthy Transportation Choices. Pervasive Computing, IEEE 2013;12(1):44-47;
7. Doherty A, Marshall S, **Kelly P**, Hamilton A, Oliver M, Badland H, Kerr L, Foster C. Identifying sedentary behaviour types using SenseCam: A pilot study. Journal of Science and Medicine in Sport 2012;15:S296-S297;
8. Doherty A, **Kelly P**, Oliver M, Hamilton A, Badland H, Marshall S, et al. Using SenseCam to categorise type and context of accelerometer-identified episodes. Journal of Science and Medicine in Sport 2012;15:S92-S93;

9. Doherty AR, **Kelly P**, Kerr J, Marshall S, Oliver M, Badland H, Foster C. Use of wearable cameras to assess population physical activity behaviours: an observational study. The Lancet 2012;380:S35.

Conference and seminar presentations I have given

1. Innovative technologies to study environment-health behaviour relationships; Use of digital life-logging to study environment health relationships; 13th annual meeting of the International Society for Behavioural Nutrition and Physical Activity (ISBNPA); **Ghent, Belgium; May 2013**
2. International Conference on Diet and Activity Methods (ICDAM); Can we use digital life-log images to investigate the error on self-reported journey behaviour in school children? Results from a pilot study; Rome, Italy; May 2012
3. Seminar Presentation at Department of Family and Preventive Medicine, UC San Diego; **San Diego, USA; July 2011**
4. Seminar presentation at School of Public Health, Auckland University; **Auckland, New Zealand; June 2011**
5. Innovative research in the fields of public health; ISBNPA satellite meeting; 'Beyond 2011: New Paradigms to Improve Physical Activity and Nutrition'; **Queenstown, New Zealand; June 2011**
6. Emerging Technologies for Assessing Physical Activity: Using Digital Images to Investigate Active Travel Behaviour; 10th annual meeting of the International Society for Behavioural Nutrition and Physical Activity (ISBNPA); **Melbourne, Australia; June 2011**
7. Invited speaker at the ASO (Association for the Study of Obesity) annual conference. "The use of objective assessment of physical activity and nutrition behaviour in the fight against obesity"; **London, UK; April 2011**

8. Invited presentation at Unit for Bio-cultural Variation and Obesity (interdisciplinary research unit based at the University of Oxford). “Digital Image Capture in Public Health Surveillance for Physical Activity and Food Behaviour Assessment”; **Oxford, UK; March 2011**
9. Presentation at UKTRC Transport, Physical Activity and Health Group at UCL, funded by Department for Transport; **London, UK; February, 2011**
10. Using digital images to estimate the error on self-reported travel behaviour; Sixth International Conference in Movement and Health; **Olomouc, Czech Republic; November 2010**
11. Seminar Presentation at National Centre for Social Research; **London, UK; September 2010**
12. Investigating Movement and Sedentary Behaviour using SenseCam; Second Annual SenseCam Symposium; **Dublin, Ireland; September 2010**