Improving our understanding of the influences of the environment on physical activity

Lindsey Gail Smith

Wolfson College

UNIVERSITY OF CAMBRIDGE

This thesis is submitted for the degree of Doctor of Philosophy

Physical Activity and Public Health Group, MRC Epidemiology Unit & Centre for Diet and Activity Research, School of Clinical Medicine, University of Cambridge

January 2020

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the statement outlining my contribution, stated in the acknowledgements, and specified in the text.

It is not substantially the same as any work that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared and specified in the text.

This thesis does not exceed the word limit of 60 000 words (excluding figures, photographs, tables, appendices and bibliography) as prescribed by the School of Clinical Medicine Degree Committee.

Lindsey Smith

23 January 2020

Abstract: Improving our understanding of the influences of the environment on physical activity

Lindsey Gail Smith

Physical activity reduces the risk of many chronic diseases, and much of the population would benefit from being more active. Public health strategies increasingly identify the environment as an important influence on activity and therefore a target for intervention. However, it is difficult to draw conclusive evidence from the existing literature as the range of environments and spaces individuals are exposed to are rarely accounted for. This thesis aims to further develop our understanding of the ways in which environmental characteristics influence physical activity.

The first part examines cross-sectional associations between environmental characteristics such as street connectivity, air pollution, and deprivation, and self-reported and objective measures of physical activity. Data are used from the national multicentre UK Biobank study. The findings suggest environmental characteristics have the potential to contribute to different physical activities but interventions which focus on a single environmental attribute may not have the greatest benefits. The UK Biobank study uses measures of environments around residential addresses which may not capture all locations where participants are active.

The second part of my thesis provides a more representative picture of environments that people are exposed to by focusing on activity spaces: locations accessed by an individual as a result of their daily activities. I systematically review literature which uses the activity space concept, and discuss research questions that have been answered, the spatial and temporal methods used, and the implications for causal inference. Included studies used variable methods to assess the features of spaces themselves (such as shape or size) or features within spaces (such as density of destinations).

Informed by the conceptual work of the review, the latter section of the thesis uses quantitative and qualitative data from the Commuting and Health in Cambridge study to understand how new transport infrastructure might give rise to changes in use of space. The development of a replicable process to clean and prepare GPS data is presented and findings show how the infrastructure provides a new space for physical activity. The final project explores the applicability of different spatial methods for assessing population levels of activity and how changes in the location of physical activity might contribute to overall levels of activity over time.

By developing and applying scalable methods to show how the spatial patterning of behaviour and physical activity changes in the context of an intervention, this thesis provides methodological and scientific contributions to the field of physical activity and public health. Future research in this topic area should aim to strengthen the basis for causal inference and develop evidence to effectively inform public health policy and action.

Acknowledgements

I am extremely grateful to my supervisors, Jenna Panter, Thomas Burgoine and David Ogilvie. All have been consistently supportive and provided thoughtful and insightful input throughout my PhD. Thanks especially to Jenna for being patient, kind, and for creating an encouraging environment for me to grow as a scientist. Under Jenna's supervision I have learnt to appreciate the joy of working on innovative projects I truly care about and the importance of collaboration and constructive mentorship in producing high quality research.

I am very grateful to the MRC Epidemiology Unit for funding my PhD and to my colleagues within the Unit who provided advice and assistance. In particular, I'd like to thank Louise Foley for helping me to navigate the activity space literature and for guiding me around a conceptually challenging topic. I have been fortunate to work alongside a cohort of brilliant PhD students whom I thank for their challenges and perspectives and for inspiring me through their exceptional research projects over the last 3 years. Thanks also to Simon Griffin, and to Andy Jones and Emma Coombes from the UEA, who recommended the MPhil and PhD programme at the Unit and who supported my application back in 2014.

I am thankful to my Mum, Nicky, whose support and steadfast calmness has fuelled a determination in me, and to my Dad, Bob, for knowing when a screen break and a pint in the Cumbrian fells was in order.

Lastly, I am grateful for my friends, old and new, who provided encouragement along the way. A special thanks to Steph and Richard for their unwavering generosity; providing a place to stay and hearty packed lunches for the long days working in Cambridge. I also thank Kathrine, Sarah, Rachel, and Kelly for always listening, and Al who brought brightness during the busy days of thesis writing.

Funding

This work was supported by the Medical Research Council [Unit Program number MC_UP_12015/6]. The work was also supported under the auspices of the Centre for Diet and Activity Research (CEDAR), a UKCRC Public Health Research Centre of Excellence at the University of Cambridge, for which funding from the British Hearth Foundation, Economic and Social Research Council, Medical Research Council, National Institute for Health Research and the Wellcome Trust, under the auspices of the United Kingdom Clinical Research Collaboration, is gratefully acknowledged.

Contributions

The general topic for the PhD was conceived following discussions with Andy Jones and Emma Coombes, and was built on key gaps in the literature identified by Andy and Emma, and areas of personal interest and expertise. The final proposal was drawn up by myself, and further refined with assistance from Jenna Panter.

The analysis reported in Chapter 2 was planned and undertaken by me, following advice and guidance from Jenna Panter and David Ogilvie. Katrien Wijndale and Tom White provided advice on processing and defining measures of physical activity and Marco Taino provided guidance regarding measures of air pollution. I wrote the paper, receiving critical feedback from Jenna Panter and David Ogilvie. I was not involved in any collecting of the data by UK Biobank and received assistance in submitting an application to access the data from Jenna Panter.

I had planned to use the concept of the activity space when submitting my PhD proposal and the idea to perform the systematic review in Chapter 3 was conceived following discussions with Jenna Panter, Esther van Sluijs and David Ogilvie. A substantial amount of planning was conducted with advice from Jenna Panter and Louise Foley. I performed the searches for the review, the screening of articles and the data extraction. Louise Foley independently screened 20% of the articles identified for full text review and where there were any discrepancies in agreement relating to reasons for exclusion, articles were referred to Jenna Panter for a majority decision. Fumiaki Immamura assisted with the translation of one identified article into English. I wrote the final paper, receiving feedback from Louise Foley, Jenna Panter and Thomas Burgoine.

The idea to use GPS data, physical activity data, and qualitative data to investigate changes in activity spaces and locations of physical activity arose following discussions between myself, Jenna Panter and Thomas Burgoine. Analysis plans for Chapter 4, Chapter 5 and Chapter 6 were written by me, receiving input from Jenna Panter, Thomas Burgoine, and Andy Jones, and were circulated to the Principal Investigators on the Commuting and Health in Cambridge study by David Ogilvie. I wrote the Python code to clean and prepare the GPS data (Chapter 4) and Susie Boatman was responsible for matching the GPS data to questionnaire data to maintain participant confidentiality. Emma Coombes provided assistance matching GPS and physical activity data and Stefanie Hollidge advised on measures of physical activity. I performed the analyses and wrote the text presented in Chapter 4, Chapter 5 and Chapter 6, receiving feedback from Jenna Panter, Thomas Burgoine and Louise Foley. I was not involved in the design of the Commuting and Health in Cambridge study nor collecting of the data.

List of abbreviations

BMI	Body mass index
CI	Confidence interval
CIR	Colour infrared
DPA	Daily path area
DTM	Digital terrain model
e.g	Exempli gratia or “for example”
EPIC	European Prospective Investigation into Cancer and Nutrition
GIS	Geographic information system
GPS	Global positioning system
HDOP	Horizontal dilution of precision
HHI	Herfindahl Hirschman index
i.e	Id est or “That is”
IPAQ	International Physical Activity Questionnaire
KDE	Kernel density estimation
KHz	Kilohertz
km	Kilometres
km/h	Kilometres per hour
LUR	Land use regression
m	Metres
MCP	Minimum convex polygon
MET	Metabolic equivalent
m*g*	Milli-gravity units
MVPA	Moderate to vigorous physical activity
n	Number (used to indicate number of participants in a study)
n.i	Not included in regression model
n.s	Not significant
NDVI	Normalised deviation vegetation index
NHS	National Health Service
NO_X	Nitrogen oxides
np	No parameters given

NO_2	Nitrogen dioxide
OS	Ordnance Survey
p	p-value
$PM_{2.5}$	Particulate matter with aerodynamic diameter ≤ 2.5 µm
PROSPERO	International Prospective Register for Systematic Reviews
RCT	Randomised control trial
RRR	Relative risk ratio
SDE	Standard deviational ellipse
UGCP	Uncertain geographic context problem
UK	United Kingdom
UKBUMP	UK Biobank Urban Morphometric Platform
†	Indicates significant ($p < 0.05$)
*	Indicates significant ($p < 0.01$)
**	Indicates significant ($p < 0.001$)
<	Less than
>	Greater than
≤	Less than or equal to
≥	Greater than or equal to
ß	Regression coefficient
µm	Microns
μgm^{-3}	Micrograms per cubic metre

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Physical activity and health

Engaging in regular physical activity has been associated with a reduced risk of chronic and preventable diseases including cardiovascular disease, type 2 diabetes, some cancers [1], and mental health conditions such as clinical depression [2]. Physical activity also plays a fundamental role in the balance of energy expenditure and weight control and can improve cardiorespiratory fitness as well as musculoskeletal health [3].

To attain significant health benefits, it is recommended that adults aged between 18 and 64 years accumulate at least 150 minutes of moderate-intensity aerobic activity per week, supported by weight-bearing activities on two or more days [4]. Moderate-intensity activities include walking and cycling performed in bouts of at least 10 minutes. As well as benefiting health, displacing car use through walking or cycling could foster social interactions, promote social equity [5], [6], and contribute to more environmentally sustainable communities through reduced traffic congestion and carbon emissions [7], [8].

However, many adults do not achieve the recommended levels of activity. Globally, around 23% of adults were physically inactive in 2010 with physical inactivity reported as the fourth leading cause of global mortality [9], [10]. In the UK, the proportion of adults who were physically inactive was comparatively high at 37% [10], affecting the general health of the population and contributing to the burden of chronic diseases. It was estimated that a lack of physical activity cost the NHS £1.06 billion in 2011 through treatment of inactivity-related diseases with a further cost of £6.5 million to the national economy in England due to loss of productivity [11].

Levels of physical activity have declined in the past 20 years. Technological advances have prompted changes in occupational, transport-related and recreational behaviours with a shift towards non-manual jobs and passive modes of travel. Estimates predict a further 15% reduction of activity in the UK by 2030 [12]. Similar effects are being replicated in developing countries with populations spending more time sedentary and less time active. This continued trend in inactivity and its associated health burden indicate its importance as a public health issue and provides a strong case for directing interventions to target the promotion of sustained physical activity.

1.2 A public health perspective on physical activity

Physical activity is undertaken in different domains of life including at home, at work, in transport, and in leisure time. It comprises a range of health behaviours including activities for utilitarian purposes, such as walking or cycling for transport, and leisure activities, such as sports participation or structured exercise. The intensity and time spent active varies across domains and behaviours, and each have their own set of determinants which interrelate [13], [14].

The promotion of physical activity is a recognised public health priority and a number of strategies have been proposed to drive changes in activity behaviours. Early research focused on physical activity as a lifestyle target for preventing coronary heart disease and improving cardiovascular health [15]. Traditional high risk strategies therefore target those who are least active and at greatest risk of developing, or have developed, chronic disease [16]. These typically involve individual-level clinical interventions which focus on cognitive components of health behaviour, including attitudes and personal views, to advise, educate or motivate people to adopt an active lifestyle [17]. Whilst shown to be effective for some, these interventions do not address widespread risks beyond the control of the individual and require a high level of agency from individuals if they are to benefit [18], [19]. These are likely to be less effective than those that require less agency [20], [21]. The long-term sustainability and applicability of these targeted approaches for all individuals therefore remains in question [16], [18], [19].

Rose's population approach focuses on prevention and interventions that target whole populations without identifying high risk individuals [22], [23]. Population strategies complement individual-level approaches by recognising that broader economic, environmental, and social factors can influence behaviours, alongside individual lifestyle factors, and consequently health inequalities [24]. As there is a high prevalence of physical inactivity globally, population approaches appear well-suited to promoting physical activity. By aiming to change many people's behaviour by a small amount, these may have a larger impact on society's inactivity-related health outcomes [23]. For example, an increase of only a few minutes in physical activity at the individual level translates to a large change in levels of activity at the population level.

Some population interventions target groups of individuals to encourage specific activities such as increasing access to facilities for exercise [25]. However, this only addresses a small component of physical activity for some of the population and may not be accessible or appealing to many due to economic and time costs. Focusing on specific types of activity or exercises undertaken in one place may not therefore be the most beneficial for increasing global levels of physical activity [26]. Alternative interventions target whole communities to incorporate physical activity into everyday living through changes to the environment in which people live [26]–[29]. These have the potential to achieve larger and more sustainable changes by focusing on determinants of behaviour for whole populations [30]. Consequently, understanding how broader factors influence behaviour has become a key focus in physical activity and population health literature.

1.3 The environment and physical activity

Socio-ecological models suggest that the environment and social context in which people live is related to health [31], [32]. The environment encompasses the physical urban form, natural elements, economic conditions as well as societal norms.

Although there are many determinants of physical activity, including natural and economic environments which are difficult to change, some elements of the physical urban form are modifiable and may inhibit or encourage habitual behaviour at scale. For example, modifying urban environments to create better connected neighbourhoods with more appealing infrastructure for walking and cycling may lead to increases in active over passive modes of travel at the population level. Structural interventions that allow for people to live in environments more conducive to healthy behaviours therefore have the potential to effectively and equitably promote physical activity [25], [33].

The impetus to promote physical activity through environmental changes is further supported in the wider health literature. Encouraging active lifestyles is widely acknowledged as a means to reduce morbidity and mortality [34] and is important for mental and social well-being [35], [36]. The location of physical activity and quality of environment in which it occurs further plays an important role for related health outcomes. For example, outdoor activities and physical activity undertaken in nature have been associated with enhanced mental health [37], [38] , which has been shown to have a bidirectional association with physical activity [39], [40].

Lack of crime and the presence of greenness have been associated with longevity and the reduced risk of type 2 diabetes where physical activity is hypothesised to be a mediating factor [41], [42]. This is particularly the case in densely populated urban areas where parks and other green spaces are increasingly viewed by policymakers and planners as a lever for promoting healthier, more active lifestyles. Additionally, the co-benefits of substituting passive for active modes of travel in relation to reduced traffic and air and noise pollution are well documented and have been associated with cardiorespiratory health and premature mortality [43], [44]. The environments which individuals interact with whilst participating in physical activity therefore have the potential to amplify or diminish the beneficial effects on physical and mental health outcomes brought about by physical activity.

Changes to the built environment are increasingly being recognised in policy and practice guidelines as important levers for increasing physical activity [25], [33]. In the Lancet's series on urban design, transport, and health, the importance of integrating the reduced demand for driving relative to active modes of travel into urban planning policies is highlighted [45]. In order to deliver policies effectively, there is a need for a strong evidence base to understand environmental effects on physical activity and which types of built environment interventions are most and least effective for whom and where, and why [46].

1.3.1 Strengthening the evidence base

The World Health Organisation's Global Action Plan on Physical activity includes 'active environments' as a strategic objective for tackling inactivity within populations [47], recognising changes to the built environment as a means to promote population levels of physical activity. Such socioecological approaches go beyond the high agency, individual approaches outlined in Section 1.2 by addressing underlying risks and placing drivers of physical activity in their social and environmental context. However, due to a lack of progress in promoting physical activity, there is a need to move beyond the traditional linear models of cause and effect that have underpinned much of the existing evidence base to date [48].

Systems approaches are based on the theory that individuals and variables do not operate in isolation [49]. They build on socioecological approaches when conceptualising health outcomes and inequalities by accounting for connections between factors and the ways in which actors interact with them [50]. Systems are defined by more than a sum of interdependent factors. Systems comprise feedback loops and adaptation, whereby changes to a system reinforce further changes and responses to adjustments in behaviour. For example, the prioritisation of cycle lanes increases the safety and convenience of cycling, making it a more viable travel option. An uptake of cycling in response may lead to a provision of cycle parks which further increases the convenience of cycling, encouraging the prioritisation of more cycle lanes in a reinforcing loop.

From a systems perspective, physical inactivity and associated comorbidities have emerged as a product multiple factors and consequently cannot be solved with a single short-term solution. Instead, evidence should account for multidimensional circumstances in which people live and interventions are implemented [50], [51]. While multilevel methods account for factors at multiple levels to some extent, they do little to improve understanding of the most plausible and modifiable determinants of physical activity, how and why actors interact with interventions, and potential permeations of behaviour change for feedback loops and adaptations. Modern concepts of causality and strong research designs, including the use of multi-method research can shed light on potential associations and feedback mechanisms within the system [52]–[54]. Drawing on this evidence, actions must be wide ranging across multiple sectors and policies coordinated in order to create effective shifts within interacting factors within the systems and to maximise potential change [50], [55].

1.3.2 Current evidence

Several reviews have summarised the evidence on the relationship between the environment and different physical activity outcomes for children and adolescents [56], [57], adults [58]–[62], and older adults [63], [64]. Although findings were generally mixed across reviews, some consistent observations have been made. Repeated findings were shown for neighbourhoods with greater walkability, land use mix, and residential density being supportive of physical activity [57]–[59], [64]. Associations were often shown for specific activity domains or behaviours, rather than overall levels of physical activity [59]–[61], which suggests that some environments are associated with certain activities only. Associations also varied for different populations, whether geographically or by age, which may indicate contextual and lifestyle factors influence the causal pathway. Van Holle and colleagues suggest that other contextual factors such as the quality of the environment may moderate observed associations [59].

The reviews highlight the dominance of cross-sectional study designs in the field. Most reviews include studies from different locations worldwide, however, the majority of studies have been undertaken in North America or Australia [58], [60]–[62] with one review collating studies from European settings [59]. The majority of studies included for review use sample sizes of less than 4000 [57], [59], [62], [64] and focus on a single city or region [63].

Although perceived measures of the environment are often used, an increasing number of studies are using objective measures to assess environmental exposures with some using both. Objective measures include quantitative representations of the neighbourhood environment using Geographic Information Systems (GIS) and audit data [58], [63]. Environmental characteristics have predominantly been quantified within residential neighbourhoods with the size and definition of the neighbourhood varying across studies. Neighbourhood delineations have been defined by administrative boundaries, radial and network buffers of 400 m to 1.6 km, and perceived areas within a 10-15 minute walk of the home address [58], [63]. The environmental characteristics included for analysis tend to be of micro-level, assessing features such as recreational facilities or walkable features [56], [59], with few studies accounting for a range of environmental features, wider sociodemographic characteristics and potential moderating factors in the analysis [61], [63].

Theories used in health geography have seen a shift away from understanding environmental influences using static measures, such as characteristics of the neighbourhood or around the workplace, toward more dynamic conceptualisations of space [65]. Coupled with recent technological advances and the availability of high precision location data, such as Global Positioning System (GPS) data, more specific measures of environmental exposures and the spatial context of physical activity are increasingly being applied in health studies [66].

Linking location and physical activity data allows for the spatial and temporal context of an individual's movement and activity to be measured more accurately [67], [68]. This has given rise to a range of new data collection methods, approaches to analysis, and research questions that can be explored [68]. For example, a recent review by Yi and colleagues identified a number of descriptive studies that combine GPS and accelerometer data to identify environments or domains where the greatest amount or intensity of physical activity occurs [69]. The concept of the activity space, whereby a set of spatial locations visited by an individual are assessed, is an emerging approach being used in studies of the environment and physical activity [70]. However, there is little consistency across the design of these studies. The different research questions that may be answered, the methods used to delineate activity spaces, and what their use means for causality has also received little attention.

A range of measures and categorisations of physical activity have been employed across studies to capture global levels of physical activity as well specific intensities, domains, and behaviours. Assessments of activity are largely self-reported [57], [58], [63], with little standardisation across measures [60], [62]. Although accelerometer and pedometer assessments of movement and step counts have been used to assess physical activity [56], their use to investigate associations with environmental exposures has only recently begun to be reported. Furthermore, few studies investigate a range of activity outcomes or complement objective measures with self-reported domains and behaviours [59].

1.3.3 Limitations of current evidence

A number of limitations make it difficult to draw conclusive evidence on the relationship between the environment and physical activity from the reviews. Limitations relate to study design and the assessment of exposures and outcomes.

1.3.3.1 Setting and study design

Small sample sizes and limited geographical and environmental heterogeneity limit the ability to draw generalisable conclusions from datasets. A lack of geographic coverage further limits the generalisability of findings outside of North America and urban landscapes therein, due to differences in environments and behaviours across settings. Large nationwide studies with a mix of environments and population groups may therefore be an important contribution to the field to provide more transferrable findings.

Cross-sectional studies indicate associations between environmental characteristics and physical activity where people might be more likely to be active in more supportive environments. However, it is not possible to infer causality from these studies and there is a risk of reverse causation. Randomised control trials (RCTs), whereby individuals are randomly assigned to one of two groups with one group exposed to the intervention, may strengthen the basis for causality as they are less susceptible to bias and confounding. However, exposure to built environment interventions may be impractical or unethical to manipulate [71]. McCormack and colleagues extend prior findings by assessing quasi-experimental studies that assess activity behaviours before and after relocation to a new neighbourhood [58]. However, a number of factors are associated with moving and intentions to adopt different behaviours may be tied to reasons for relocating rather than the change of environment. Investigating a change in the environment where people have not moved, or made any other key changes to their lifestyle, provides a potentially more powerful study design.

Natural experiments, whereby exposure to an intervention cannot be randomised, widen the range of interventions that can usefully be evaluated and may further provide more robust evidence of causality than cross-sectional non-intervention studies [71]. They enable investigation into whether and how changing the built environment changes physical activity and behaviours. However, the application and assessment of such studies is in its infancy and careful consideration is required when designing methods to implement studies as effects may take time to emerge [28]. Consequently, there is a need for longitudinal study designs and studies which effectively evaluate environmental interventions and their influence on physical activity to help inform neighbourhood design and planning strategies to improve public health.

1.3.3.2 Assessment of exposures

A range of environmental characteristics have been investigated and different data collection procedures and measures have been used. While objective measures overcome issues of recall bias and differences in perceptions of space across individuals, they are more difficult and costly to incorporate into large scale population studies. Spatially referenced measures in GIS may also be quantified in a myriad of ways such as density of features and relative coverage of land uses. This makes it difficult to compare findings across studies and to identify consistent patterns of association from the current evidence base.

Much of the literature focuses on micro-level aspects of the physical built environment such as proximal characteristics including walkability and land use mix. Applying a public health perspective and reflecting on the growing recognition of broader determinants of health indicates that a range of wider social and environmental attributes such as deprivation, rurality and pollution are important [24], [50], [72]. Studies have typically focused on a single characteristic of the environment, and whilst this is useful for identifying associations with more specific physical activity outcomes (such as walkability and walking), environmental characteristics coexist and interrelate. To limit confounding and to identify moderating factors such as the quality of walkable neighbourhoods and greenspaces, studies that examine a range of environments of different scales may be appropriate.

Environmental measures have also been limited to the neighbourhood environment and typically rely on static measures of exposure using inconsistent spatial units. The Uncertain Geographic Context Problem (UGCP), described by Kwan and colleagues [73], highlights the need to represent the spatial area in which behaviours occur in order to match specific environments and physical activity behaviours. If the measure of exposure differs from the causally relevant context for an outcome, inferential errors may arise through misclassification of the environments to which an individual is exposed [74], [75]. Assessing only the residential environment and total physical activity, for example, may be inappropriate because much physical activity occurs away from home and measuring environments that have little or no bearing on physical activity may bias findings or dilute the true effect [76], [77]. The increasing use of location-specific data and application of the activity space addresses this by providing a more representative measure of where people spend time and more specific information about environmental contexts of activity [68], [70]. However, due to the novelty of their application within the field, there are a number of technical limitations associated with spatially referenced data [67] and the role of the activity space in strengthening the basis for causal inference has not yet been formally reviewed.

1.3.3.3 Assessment of outcomes

The majority of evidence relies on self-reported outcomes of physical activity. Whilst this is useful for providing information on domain and behaviour-specific activity, data are often subject to social desirability and recall bias. The use of objective measures allows for greater confidence in the validity of the assessment but current studies using objective measures largely use small samples due to financial and logistical constraints.

Studies which have combined objective physical activity measures with GPS data have typically focused on the spatial element of activity, describing the quantity of physical activity across locations [69]. However, the capabilities of linked GPS and physical activity data extend beyond this. The data allow for environments to be weighted by time spent active in them and for changes in spatial and temporal patterning of movement and physical activity in response to a built environment intervention to be assessed [67], [68]. More testing and investigation into appropriate space-time modelling of physical activity is required as well as thoughtful consideration into how the use of GPS and physical activity data can strengthen the basis for causal inference. Combined with self-reported and qualitative information, it may be possible to shed light on how and why spaces are used for different behaviours.

In summary, there is uncertainty in the existing literature on environment-physical activity relationships as the range of environments and spaces individuals are exposed to have rarely been accounted for. Key limitations relate to a narrow focus on environments and physical activity outcomes, assessing a single characteristic or outcome in isolation, and the use of static neighbourhood measures of the environment. The concept of the activity space is increasingly being used as a more dynamic measure of environmental context of activity. However, its application is associated with conceptual issues which have implications for causality. Despite the emergence of studies which combine GPS and objective physical activity data, there are no standard tools for processing data or recommended methods for analysis and few longitudinal, qualitative and intervention studies.

1.4 Thesis overview

1.4.1 Aims and scope

This thesis aims to address the gaps in the literature highlighted in the previous section in order to further develop our understanding of the ways in which environmental characteristics influence physical activity. The following aims form the key components of the thesis:

a) To investigate a broad range of environmental characteristics in combination and their relationship with different measures of physical activity and behaviours;

b) To develop an understanding of the role of the activity space in studies of the environment and physical activity and its implications for causality;

c) To test and develop replicable data cleaning processes for GPS data;

d) To implement the activity space concept using a combination of qualitative and quantitative data to investigate the effect of a built environment intervention on the spatial patterning of movement and physical activity.

The thesis aims to provide methodological and scientific contributions to the field of physical activity and public health using data collected from two independent samples of adults in the UK. I restricted analyses to samples of adults as they form the largest proportion of the population and their movement, travel and activity patterns are generally more autonomous and sensitive to change than other population groups such as children and older adults.

1.4.2 Data used in thesis

1.4.2.1 UK Biobank

UK Biobank is a population-based prospective cohort study of 503 317 participants recruited in the UK between 2006 and 2010. The study was established primarily to investigate the genetic and lifestyle determinants of a range of diseases in middle and older aged adults. The study design and survey methods are described in detail elsewhere [78].

Individuals aged between 40 and 69 years were invited to participate if they lived within 35 km of one of 22 assessment centres located throughout the UK. Extensive questionnaire data, physical measurements, and biological samples were collected at recruitment between 2006 and 2010. Additional data have been and continue to be collected for large sub-samples of the cohort. Between 2013 and 2015, 236 519 participants were invited to partake in objective physical activity monitoring [79]. Those who agreed to participate wore a wrist-worn accelerometer for 7 days.

Objective data characterising environmental conditions around participants' home addresses are available for over 70% of the full dataset. Measures were derived for characteristics considered to be important for physical activity, diet, alcohol consumption and general health. The processes and available measures have been previously published [80], [81].

Most previous research in the field has relied on small samples, self-reported measures of activity or comparatively narrow conceptions of environmental exposures. Key strengths of the dataset include its large sample size, heterogeneity in geographical locations across the UK, and generalisability of risk factor associations which are concordant with other nationwide cohort studies [82]. Diverse measures of environmental characteristics and physical activity outcomes for each individual provide additional strengths. Accelerometer data allow for the volume of physical activity, and therefore the dose-response relationship with health outcomes, to be measured more sensitively [83]. Although UK Biobank has a low response rate for its baseline survey (5.5%), 44.8% of participants invited to wear an accelerometer accepted the invitation creating potential sample of over 100 000 participants with objective physical activity data [79]. This sample size far exceeds that of comparable studies [84]–[88], making it the largest accelerometer cohort to date. A large range of complementary physical activity measures are also available, including self-reported behaviours, such as walking, and objective measures of overall physical activity. This allows for specific types of physical activity to be investigated in line with total volumes of activity, further clarifying the Biobank dataset as a suitable choice for analysis and for meeting aim (a) of the thesis.

Research in this thesis was conducted using the UK Biobank Resource under Application Number 20684. All participants provided UK Biobank with explicit consent to link to any health-related records. Any participant can withdraw at any time without giving a reason. UK Biobank and its funders (principally, the Medical Research Council and Wellcome Trust) were advised throughout the development of UK Biobank's information materials and consent form by, amongst others, the independent Ethics & Governance Council. This process also involved assessing participants' understanding of the consent that they were giving to UK Biobank. The study protocol has been approved by the North West Research Ethics Committee.

1.4.2.2 *Commuting and Health in Cambridge*

The Commuting and Health in Cambridge is a quasi-experimental cohort study conducted in four annual phases between 2009 and 2012. The Cambridgeshire Guided Busway opened in 2011 and comprised a new bus route and traffic-free pathway for walking and cycling. The primary purpose of the study was to investigate the effect of the busway, a major change to the built environment, on travel behaviour, physical activity, and health. A detailed description of the study has been published elsewhere [89], [90].

Participants were adults who worked in Cambridge, UK and lived within 30 km of the city centre. Recruitment primarily occurred through workplaces which varied in type of employment and geographic setting. For each year, participants completed a questionnaire collecting information on sociodemographic characteristics, travel behaviour, physical activity and health. A subset of participants was also invited to participate in objective physical activity monitoring at the second, third and fourth phase of the study. Participants wore a GPS receiver and combined heart rate and movement sensor for 7 days at the same time each year. The study comprised qualitative data in the form of semi-structured and photo-elicitation interviews. I used interview data collected post-intervention designed to elucidate the perceived impact of the busway on travel behaviour and reasons for and against its use.

There were several advantages to using this dataset. The study was longitudinal and repeat GPS and objective physical activity data were available before and after the opening of a major transport infrastructural intervention. In contrast to many of the studies which have used GPS and accelerometer to locate physical activity, qualitative data were also available, enabling the exploration of how qualitative and quantitative data can be used together to understand ways in which the intervention was used.

The dataset is managed by the Medical Research Council Epidemiology Unit at the University of Cambridge. The Hertfordshire Research Ethics Committee approved the baseline data collection (reference number 08/H0311/208) and the Cambridge Psychology Research Ethics Committee approved the follow-up data collection (reference number 2014.14). All participants provided written informed consent.

1.4.3 Thesis structure

The thesis is presented as a series of chapters which build on one another to realise the aims of the thesis (Section 1.4.1). Chapter 2 addresses aim (a) and uses data from UK Biobank to understand the relationships of different types of physical activity with different aspects of the neighbourhood environment. Cross-sectional associations between a broad range of environmental characteristics, including walkability, air pollution and deprivation, and self-reported walking and objective measures of physical activity are examined. The chapter discusses the trade-offs between environments conducive and limiting for physical activity and the implications of focusing on a single environmental characteristic for neighbourhood design and public health planning strategies.

Chapter 2 uses static measures of environments around residential locations. Chapter 3 provides a more representative picture of where people may be active by focusing on activity spaces. In this chapter, I address aim (b) by systematically reviewing qualitative and quantitative studies which apply the activity space concept to investigate the relationship between the environment and physical activity. The spatial and temporal methods used to define activity spaces, research questions answered, and the implications for causal inference are reviewed. Future directions for research are highlighted including ways in which studies may limit the risk of bias.

Informed by the conceptual work of the review in Chapter 3, in the remainder of the thesis I use qualitative and quantitative data to address aims (c) and (d). I use GPS, questionnaire, interview, and objective physical activity data from the Commuting and Health in Cambridge study to understand how new transport infrastructure might give rise to changes in use of space and physical activity.

Chapter 4 details the development and testing of data processing methods to clean GPS data and derive activity spaces. Using the processed data, in Chapter 5 I apply an exploratory approach to understand how activity spaces change in response to the opening of new transport infrastructure. Qualitative data is also integrated with visualisations of mapped GPS data to elucidate how the infrastructure is used for physical activity, walking, and cycling.

The cleaned GPS data from Chapter 4 are matched to combined heart rate and movement sensor data in Chapter 6. In this chapter I evaluate the applicability of different methods to locate physical activity and assess whether the use of new infrastructure contributes to an increase, decrease, or displacement of physical activity over time.

Chapter 7 presents the key findings and the strengths and limitations of work presented in the thesis. The implications of the wider research and recommendations for future research are discussed.

Chapter 2

Characteristics of the environment and physical activity in midlife: findings from UK Biobank

2.1 Introduction

2.1.1 Chapter overview

This chapter describes cross-sectional associations between neighbourhood environments of varying scales and physical activity outcomes measured using objective and self-reported assessments in the UK Biobank dataset. The rationale for choosing each environmental characteristic and methods used to derive measures are described in detail in the methods section, before presenting the study results and discussing the implications of the findings. A shortened version of the study has been published in Preventive Medicine [91].

2.1.2 Background

Physical inactivity accounts for 9% of premature mortality worldwide and engaging in regular physical activity reduces the risk of non-communicable diseases including cardiovascular disease, type 2 diabetes, and some cancers [1]. Moderate to vigorous activity (MVPA) confers health benefits and allows for comparisons with activity recommendations [4]. Activities such as walking could also foster social interactions, promote social equity, improve air quality, and lead to more environmentally sustainable communities by displacing car use [6], [8]. However, many adults do not achieve sufficient levels of activity [10].

It is hypothesised that the environment and social context in which people live is related to physical activity [32]. The number of studies exploring these associations has increased in the past 20 years with much of the literature focused on micro-level attributes of the physical built environment which may provide spaces for use and improve destination accessibility [14], [45], [59].

Applying a public health perspective and embracing the notion of the wider social determinants of health [45], [92] suggests a range of micro- and macro-level environmental attributes might be important. Contextual conditions such as deprivation and rurality are likely to influence health behaviours, as well as more immediate conditions of environmental disturbance or the natural environment which affect the desirability to use space. It is postulated that high levels of pollutants increase the perception of risk [93] and discourage outdoor activity. Additionally, poorer communities are often disproportionally exposed to air pollution [94]. Few studies have examined the role of air pollution and its association with physical activity [95] and none have assessed contextual characteristics, such as deprivation and air pollution, and micro-level characteristics of urban form. Investigating these simultaneously may help provide a broader perspective on the role of the residential environment as it relates to physical activity. This is important for better understanding the trade-offs between characteristics more or less conducive to physical activity and the implications for public health.

Objective measures enable precise data to be collected on duration and intensity of activity [96]. A large-scale study of participants living in 14 cities found that parks and greater residential density in the neighbourhood were positively associated with objectively measured MVPA [97], however, specific behaviours were not investigated. The most consistent associations are drawn from studies where domain or activity-specific outcomes and exposure measures are well-matched [14]. For example, a UK study found that greenness was associated with active commuting and walking which contribute to overall MVPA [98]. However, it is difficult to identify these activities accurately from objective physical activity data alone. Combining objective with self-reported measures of activities such as walking can therefore complement precise estimates of total activity with information on specific activity behaviours.

2.1.3 Aims and scope

Using a large dataset with geographical heterogeneity, this chapter seeks to assess the associations between environmental characteristics in the residential neighbourhood and a range of objective (‘recorded’) and self-reported (‘reported’) measures of physical activity and walking. Characteristics are described under five broad facets (spaces for physical activity, walkability, disturbance, the natural environment, and the sociodemographic environment) which range from micro-level environments considered to encourage specific types of activity, to macro-level environments which may affect levels of activity more generally. Physical activity measures increase in specificity from recorded total activity to reported time in walking behaviours.

2.2 Methods

2.2.1 Study design

Cross-sectional data were used from the UK Biobank study, collected from 502 656 participants aged 37-73 years at recruitment. Respondents were invited if they were registered with the National Health Service (NHS) and lived within 35 km of one of 22 Biobank assessment centres (Figure 2.1).

Figure 2.1: Distribution of UK Biobank assessment centres and total number of participants at baseline

Baseline data including sociodemographic, lifestyle, and physical activity information were self-reported between March 2006 and July 2010 [99]. A random sub-sample of participants (n=236 519) who provided a valid email address were invited to take part in objective physical activity measures [79]. Accelerometers (Axivity AX3) were posted to those who agreed to participate (44.8%, n=106 053) and worn between June 2013 and December 2015.

The UK Biobank study has ethical approval from the North West Multi-centre Research Ethics Committee (MREC), Information Advisory Group (IAG), and the Community Health Index Advisory Group (CHIAG). Details on the Biobank study design and survey methods are described in a full protocol and accompanying paper [78], [100].

2.2.2 Inclusion criteria

Analysis was restricted to participants who had data on all environmental characteristics, covariates and at least one physical activity outcome. Two samples were therefore created: those with accelerometer data and those who provided information for reported activity.

Environmental measures were derived for home addresses collected at baseline. Information on home location was collected at two further time points for a sub-sample; firstly, between December 2009 and June 2013 (n=20 346) and secondly between April 2014 and November 2016 (n=11 923). As measures of physical activity were recorded after baseline, I identified participants who had moved home to ensure environmental exposures were appropriate for analysis. Locations of residential addresses between baseline and follow-up were subsequently compared. The scale at which coordinates were presented were different for both time points so the follow-up data was rounded to match the coarser scale used at baseline. If follow-up coordinates were different to baseline and there was assumed to be no rounding error, participants were classified as movers and excluded from analysis. As follow-up data were not available for all participants, it is unlikely all movers are captured but this is the most reasonable approach given the data available.

I also considered excluding those whose mobility, and therefore physical activity, was limited. However, the data available on the presence of pain in the leg or chest when walking was only available for a sub-sample of participants and did not differentiate between musculoskeletal and cardiovascular issues. With this limited information, it is difficult to understand the nature of confounding clearly. Although it may be more difficult to walk, increased walking may be part of a rehabilitation program or self-selection may occur whereby participants live closer to facilities for ease of access or where reliance on private transport is easier. I therefore chose not to exclude participants from the sample if they had reported pain.

2.2.3 Physical activity

Five physical activity outcomes were included for analysis based on the accuracy and specificity of the measure: mean acceleration, recorded time spent in MVPA, reported MVPA, walking, and walking for pleasure.

2.2.3.1 Recorded physical activity

Objectively-measured physical activity data were collected for a random sub-sample of participants from all assessment centres except those in the North West of England, which were excluded due to concerns of participant burden from trialling other new projects in this region. Participants wore an accelerometer on their dominant wrist for 7 days, including night-time. Data from wrist-worn devices have been validated against established measures of physical activity energy expenditure [101].

Doherty and colleagues describe the calibration and data processing in detail [79]. Briefly, measures of acceleration were collected in 5 second epochs, maintaining the mean acceleration over the duration of the epoch. The percentage of time spent in different ranges of acceleration for the week are available in the dataset. Given the fractional measures of acceleration are derived from a cumulative distribution function of all 5 second epochs, sustained bouts of acceleration are not accounted for in the data. Non-wear time was previously identified as stationary episodes of at least 60 minutes and removed from the data in line with protocol for processing physical activity used elsewhere [102], [103]. Only participants with more than 72 hours of wear time were included in the sample and periods of non-wear time had been imputed using the average magnitude from a similar time on a different day of measurement.

I used two measures of recorded physical activity for analysis: mean acceleration and time spent in MVPA. Mean acceleration which assesses average volume of activity in milli-gravity units (m*g*) for the week was already available in the dataset, calculated by averaging worn and imputed values. This value was used to indicate a global measure of activity. Data were also available as fraction of time spent over different acceleration thresholds which I used to estimate the total minutes spent in MVPA over the course of the week. MVPA equates to 3 METs which is equal to 134 m*g* of acceleration captured by the dominant wrist [101]. Based on the available data and discussions with colleagues in the MRC Epidemiology Unit, I used the fraction of time spent above the closest available acceleration threshold in the processed dataset (125 m*g*). The total time spent in MVPA was then divided into tertiles. Previous studies have used daily measures of bouted MVPA based on established cut-points of counts per minutes collected from hip-worn accelerometers [97], [104]. The advantage of using wrist-worn over hip-worn devices is that they can be worn continuously day and night, are waterproof, and result in higher levels of participant compliance. Although the MVPA metric I computed is not directly comparable with existing studies, it was considered appropriate for the assessment of patterns across the range of environmental characteristics and physical activity outcomes in this study.

2.2.3.2 Reported physical activity

Self-reported physical activity data were collected from a touchscreen questionnaire completed at an assessment centre. The full questions are available online [99] and are similar to those used in the short form of the International Physical Activity Questionnaire (IPAQ) [105].

Using the following questions relating to MVPA and walking, participants were asked how many days in a typical week they did each type of activity for at least 10 minutes and the duration of each episode.

<u>For moderate physical activities</u>
"In a typical WEEK, on how many days did you do 10 minutes or more of moderate physical activities like carrying light loads, cycling at normal pace? (Do not include walking)"
"How many minutes did you usually spend doing moderate activities on a typical DAY?"

<u>For vigorous physical activities</u>
"In a typical WEEK, how many days did you do 10 minutes or more of vigorous physical activity? (These are activities that make you sweat or breathe hard such as fast cycling, aerobics, heavy lifting)"
"How many minutes did you usually spend doing vigorous activities on a typical DAY?"

For walking

"In a typical WEEK, on how many days did you walk for at least 10 minutes at a time? (Include walking that you do at work, travelling to and from work, and for sport or leisure)"

"How many minutes did you usually spend walking on a typical DAY?"

To calculate the weekly time spent in each activity, I multiplied the number of reported days by the duration. Participants who did not report a frequency but a duration or a frequency but not a duration were assigned to the median frequency or duration for that activity based on responses from other participants in the sample. To calculate minutes spent in MVPA, the weekly time spent in both moderate activity and vigorous activity were generated and summed.

Walking for pleasure was assessed in a similar way, except that categorical response items were used. Using the following questions, participants were asked if they had spent any time walking for pleasure, not as a means of transport, within the last 4 weeks alongside a list of other activities (such as swimming or cycling), light DIY or heavy DIY. Those who responded positively to any of the activities were prompted to report the duration of activity using the available options.

For walking for pleasure

"How many times in the last 4 weeks did you go walking for pleasure?"

Once in the last 4 weeks

2-3 times in the last 4 weeks

Once a week

2-3 times a week

4-5 times a week

Every day

"Each time you went walking for pleasure, about how long did you spend doing it?"

Less than 15 minutes

Between 15 and 30 minutes

Between 30 minutes and 1 hour

Between 1 and 1.5 hours

Between 1.5 and 2 hours

Between 2 and 3 hours

Over 3 hours

I converted the categorical responses for frequency to the frequency per week (i.e. those reporting engaging in activity ‘once a week’ were assigned a frequency of ‘1’ and those reporting activity ‘every day’ were assigned ‘7’). For duration in minutes, activities were assigned to the median of that category (i.e. ‘between 30 minutes and 1 hour’ was assigned ‘45 minutes’. Those responding ‘less than 15 minutes’ and ‘over 3 hours’ were assigned 7.5 minutes and 210 minutes respectively). These assignments match the ones used in the processing of self-reported physical activity from several large cohorts including EPIC-Norfolk [106]. I computed weekly time spent walking for pleasure by multiplying the reported number of days by the median duration.

Given the variation in questions used to measure time spent in different types of activity, I chose to divide all self-reported times into tertiles. I considered this appropriate to allow for broad comparisons and the identification of patterns across the physical activity outcomes.

2.2.4 Environmental data

The UK Biobank Urban Morphometric Platform (UKBUMP) is a nationwide resource and uses objective data to characterise environmental conditions that influence health using a range of buffer sizes around each participant’s home location [80]. Variables were based on a conceptual model [80] and previously derived to serve a range of research questions related to physical activity, diet, alcohol consumption and general health. The processes are described in detail elsewhere [80], [81]. Briefly, measures of environmental conditions are available for each participant, based on the characteristics within a defined straight-line or network distance of their residential address. I chose to use measures which characterised the area within 1 km, or closest available distance, as this corresponds to a 10-15 minute walk, and 0.8-1 km is commonly used and broadly accepted in the literature [107], [108].

I selected fifteen variables conceptually and most plausibly related to physical activity (excluding those related to diet and alcohol consumption) for analysis. I grouped these variables into five broad facets (spaces for physical activity, walkability, disturbance, natural environment, and the sociodemographic environment) based on theme and their influence on different activity types (Table 2.1). I performed exploratory analyses based on the distribution of the environmental data and consulted previous studies and recommended levels of pollution to identify suitable cut points for each environmental factor.

Facets ranged from micro-level environments such as facilities designed for physical activity, which are considered to encourage specific types of activity, to macro-level environments such as urban-rural status, which may affect levels of activity more generally. The hypothesised pathways between the environmental variables and physical activity outcomes are detailed in Figure 2.2.

Table 2.1: Description and classification of objectively measured environmental variables

Variable	Description	Spatial scale *Buffer type*	Data source[a], *Year*	Classification
Spaces for physical activity				
Facilities for physical activity	Presence of facilities for physical activity	1 km[b] *network*	UK OS AddressBase Premium point data, *2013*	No / Yes
Parks	Presence of parks	1 km[b] *network*	UK OS AddressBase Premium point data, *2013*	No / Yes
Walkability				
Walkability	Composite measure of street connectivity, residential density and land use mix Z scores of component measures were generated and summed	n/a	Derived from UK OS ITN, *2010* and UK OS AddressBase Premium point data, *2013*	Quartile
Disturbance				
Air pollution	Annual average for concentration of nitrogen oxides (NO_X)	Interpolated from model at residential address	European Study of Cohorts for Air Pollution Effects (ESCAPE) Land Use Regression model, *2010*	<26 μgm^{-3} / ≥26 μgm^{-3}
Noise pollution	Average daytime sound level pressure over 12 hour period (07:00 to 19:00)	Interpolated from model at residential address	Common NOise aSSessment methOdS (CNOSSOS-EU) model, *2009*	<54 kHz / ≥54 kHz
Distance to major road	Inverse distance to the nearest major road based upon a local road network where a major road is a road with traffic intensity >5000 motor vehicles per 24 hours	n/a	Road network: OS Meridian 2 road network, *2009* Traffic data: Eurostreets (vs 3.1) digital road network, *2008*	Quartile
Natural environment				
Terrain	Mean slope angle	1 km[b] *Circular*	Landmap DTM (5 m resolution) Stereo aerial photography *1998–2008*	<3° / ≥3°
Greenness	Mean normalised deviation vegetation index (NDVI)	0.5 km *Circular*	CIR Landmap satellite data (5m resolution), *2006-2010*	Quartile
Sociodemographic environment				
Urban-rural status	Based on population density	Postcode	Office for National Statistics Postcode Directory (ONSPD) and UK Census data, *2001*	Urban / Fringe / Rural
Area-level deprivation	Townsend deprivation index	Census output area	UK Census data, *2001*	Quintile

OS = Ordnance Survey; ITN = Integrated Transport Network; DTM = Digital Terrain Model; CIR = Colour Infrared
[a]For further details on data sources, please refer to UKBUMP data analysis and specification manual [81]
[b]0.5 km distance used for sensitivity analyses to investigate the effects of smaller neighbourhood measures

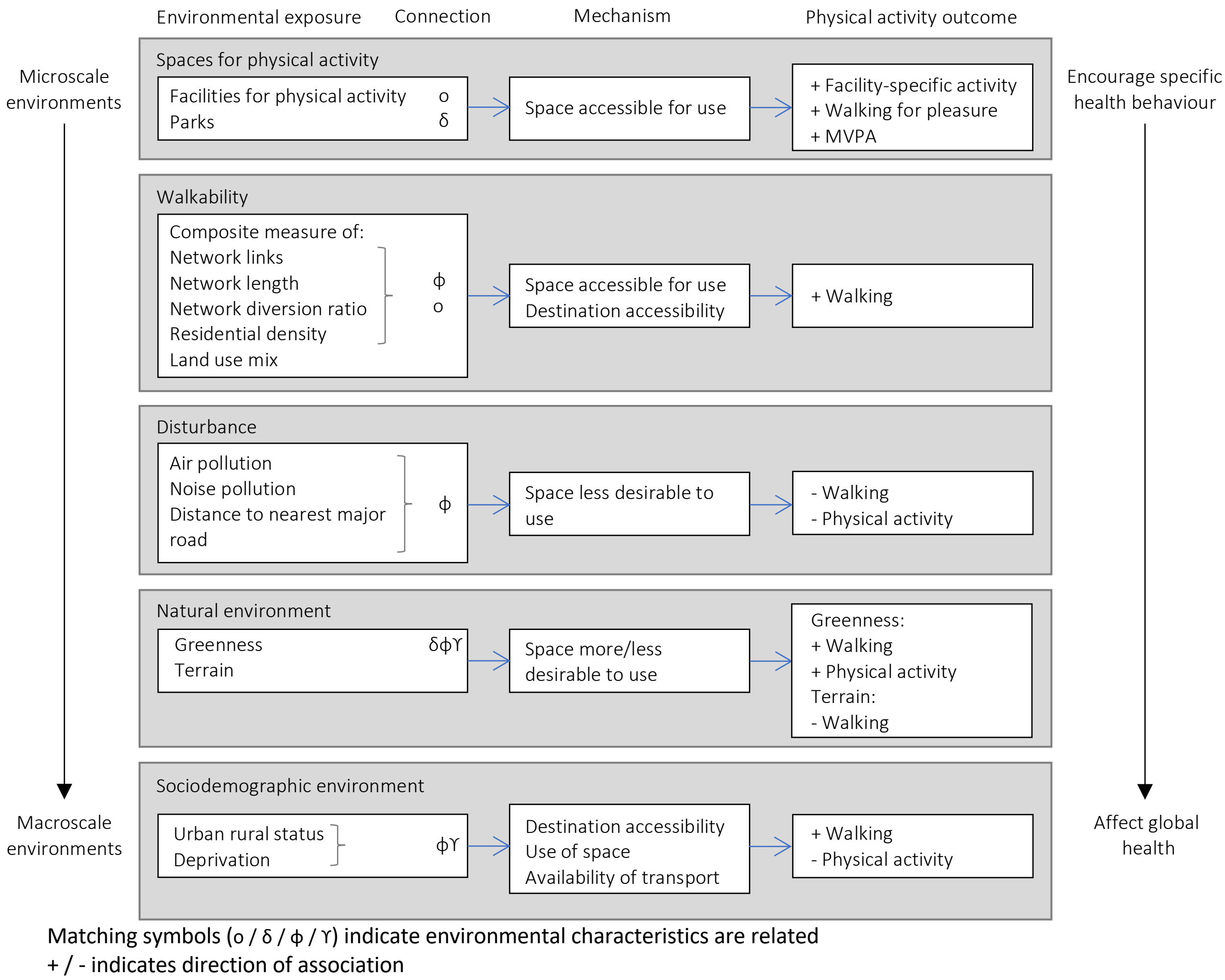


Figure 2.2: Hypothesised pathways between environmental characteristics and physical activity

2.2.4.1 Spaces for physical activity

Land use feature data were used to identify i) facilities designed for physical activity to take place in and ii) public parks. Full details of all features included in the classification are detailed in Table 2.2. The distribution of the data showed that relatively few participants had these facilities around their home, a binary classification was therefore used where neighbourhoods were categorised as having access to spaces for physical activity or not. I also performed a sensitivity analysis performed to include facilities where activity could be undertaken but have not been purposefully designed for this, such as community centres (Table 2.2).

Table 2.2: Features included as spaces for physical activity

	Included for main analysis	Included for sensitivity analysis
i) Facilities for physical activity[a]		
Indoor/outdoor leisure centre	✓	✓
Bowls facility	✓	✓
Cricket facility	✓	✓
Swimming facility	✓	✓
Equestrian facility	✓	✓
Football facility	✓	✓
Golf facility	✓	✓
Leisure/sports centre	✓	✓
Racquet sports facility	✓	✓
Playing field	✓	✓
Recreation ground	✓	✓
Rugby facility	✓	✓
Tenpin bowling	✓	✓
Water sports facility	✓	✓
Public hall/Community facility		✓
Church hall		✓
Private social club		✓
Arena/stadium		✓
ii) Public parks		
Park	✓	✓
Public park/garden	✓	✓
Open space		✓
Public open space/nature reserve		✓

[a]Skate parks and winter sports facilities were not included as data for these features were only available for a limited portion of the sample (n=20 152 participants)

2.2.4.2 Walkability

For the main analysis, I derived a composite score for walkability, based on measures of street connectivity, residential density and land use mix [109]–[112]. The separate components are described in Table 2.3.

Table 2.3: Description of objectively-measured walkability component variables

	Variable	Description	Spatial scale *Buffer type*	Data source, *Year*	Classification
Walkability components	**Street connectivity**				
	Network links	Total number of network links where a link is a street joining a junction or dead-end to a junction or dead-end	1.2 km[a] *network*	UK OS ITN, *2010*	Quartile
	Network length	Total length of links	1.2 km[a] *network*	UK OS ITN, *2010*	Quartile
	Network diversion ratio	Mean difference between crow-flight path and actual path for all links	1.2 km[a] *network*	UK OS ITN, *2010*	Quartile
	Residential density	Total number of residential addresses divided by total neighbourhood area (no. features/square km)	1 km *network*	UK OS AddressBase Premium point data, *2013*	Quartile
	Land use mix	Proportion of land use squared and summed	1 km *network*	UK OS AddressBase Premium point data, *2013*	Quartile

[a]0.4 km distance used for sensitivity analyses to investigate the effects of smaller neighbourhood measures

Land use density data were available as the number of features per square kilometre. To measure land use mix, I grouped features considered to be walkable destinations into five categories: residential, retail, office, community, and recreational space based on literature, locale, and available data [113]–[116]. I created a land use mix score using the Herfindahl Hirschman Index (HHI) (Equation 1), as used in similar studies [117]–[120]. The HHI was considered an appropriate calculation of mix as the density measure of the data refers to the number of features, rather than the proportion of land cover. It is unlikely that there will be an equal distribution of residential features to retail features, for example, which would score highly in an entropy formula. Instead, the HHI assesses the range of land uses, with a greater number of categories per square kilometre scoring better than an equal distribution of fewer categories. HHI scores ranged from zero to 10 000 (100^2) where a high score indicates a low level of land use mix.

Equation 1: $\Sigma(p_i^2)$

p represents the proportion of features devoted to a specific land use (i) per square km of 1 km network buffer

p is calculated by dividing the number of features in land use (i) by the total number of features of all present land uses per square km of 1 km network buffer

All network measures were divided into deciles and summed to create a street connectivity score out of 30. As in previous research[109]–[112], Z scores were generated for the combined street connectivity score, residential density, and land use mix, then summed to create a walkability score for each participant. A higher score indicated greater walkability. I also investigated the walkability components separately in sensitivity analysis.

2.2.4.3 Disturbance of the environment

<u>Air pollution</u>

Air pollutants have been measured at 36 sample areas across Europe and modelled using a land use regression (LUR) model [121]. The LUR model accounts for predictors of air pollutants including land use, traffic, and geographic characteristics and is used to estimate outdoor air pollution at participant's addresses [122]. Annual average concentrations of particulate matter with aerodynamic diameter ≤ 2.5 µm ($PM_{2.5}$), nitrogen dioxide (NO_2), and nitrogen oxides (NO_X) are available as continuous measures within the UK Biobank dataset. Given the data available, I selected the variables to include based on a combination of preliminary analysis and my a priori conceptual rationale.

Based on the hypothesis that low air quality deters use of greenspace and the neighbourhood in general, people's perceptions of the neighbourhood may impact their choice to be active. Measures of pollutants where traffic is the main source (such as NO_X) may be important as people may avoid main or busy roads which have high volumes of traffic. Using $PM_{2.5}$ which is the most harmful to health [123] may be problematic as these pollutants may be difficult for people to perceive [124].

To avoid issues of collinearity and help in the decision about which measure to carry forward, I examined the univariate associations between each pollutant, and between each pollutant and physical activity outcome. I found that NO_X showed the strongest and most consistent results with all physical activity outcomes (high concentrations associated with lower levels of activity). NO_X has been used in other epidemiological studies [93], [125] and an additional data site in the UK was used to generate the NO_X model than $PM_{2.5}$ [121]. I therefore chose to use NO_X as an indicator of air pollution. As data in UK Biobank are skewed towards lower levels of air pollution it was not sensible to use recommended levels of air quality to classify the data [126]. Instead, measures of NO_X were dichotomised based on the median. A measure of distance to the nearest major road was also included in the models to capture the further impact of traffic volumes.

Noise pollution

Levels of noise pollution have been mapped based on variables including road traffic, railway traffic and industrial noise sources in Europe [127]. A measure of average daytime sound was chosen for analysis as it was assumed that most physical activity in the neighbourhood would take place during the day. Levels of noise pollution were dichotomised based on the median for analysis.

2.2.4.4 Natural environment

Terrain

As outlined in Figure 2.2, I hypothesised that hilly environments may not be conducive to activities such as walking. Based on the distribution of the data, I categorised terrain into the least and most hilly environments based on the median of the data.

Greenness

Normalised deviation vegetation index (NDVI) was calculated based on 0.5 cm by 0.5 cm resolution colour infrared (CIR) imagery [98]. The satellite images were collected during summertime across the baseline phase of the UK Biobank study (2006-2010) and values were averaged to calculate mean NDVI to minimise temporal misclassification. Measures of greenness were classified for all participants in the same way, reducing confounding by seasonal variation.

2.2.4.5 Sociodemographic environment

Urban-rural status

Seventeen categories for urban-rural status were provided in the UKBUMP dataset based on country and home area population density. For the purposes of analysis, these categories were collapsed into three groups: urban, town and fringe, and rural.

Area-level deprivation

The Townsend deprivation index is a composite measure of deprivation based on unemployment, non-car ownership, non-home ownership, and household overcrowding; a negative value represents high socioeconomic status. This was calculated before participants joined the UK Biobank and was based on the preceding national census data, with each participant assigned a score corresponding to the postcode of their home dwelling. I categorised deprivation scores into quintiles for analysis.

2.2.5 Covariates

All covariates were derived or self-reported in the lifestyle questionnaire during baseline assessment and comprised age, sex, ethnicity, assessment centre, highest educational qualification, income, employment status, housing tenure, number of vehicles in household, whether children lived in the household, urban-rural status, and area-level deprivation.

2.2.6 Statistical analysis

Environmental data were available for participants from all assessment centres except Stockport where the pilot study was conducted. Both the full and the potential sample (those with environmental data) were compared with the final analytic samples (those with either recorded or reported activity) to investigate attrition through the exclusion process of this study. Descriptive analyses were undertaken to assess the characteristics of the samples, and Wilcoxon rank sum tests were used to compare recorded and reported activity.

To test for collinearity, the correlation between each environmental variable was examined and where correlations were greater than 0.5, the variable most strongly related to physical activity was used. Linear regression models were used to assess the associations between the environmental characteristics and mean acceleration. Multinomial logistic regression models were used for tertiles of recorded and reported time spent in MVPA, walking, and walking for pleasure as preliminary analyses indicated that assumptions of linear regression could not be satisfied. First, univariate regression analyses were conducted for each environmental characteristic, adjusting for covariates (Model 0). All significant characteristics ($p<0.05$) were carried forwards into a single adjusted model for each activity outcome (Model 1). Significance was assessed with tests for trend across each activity tertile.

Following the regression analyses, I looked for differences in directions of association and significance between Model 0 and Model 1 to check whether multicollinearity was driving associations seen in Model 1.

All analyses were conducted using STATA/SE 14.1.

2.2.7 Sensitivity analysis

Sensitivity analyses were run to explore which components of the walkability scores (street connectivity, land use mix, and residential density) contributed most to any associations observed. To investigate the effects of using smaller neighbourhood measures, further sensitivity analyses were performed by repeating the process with smaller buffer sizes for facilities for physical activity, parks, walkability, and terrain.

2.3 Results

2.3.1 Sample

Environmental data for all exposures of interest were available for 352 755 participants (70.2% of full sample), of whom 65 967 (18.7%) had valid recorded physical activity measures and 337 822 (95.8%) provided information on at least one of the three reported outcomes (Figure 2.3). The distribution of characteristics was similar for all samples (Table 2.4). The sample with reported physical activity data were most similar to the full sample while the sample with recorded physical activity data contained a higher proportion of women and were more likely to be educated to degree level, in paid employment, a home owner, and have access to a vehicle.

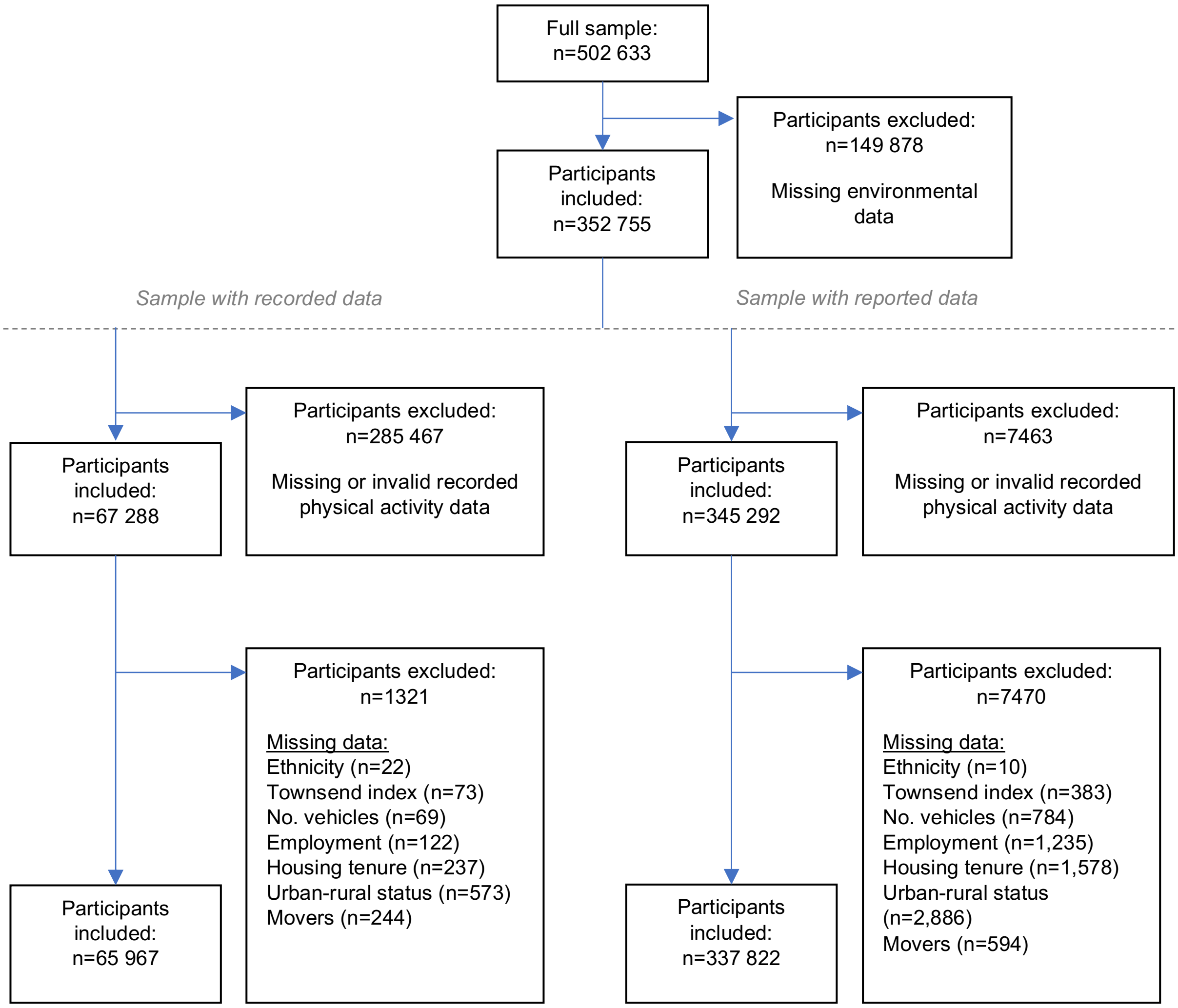


Figure 2.3: Flowchart of process for inclusion for participants with reported and recorded physical activity data

Table 2.4: Sample characteristics

	Full sample (n=502 633)	Sample who had environmental data available (n=352 755)	Sample who provided recorded physical activity data (n=65 967)	Sample who provided reported physical activity data[a] (n=337 882)
	n (%)	n (%)	n (%)	n (%)
Sex				
Male	229 171 (45.6)	160 917 (45.6)	28 718 (43.5)	154 042 (45.6)
Female	273 462 (54.4)	191 838 (54.4)	37 249 (56.5)	183 780 (54.4)
Age at baseline				
40-49	117 903 (23.5)	82 573 (23.4)	15 282 (23.2)	78 590 (23.3)
50-59	167 191 (33.3)	116 173 (32.9)	23 686 (35.9)	111 438 (33)
60-69	215 112 (42.8)	152 292 (43.2)	26 767 (40.6)	146 145 (43.3)
70-79	2427 (0.5)	1717 (0.5)	232 (0.4)	1649 (0.5)
Age at recorded physical activity assessment				
40-49	8785 (8.5)	6331 (8.8)	5544 (8.4)	n/a
50-59	29 911 (28.8)	20 758 (28.8)	18 761 (28.4)	
60-69	45 938 (44.3)	31 887 (44.2)	29 388 (44.5)	
70-79	19 076 (18.3)	13 188 (18.3)	12 274 (18.6)	
Ethnicity				
White	472 816 (94.6)	331 981 (94.2)	63 689 (96.5)	319 543 (94.6)
Non-white	27 039 (5.4)	20 267 (5.8)	2278 (3.5)	18 279 (5.4)
Weight status				
Underweight/Normal	165 073 (33.0)	114 334 (32.6)	25 556 (38.8)	110 381 (32.8)
Overweight	212 168 (42.5)	149 218 (42.6)	27 189 (41.3)	143 671 (42.7)
Obese	122 287 (24.5)	87 079 (24.8)	13 085 (19.9)	82 266 (24.5)
Urban-rural status				
Urban	428 890 (86.2)	303 764 (86.9)	56 059 (85.0)	293 056 (86.7)
Fringe	33 865 (6.8)	24 226 (6.9)	5050 (7.7)	23 613 (7.0)
Rural	34 803 (7.0)	21 676 (6.2)	4858 (7.4)	21 153 (6.3)
Highest educational qualification				
College or University degree	161 206 (32.4)	109 644 (31.1)	27 666 (41.9)	106 575 (31.5)
Other professional (e.g. teaching)	25 810 (5.2)	18 328 (5.2)	3365 (5.1)	17 623 (5.2)
Higher education (e.g. A Levels, NVQ)	88 070 (17.7)	61 692 (17.5)	12 170 (18.4)	59 627 (17.7)
Secondary education (e.g. GCSEs)	132 113 (26.5)	97 224 (27.6)	16 794 (25.5)	93 810 (27.8)
Other	90 787 (18.2)	65 381 (18.6)	5972 (9.1)	60 187 (17.8)
Employment status				
Paid employment or self-employment	287 225 (57.2)	199 930 (56.8)	40 229 (61.0)	193 972 (57.4)
Retired	167 013 (33.3)	118 909 (33.8)	21 171 (32.1)	114 604 (33.9)
Unable to work	16 836 (3.4)	12 009 (3.4)	1123 (1.7)	10 408 (3.1)
Unemployed	8265 (1.6)	5880 (1.7)	780 (1.2)	5481 (1.6)
Home duties, carer, student, volunteer, or other	22 423 (4.5)	15 541 (4.4)	2664 (4.0)	13 357 (4.0)
Housing tenure				
Home owner	442 566 (89.6)	312 526 (88.9)	62,232 (94.3)	304,046 (90.0)
Renting	46 462 (9.4)	31 452 (8.9)	3,066 (4.6)	28,747 (8.5)
Other	5123 (1.0)	7449 (2.1)	669 (1.0)	5,029 (1.5)
No. vehicles in household				
Two or more	245 129 (49.0)	170 355 (48.5)	34 839 (52.8)	165 238 (48.9)
One	208 636 (41.7)	149 192 (42.5)	27 420 (41.6)	143 131 (42.4)
Other	46 606 (9.3)	31 878 (9.1)	3708 (5.6)	29 453 (8.7)
People in the household				
One	92 942 (18.6)	63 395 (18.1)	10 691 (16.2)	60 478 (18.0)
Two	232 811 (46.6)	164 856 (47.1)	31 655 (48.1)	159 104 (47.2)
Three or more	172 324 (34.5)	121 638 (34.8)	23 527 (35.7)	117 178 (34.8)
Children in household				
No	324 331 (64.8)	227 131 (64.6)	41 756 (63.3)	217 580 (64.4)
Yes	176 040 (35.2)	124 294 (35.4)	24 211 (36.7)	120 242 (35.6)

[a]This sample included any participant who provided information on any of the three reported outcomes (time spent in MVPA, total walking, or walking for pleasure).

2.3.2 MVPA, total walking and walking for pleasure

For each tertile of recorded MVPA, the greatest proportion of participants was in the corresponding tertile of reported MVPA (Table 2.5, Panel A). Similar and more convincing patterns are shown for reported MVPA and walking (Panel B), and walking and walking for pleasure (Panel C). Tests for trend indicated each pair of measures were related (p<0.001).

Table 2.5: Comparing reported and recorded physical activity and walking behaviours

		Lower tertile n (%)	Middle tertile n (%)	Upper tertile n (%)
Panel A:				
		Recorded time spent in MVPA		
Reported time spent in MVPA	Lower tertile	8357 (39)	7000 (32)	5102 (33)
	Middle tertile	7337 (34)	8095 (37)	7868 (36)
	Upper tertile	5892 (27)	6926 (31)	8856 (41)
	Total	21 586 (100)	22 021 (100)	21 826 (100)
Panel B:				
		Reported time spent in MVPA		
Reported time spent walking	Lower tertile	57 467 (52)	39 264 (35)	18 814 (16)
	Middle tertile	33 723 (30)	43 012 (38)	33 199 (29)
	Upper tertile	20 057 (18)	29 984 (27)	62 302 (55)
	Total	111 247 (100)	112 260 (100)	114 315 (100)
Panel C:				
		Reported time spent in walking		
Reported time spent walking for pleasure	Lower tertile	58 570 (51)	33 460 (30)	30 647 (27)
	Middle tertile	39 467 (34)	34 658 (32)	26 996 (24)
	Upper tertile	17 508 (15)	41 816 (38)	54 700 (49)
	Total	115 545 (100)	109 934 (100)	112 343 (100)

Panel A: Percentages given are of participants in reported MVPA strata for recorded MVPA tertile
Panel B: Percentages given are of participants in reported time spent walking strata for reported MVPA tertile
Panel C: Percentages given are of participants in reported time spent walking for pleasure strata for reported total walking tertile

2.3.3 Associations between environmental characteristics and physical activity

Associations between environmental characteristics and physical activity were broadly similar in terms of magnitude and statistical significance between Model 0 and Model 1. The results from Model 1 are therefore presented and discussed (Figure 2.4 and Appendix A1, Table A.1).

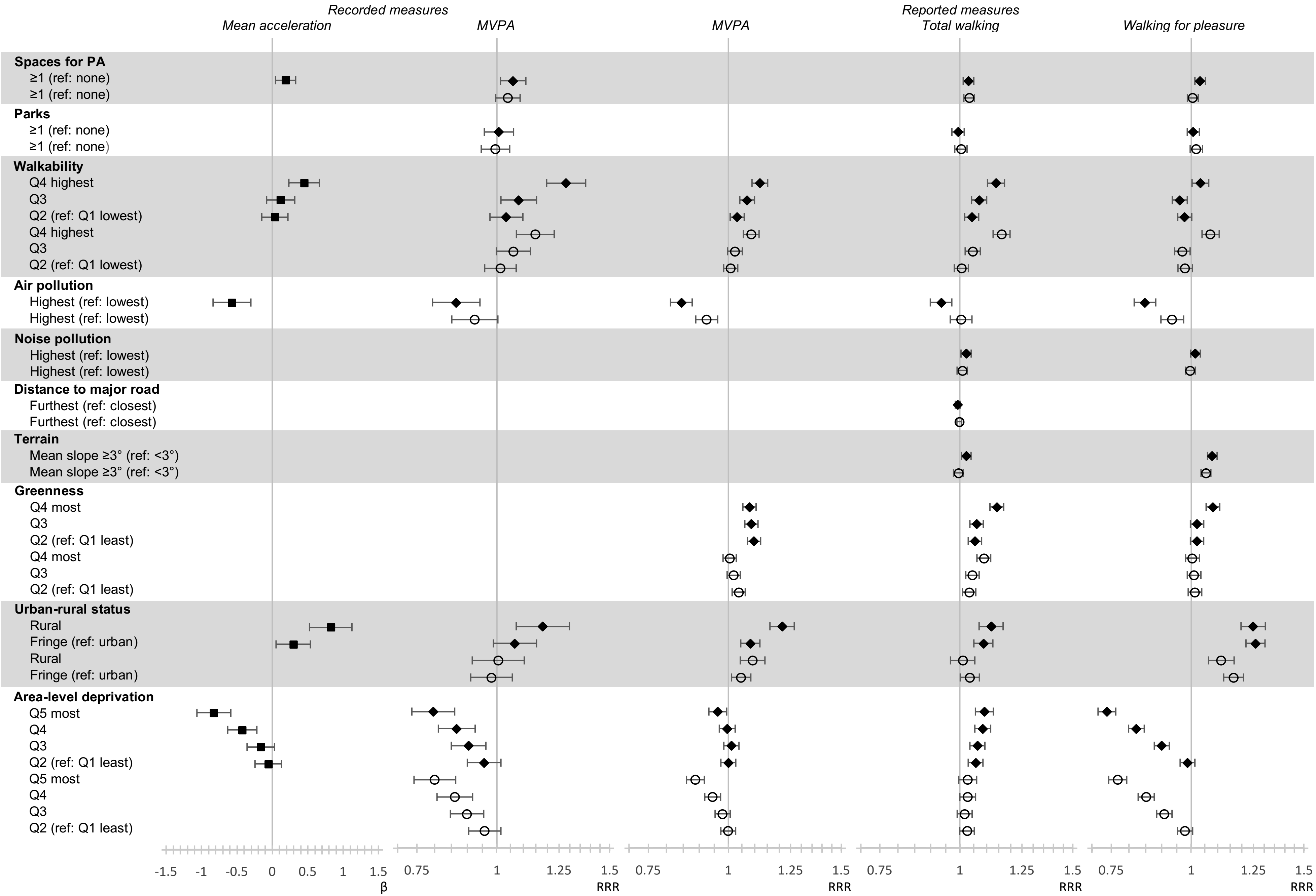


Figure 2.4: Adjusted associations between environmental characteristics and activity outcomes (Model 1)

Outcome variables: ■Continuous data; ◆ Upper tertile; ○ Middle tertile; ⊢⊣95% Confidence interval. White space is where variables have not been included in Model 1

β = regression coefficient presented on linear scale; RRR = relative risk ratio presented on log scale; MVPA = moderate-to-vigorous physical activity

2.3.3.1 *Spaces for physical activity*

Access to facilities for activity was associated with higher mean acceleration (β: 0.19, 95% CI: 0.05, 0.33), higher levels of MVPA (upper tertile RRR: 1.06, 95% CI: 1.01, 1.11), total walking and walking for pleasure. Participants with access to a park, compared to those without, were more likely to report higher levels of walking for pleasure (middle tertile RRR: 1.02, 95% CI: 1.00, 1.04).

2.3.3.2 *Characteristics of walkability*

Neighbourhood walkability was associated with higher levels of reported and recorded activity (all $p<0.001$), except for the upper tertile of walking for pleasure. When comparing the most walkable neighbourhoods with the least, associations were largest for recorded MVPA (upper tertile RRR: 1.28, 95% CI: 1.20, 1.38) and total walking (upper tertile RRR: 1.14, 95% CI: 1.10, 1.17).

2.3.3.3 *Characteristics of disturbance*

Participants living in areas with highest concentrations of air pollution recorded a lower mean acceleration (β: -0.57, 95% CI: -0.84, -0.30). The direction and magnitude of the association were consistent across all other outcomes with a weaker association for total walking. Those living in areas with highest levels of noise pollution were more likely to report higher levels of walking (upper tertile RRR: 1.02, 95% CI: 1.00, 1.04) than those in areas with lowest noise pollution. No significant associations were shown for distance to the nearest major road.

2.3.3.4 *Characteristics of the natural environment*

Participants living in areas with steepest terrain were more likely to report higher levels of walking (upper tertile: RRR 1.02, 95% CI: 1.01, 1.04) and walking for pleasure (upper tertile RRR: 1.08, 95% CI: 1.06, 1.10). Greener neighbourhoods were generally associated with higher reported levels of MVPA, walking and walking for pleasure ($p<0.001$).

2.3.3.5 *Sociodemographic characteristics*

Clear dose-response relationships were shown for characteristics of the sociodemographic environment and all activity outcomes. Participants living in rural areas typically recorded and reported higher levels of activity. Compared to urban dwellers, those in rural areas were more likely to report higher levels of walking for pleasure (upper tertile RRR: 1.25, 95% CI: 1.20, 1.31) which appears to explain the association shown for reported MVPA. Compared to those living in less deprived areas, participants in more deprived areas were less likely to record and report higher levels of activity with a strong negative gradient shown for walking for pleasure (upper tertile: RRR 0.74, 95% CI: 0.72, 0.76). Findings for total walking were in the opposite direction.

2.3.3.6 Sensitivity analyses

Results for the individual walkability components indicated that land use mix was the biggest driver of these associations (Appendix A1, Table A.2). For MVPA, measures of street connectivity appeared to be important, as did residential density for total walking and walking for pleasure.

The results of the adjusted models using smaller distances for facilities for physical activity, parks, walkability, and terrain indicated findings were qualitatively consistent with the original analysis (Appendix A1, Figure A.1).

2.4 Discussion

2.4.1 Principal findings

The study showed that characteristics of the neighbourhood environment were associated with recorded and reported physical activity in a large UK sample of adults. Walkability, disturbance, and the sociodemographic characteristics showed the strongest associations with physical activity, even after adjusting for other characteristics. There were some differences between the associations observed for global measures of activity and more specific behaviours. For example, associations between walkability appeared stronger for total walking than walking for pleasure except for in more deprived areas where a strong negative association for walking for pleasure was shown.

2.4.2 Comparisons with existing evidence

My findings were generally consistent with previous research [14], [58], [59] but some differences could be attributed to the methods used to assess outcomes and exposures or the characteristics of the sample.

The associations for access to facilities for physical activity were most strongly associated with total walking and walking for pleasure and access to parks was weakly associated with walking for pleasure. Mixed findings have been shown for different activity outcomes in the literature [58], [63], [107]. My study focused on physical proximity to facilities whereas others consider convenience, satisfaction and availability but tend not to give a detailed breakdown of the facilities under consideration [14], [58], [60], [63]. When analyses were re-run to include a broader range of recreational facilities not designed specifically for activity (e.g. church halls) the results were not attenuated (data not shown). The weak associations for parks may be because neighbourhood parks are not always the destination for physical activity, or that previous studies explored the size, perceived accessibility or quality of parks [58], [63], [107]. By simultaneously including measures of disturbance and greenness in the analysis, I go some way towards accounting for this. Further studies could investigate the role of factors that moderate the associations between environmental characteristics and activity, such as quality of the environment [128].

Strong positive associations with walkability and mean acceleration, MVPA and walking were found which is consistent with the literature [14], [58], [59]. Land use mix contributed most to the positive associations and this is recognised as an important determinant of total physical activity, MVPA, and walking [58], [59]. Greater residential density may be important for MVPA and walking, but this could be dependent on the availability of other land uses in the neighbourhood, such as places to walk for pleasure. In contrast, while street connectivity may facilitate walking, connectivity alone may be less important for increasing levels of activity.

My findings for disturbance of the environment showed that those living in more polluted areas were less likely to record or report higher levels of MVPA and walking for pleasure but the associations were less consistent for total walking. These findings may be attributed to walking for transport which often takes place in inner city areas where walkability is high but concentrations of particulate matter are also highest [93], [129]. Although a relatively coarse measure of annual NO_X was used, few other studies have assessed the relationship between air pollution and physical activity. There is some evidence that exposure to air pollution may discourage other activities such as walking for pleasure [130] which is consistent with my findings for urban-rural status.

Greenness was associated with reported but not recorded activity. Although the number of studies using both objective measures of physical activity and greenness is limited, one other study found strong non-linear associations [128]. Those authors concluded that the greenness-physical activity relationship was weakened in areas of high walkability which may explain the lack of associations in my study. I found associations between steep terrain and walking for pleasure. As hilliness has rarely been assessed in the literature before, there are inconsistencies about the direction of association with different domains of activity [59], [60]. Although I cannot be certain why and where activity takes place, one possible explanation could be that participants with a preference for walking choose to live in hillier neighbourhoods or that activity in greener or hillier areas may be perceived to be longer due to aesthetics or a greater exertion of energy [131]. This potential effect of preferences or self-selection area warrants further investigation.

Most of the literature on environmental associations of physical activity is from the USA or other areas of Europe [14], [58], [59] and so the differences between my findings and previously published work may be due to differences in settings or the prevalence of baseline behaviours. Contradictory to current research [58], [59], my study suggests those living in more rural areas report higher levels of walking for pleasure, even after adjusting for area-level deprivation and income. Participants who lived in more deprived areas generally recorded lower levels of activity, however, the same group were more likely to report the highest levels of total walking, possibly having done so out of necessity rather than choice. This suggests that while the physical built environment may be a necessary condition for physical activity, it does not wholly explain the patterns shown. Residents must choose to use built environment features in health-promoting ways and these decisions are rooted in socioeconomic disparities in access [132] and elements of culture. Cultures create a structure for interpreting and participating in activities by defining resources and norms available in social settings [133]. My study showed a marked gradient between deprivation and walking for pleasure which may highlight a shared set of values and habits informed by local culture in relation to outdoor exercise. For particular populations, walking for pleasure may be considered elitist or distant while widely accepted by others – constraining or enabling the available choices for physical activity [133]. Decisions to engage in specific behaviours therefore, are structured not only by the built environment and socioeconomic factors but reinforced by place-based culture in which individuals reside [132], [134].

2.4.3 Strengths and limitations

The key strengths of the study were the large sample size and the combination of objective and self-reported measures of activity which allowed me to examine and compare different environmental associations for global and specific outcomes. Whilst similar studies have used objective physical activity data from multiple countries [97], they use data from one locality within each country and focus on a single outcome. Geographically heterogeneous data from across the UK were used. Recognising the importance of understanding a range of place-based determinants of health, I included immediate and contextual characteristics of the residential neighbourhood, organised around five facets important for physical activity and public health.

An additional strength of the study was the ability to examine environmental characteristics simultaneously and control for potential confounders. It is widely acknowledged that environments do not exist in isolation and an existing body of literature by Richardson and colleagues considers multiple environmental factors, including air pollution, as a composite index measuring 'environmental deprivation' [135], [136]. Studies which use the index suggest that environmental deprivation is associated with income deprivation and area-level health [137] and where the environment is adverse, higher income individuals are more likely not to choose active travel [138]. In my study, unique contributions are shown for different environmental variables which cannot be elucidated from an index. The study therefore builds on the existing literature by providing insight into the specific types of concurrent environmental strategies that may be employed to benefit public health. For example, schemes to improve the walkability of more deprived neighbourhoods may lead to increased levels of utilitarian walking, however, these could be combined with traffic calming schemes to reduce air pollution and its associated affects.

Limitations of my study include the use of cross-sectional data meaning that causal inferences cannot be made and there is a risk of reverse causation. Although the sample is uniquely large and heterogeneous [139], the included sample contained a high proportion of urban dwellers, homeowners, and participants educated to degree level. Participants were also clustered spatially with largely rural areas such as the East and South West of England, Scotland and Wales underrepresented (Figure 2.1). As assessment centres were restricted to urban areas, accessibility may have been an issue for certain rural populations which potentially influenced findings. For example, positive associations shown for rural dwellers and overall physical activity and hilly neighbourhoods and walking for pleasure may be have been strengthened as a result of healthy volunteer selection bias. Although the UK Biobank is not representative, the dataset has shown to be valid for assessing associations [140]. Despite this, there is still a risk of self-selection bias with more active participants possibly choosing to participate in objective monitoring or to live in environments matched to their preferences for activity. Unfortunately, I had no further information on this.

Measures of the environment were limited to static neighbourhood exposures. As there were no data available to locate physical activity or to describe environmental characteristics around other daily anchor points, such as the workplace, it was not possible to capture exposure to environments outside of the neighbourhood where participants may be active. These unmeasured exposures may lead to the Uncertain Geographic Context Problem (UGCP) and residual confounding [141]. Using previously derived data also meant that the accuracy of the underlying data is unknown and there is a temporal and spatial mismatch across variables. However, the categorization of exposures helps to minimise the risk of misclassification and the sensitivity analyses showed the size of the neighbourhood investigated made little difference to the pattern of findings.

The analyses of recorded and reported activity were not contemporaneous and used two different samples. Despite differences in age at times of assessment, the proportion of the samples employed and in retirement are similar which suggests the samples are comparable. To ensure characteristics of the neighbourhood were classified correctly at the time of assessment, where information was available, I removed participants who had moved home. As the number of movers was small, it is likely that the effect of any misclassification will be minimal. This information was not available for the entire cohort but will be in time.

Lastly, we would expect individuals sharing the same household to be similar and to depend upon and influence one another, but within-household clustering was not accounted for in the analysis. Address or household information was only available coarsely by coordinate estimates which made relevant fixed-effects variables difficult to derive. Including participants in the same household may have strengthened effect sizes in either direction. However, as the aim of the study was to identify general patterns rather than specific effects, and the large sample size, the reported findings and their interpretation remain useable.

2.4.4 Future research

Further investigation into activity domains and behaviours in relation to a range of environmental characteristics is required. Applying methods to identify specific activity behaviours from objective data will allow for these relationships to be explored further and with more confidence. The use of large-scale GPS data will also enable assessment of exposures and activity locations within and outside the neighbourhood. Combining objective measures with qualitative evidence on perceptions of space, such as aesthetics and safety, is also important for understanding how and why environments are used for physical activities. Lastly, longitudinal study designs are encouraged to understand how changes in the environment impact physical activity and to advance the field and guide interventions.

2.4.5 Policy implications

Modifying attributes of the physical environment may promote changes in physical activity. However, the evidence highlights the potential complexity in designing neighbourhoods to support physical activity and encourage wider health benefits. My study is one of the first to investigate air pollution in relation to reported and recorded physical activity. In doing so we see that whilst walkable neighbourhoods may encourage activity, particularly total walking, higher levels of walking are associated with participants living in areas with higher concentrations of air pollution and in more deprived areas. Consequently, an environment conducive to walking may not have the greatest overall benefit for physical activity or health given the adverse effects of greater exposure to air pollution and social inequalities. While modifying neighbourhoods to support physical activity may ultimately lead to sustained population changes, interventions which focus on a single characteristic of the environment or physical activity outcome are unlikely to have the greatest benefits. Instead, it is recommended that comprehensive strategies be employed to address a range of environmental characteristics in combination with careful consideration of the trade-offs for people and places.

Chapter 3

Activity spaces in studies of the environment and physical activity: a review and synthesis of implications for causality

3.1 Introduction

3.1.1 Chapter overview

Chapter 2 focuses on environments around the home address. However, such measures do not capture all possible locations where physical activity may take place. The activity space provides a more representative assessment of environments to which an individual might be exposed to by measuring a range of spaces experienced as a result of individuals' daily activities. This chapter reviews literature which applies the concept of the activity space to assess the association between the environment and physical activity. A version of the systematic review has been published in Health and Place [142].

3.1.2 Background

Physical activity reduces the risk of chronic disease [4], [143] and a substantial proportion of the population would benefit from being more active [144]. Public health strategies increasingly identify the environment as a modifiable determinant of activity. For example, the World Health Organisation Global Action Plan on physical activity identifies the importance of improvements to walking and cycle networks, road safety, and access to public open spaces and the need to understand where people live, work and play for their effective delivery [47]. Previous studies that investigated the relationship between characteristics of the environment and activity have predominantly examined the residential neighbourhood, applying static administrative boundaries or buffers around participants' addresses [58], [63], [145]. These assessments do not characterise the spaces within which people actually move and are exposed to, or account for within- and between-person heterogeneity in spatial habits [70]. Furthermore, assessing the environment around the home address can create spatial and temporal uncertainty relating to actual exposure (the uncertain geographic context problem (UGCP)) because it is unknown how much time people spend in those environments [73].

One concept which aims to more accurately capture exposure to different environments is the activity space. The general principle of an activity space is that it provides a dynamic measure of the environment by describing the locations and spaces an individual interacts with as a result of their activities [146]. It is organised around key anchor points including home and work locations and extends to locations such as food outlets, child's schools, parks, and social meeting points [147]. Locations may be weighted by the frequency, regularity, and duration at which they are visited [70]. The concept of the activity space was introduced in 1970 when space-time geography was used to assess daily travel behaviours [148] and has since been applied in a number of disciplines including transport, psychology, and food environments, using methods including diaries, GPS devices, web mapping applications and interviews [149], [150]. With an increasing shift towards objective assessment of activity and behaviours, the number of studies using GPS devices to examine the associations between environments and activity has grown in recent years [67], [151], [152].

Chaix and colleagues recognise that studies applying the concept of the activity space have the potential to strengthen the basis for causal inference between the environment and physical activity, if the methods and research question are thoughtfully implemented [152]. Some methods used to derive activity spaces capture environments potentially accessible to an individual over time but also capture environments regardless of a person's use of, awareness or exposure to that environment. Other methods describe places visited or spaces used for physical activity. However, an individual's preference to perform an activity may bias any associations observed between accessibility to these environments and physical activity because people who want to be active seek out environments or locations supportive of activity in order to be active. Using spaces used for activity as a measure of accessible environments gives rise to a circular argument as an individual would not have visited the location if they did not intend to be active there. This circularity may lead to a form of confounding called selective daily mobility bias which is likely to generate problems for causal inference because certain individuals may appear more exposed and any relationship between accessibility to these spaces and physical activity behaviours may be strengthened [152].

A previous review described the origins of the concept of an activity space, the categorisation of the disciplinary research areas, and the methods used [153]. This previous narrative review is not systematic and does not clearly outline the approach used to search for articles. It is also mainly descriptive and does not develop or evaluate the concepts in depth. Consequently, there is a gap in understanding how the activity spaces have been applied and used in existing research.

3.1.3 Aims and Scope

In this chapter, I perform a systematic review with the aim to examine the application of the activity space in studies of the environment and physical activity, identify what methods have been used, the research questions addressed, synthesise the methodological, analytical and conceptual issues, and assess the extent to which they strengthen the basis for causal inference.

3.2 Methods

3.2.1 Search strategy

To inform the design and scope of the search strategy and inclusion criteria, I completed pilot searches. Existing systematic reviews relating to GPS-located physical activity were identified [151], [154]–[157] and key words and terms of interest extracted from each. I then tested terms relating to themes common to the reviews (GPS, environment, activity space and physical activity) in PubMed. Different combinations of the themes were tested to understand if any search terms limited the results. Titles of studies from the searches were screened and potentially relevant articles were shortlisted and reviewed by all authors to inform the final search strategy and inclusion criteria.

After reviewing the outputs from the pilot searches and identifying articles which are suitable for achieving the aims of the review, the GPS theme was extended to capture other mapping methods and an additional theme relating to health and behaviour outcomes other than physical activity was added. The purpose of the additional subheading was to capture studies that apply the concept of the activity space, based on individuals' movement and accessible or accessed spaces, in relation to general health and socioeconomic influences of health.

Combinations of the search themes were tested with different Boolean operators (AND/OR) to identify which combination captured a broad range of studies, including all relevant studies identified from the pilot searches, and which returned a manageable number of results to sort through manually. In brief, using the activity space theme as an OR term returned too many records to manually sort through and using the environment theme as an OR term introduced a large number of irrelevant studies. The broad environmental terms were therefore omitted from the final search in favour of the more specific sets of terms under the mapping and activity space themes.

In January 2018, systematic searches of seven electronic databases were completed to identify potential literature, searching for articles published before 20th January 2018 (PubMed, Web of Science, TRID (Transport Research International Documentation), Scopus, Ovid MEDLINE, ProQuest, NICE (National Institute for Health and Care Excellence) evidence search). Google Scholar was proposed as an additional database, but the number of returned studies were unmanageable and the diversity of the topic was likely to have been captured by the general health and multidisciplinary databases.

The finalised search terms were based on four themes (1) mapping; (2) activity space; (3) physical activity; and (4) health/behaviour (Table 3.1) and the full search strategy as implemented in PubMed is detailed in Figure 3.1. In each section I included both broad (e.g. physical activity; exercise) and more specific search terms (e.g. walking; cycling) to ensure good coverage. Given the volume of outputs from each database, restrictions were applied to limit the results to studies of human behaviour in the multidisciplinary databases.

Table 3.1: Search terms

Theme	Search terms
(1) Mapping	GPS, GIS, map, behavioural geography, context
(2) Activity space	Activity space, potential path, daily path, destinations
(3) Physical Activity	Physical activity, walking, cycling, exercise, transport, mobility, movement
(4) Health/behaviour	Spatial behaviour, health behaviour, community, social cohesion
Search query:	1 AND 2 AND (3 OR 4)

((((((GPS OR global positioning system OR GIS OR geographical information system OR map OR mapped OR mapping OR behavioural geography OR context))

AND

(activity space OR potential path OR daily path OR destinations)))

AND

(((physical activity OR walk* OR bicycl* OR cycling OR exercise OR transportation OR mobility OR movement OR sport OR MVPA OR activity))

OR

(Spatial behaviour OR health behaviour OR community OR social cohesion)))

AND Humans[Mesh]))

NOT (chromosom*[Title/Abstract] OR hippocampus[Title/Abstract] OR nervous[Title/Abstract] OR genetic*[Title/abstract] OR cortex[Title/Abstract] OR cortical[Title/Abstract] OR chemical[Title/Abstract] OR receptor[Title/abstract] OR tumor[Title/abstract] OR tumour[Title/abstract])

Figure 3.1: PubMed search strategy

To identify additional relevant literature, eligible articles were forwards and backwards referenced searched by reviewing reference lists and papers that cited included studies. I contacted the first and last authors of eligible articles with multiple publications (n=10) via email and asked if they were aware of any other eligible articles. I also searched past editions of the GPS-Health Research Network (GPS-HRN) newsletter and emailed the editor to identify other relevant studies.

The protocol was registered with the International Prospective Register for Systematic Reviews (PROSPERO) in January 2018 (record number: CRD42018087095) [158].

3.2.2 Inclusion criteria

As I sought to understand how activity spaces had been used in studies assessing the association between the environment and activity, I took an inclusive approach embracing evidence on one potential causal pathway, drawing on the principles of a perspective articulated by Zenk [159]. This pathway might work in the following way: environmental characteristics in areas where people spend time might be associated with use of those environments (often captured through activity spaces) which might be related to levels of physical activity and subsequent health. Therefore, studies assessing environments exposed to as a result of use of space, characteristics of that space, physical activity, activity behaviours, or health outcomes were included.

All types of study designs were included, but studies had to use a spatial summary measure of movement, behaviour, activity, or locations visited and explicitly geo-locate spaces visited. The unit of analysis had to be the individual level and unique spatial summaries (activity spaces) must have been derived for each study participant. Locations could be self-reported and subsequently mapped or directly inferred from objective measures, such as GPS devices. Sub-sets of behaviour such as walking or trips made for a specific purpose were also included. No date, location, age, sample size, language or quality restrictions were applied.

Chaix and colleagues previously identified studies which assess the distribution of activity in different types of spaces or land use types, such as the time spent active in schools or parks, as descriptive and potentially limiting [152]. To focus the review, I excluded these descriptive studies. I also excluded studies that modelled or estimated routes, such as those that assumed individuals took the most direct route between two destinations, or described possible methods and did not apply them in an empirical study.

3.2.3 Study screening and data extraction

As the lead author, I (LS) screened titles and abstracts for eligibility. In phase 1 of screening, all articles with obviously irrelevant titles and abstracts were excluded. In phase 2 of screening, consideration was given to the definition and concept of the activity space by all members of the review team (LS, LF, and JP). All articles considered to provide potential context or methods of interest were initially grouped into one of six categories (Table 3.2). Articles in categories 1 to 5 were excluded and I retrieved and reviewed the full text of all articles in category 6.

Table 3.2: Study categories identified from phase 2 of screening

Category	Description
1	Studies where there was no variation in access for participants: • Those that assessed activity within a constrained area, such as internal environments (shopping centre, care home) or housing developments • Those that used non-continuous and infrequent locational data such as the mapping of social media check-ins • Those that described activity by environmental feature or land use type with no spatial summary
2	Studies that assessed non-physical activity outcomes where the causal pathway between environment and physical activity was unclear
3	Studies that assessed populations with long term limiting health conditions such as those with mobility disabilities or visual impairments
4	Non-empirical studies such as systematic reviews
5	Studies that modelled transport, such as those that map taxi or freight journeys, following the vehicle's route rather than the individual's
6	Empirical or methodological studies relating to activity spaces, environments, and physical activity

Language translation programmes were used and expertise from colleagues was sought to translate articles not written in English. 20% of articles identified for full text review were screened independently by LF for agreement. Reasons for exclusion were recorded by both LF and myself and any discrepancies in agreement were referred to JP for a majority decision.

I extracted information on study design, sample characteristics, research questions, activity space delineations, exposure and outcome measures, key findings and conceptual discussions related to causality from eligible articles into pre-designed forms (Table 3.3). In doing so, the terms and delineations used by the original authors were extracted. LF checked data extractions for accuracy and completeness for 20% of articles.

3.2.4 Data synthesis

I categorised studies according to the spatial and temporal methods used to define and delineate activity spaces, research questions addressed, and how activity spaces were applied to investigate which parts of the potential causal pathway between environmental exposure and physical activity. These categories were informed by data extracted from the studies and were designed to provide insight into the ways the activity space had been applied practically and conceptually. I synthesised results narratively to understand the consideration given to causal inference framed by Bradford Hill's principles of causation [160] and to identify any gaps in the field. Although other statistical frameworks for causality were considered [161], [162], the principles from Bradford Hill were chosen due to their broad nature and relevance to epidemiological studies.

Table 3.3: Data extraction form and example study

				Population					Study type				AS measure			
Author	**Year**	**Title**	**Link**	**Age group**	**Years**	**Sub-group**	**Location**	**Sample size**	**Data source**	**Design**	**Intervention**	**Analysis**	**Movement**	**Delineation method**	**Location data**	**Definition**
Lee	2016	Does activity space size influence physical activity levels of adolescents? A GPS study of an urban environment	[Articles\Lee_2016.pdf]	Adolescent	13.8 ±0.6	N/A	Downtown Vancouver, Canada N America	39	Active Streets, Active People-Junior study, 2012	Cross-sectional	No	Quant	All trips	DPA (200m buffer) for each person-day	GPS	Exposure: Daily AS area

Exposure			Outcome						
Description	**Measure**	**Method**	**Description**	**Measure**	**Method**	**Research Qs**	**Main findings**	**Biases and limitations of AS discussed**	**Additional notes**
Daily AS area	Objective	*See AS measures*	Daily MVPA (mins/day): Total, school day, trip based	Objective	Accelerometer	Is the size of activity spaces (geographic coverage of daily travel) associated with moderate-to-vigorous PA (MVPA; min/day) amongst adolescents?	There was no association between activity space size and school-day MVPA. School and school travel are important sources of PA in Vancouver adolescents, irrespective of activity space area covered.	No limitations of AS as a concept discussed	

3.3 Results and discussion

3.3.1 Study selection

The electronic database searches returned 12 982 records and 25 records were identified from GPS-HRN newsletters. After screening titles and abstracts and categorising articles of potential interest, 296 were identified for full text review. LF reviewed 20% of the full text articles with a 92% agreement rate. Five articles were referred to JP; there were no patterns in the reasons for referral. Three articles were identified from forwards and backwards reference searches. Eight out of ten experts responded and did not provide any additional eligible articles. In total, 47 articles met the inclusion criteria. The process of article inclusion and reasons for exclusion are detailed in Figure 3.2 and a list of included articles and a table showing the research questions answered and methods of assessment are provided in Appendix B1.

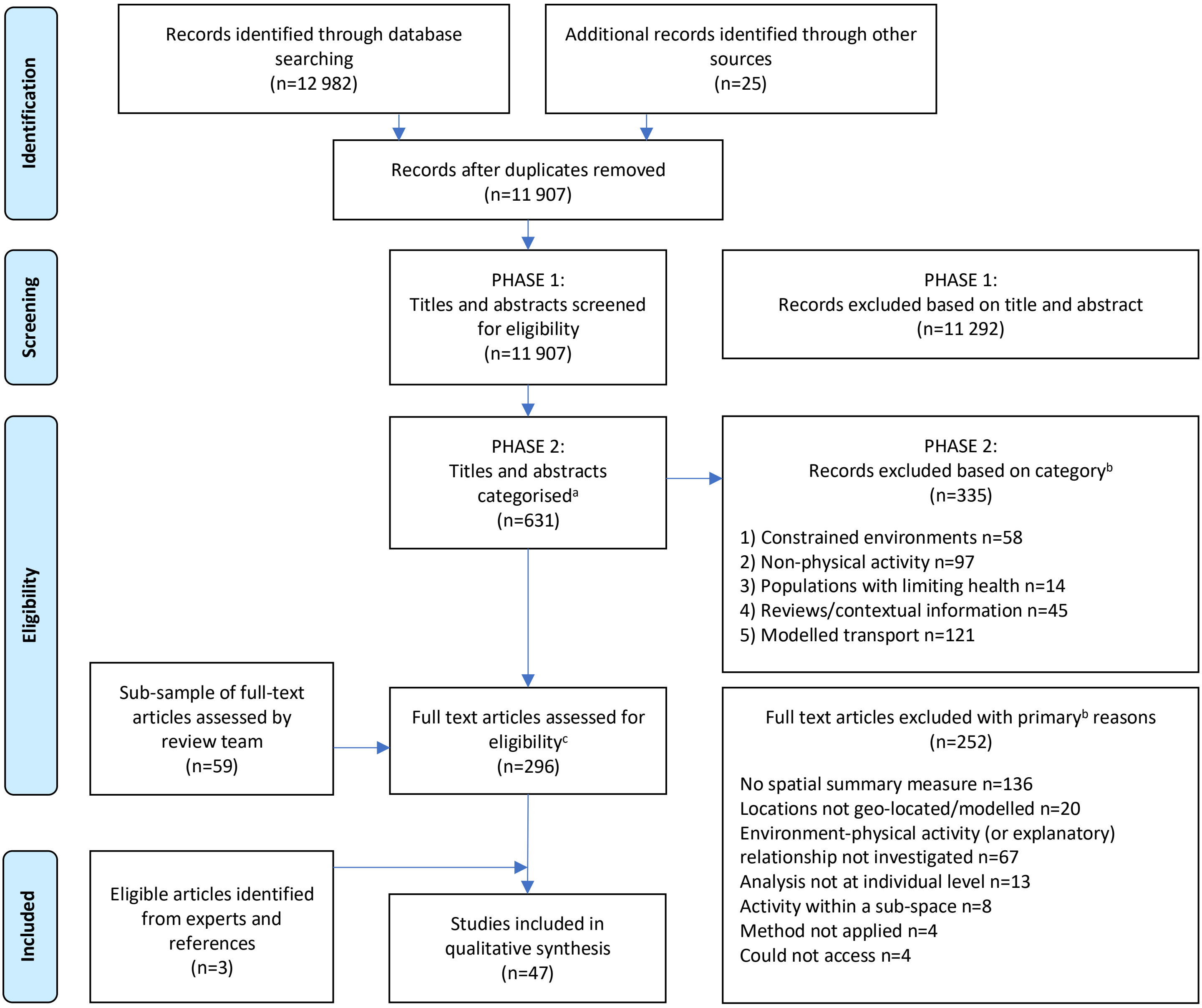


Figure 3.2: Study selection

[a]See Table 3.2 for details on categories

[b]Some studies met multiple criteria for exclusion. Categories and reasons for exclusion were ordered and only the criterion of highest order is shown

[c]All articles from category six (see Table 3.2)

3.3.2 Study characteristics

All articles were published after 2007 with 25 published within the past three years (2016-2018). The majority of study populations originated from high income countries, primarily from cities or metropolitan neighbourhoods in North America (n=24) and Europe (n=17). One study was identified from a middle income country, drawing on a sample from 28 villages near one city in India [163], and one studied rural dwellers from three towns in Northern Ireland [164]. Samples were studied for all age groups with most drawn from adult (n=22) populations.

Some targeted females [165], [166], lower income participants [167]–[169], university members [170], [171], e-bike owners [172], or those living in subsidised housing [173]. Most studies were solely cross-sectional in design (n=43) and four assessed activity spaces in relation to an intervention [171], [174]–[176]. All intervention studies examined alterations to the built environment including access to a demand responsive transport service, improvements to street safety, a covered walkway, and a modelled increase in services in the residential area. Study characteristics are detailed in Table 3.4.

Table 3.4: Summary of study characteristics

Characteristic	Reference	No.
Continent		
N America	[112] [159] [165] [166] [167] [168] [169] [173] [176] [177] [178] [179] [180] [181] [182] [183] [184] [185] [186] [187] [188] [189] [190] [191]	24
Europe	[164] [170] [171] [172] [174] [192] [193] [194] [195] [196] [197] [198] [199] [200] [201] [202] [203]	17
Australia	[204] [205] [206]	3
Asia	[163] [175] [207]	3
Age group		
Children	[177] [181] [184] [186] [187] [192] [193] [194] [203] [204] [205]	11
Adolescents	[165] [166] [173] [179]	4
Adults	[112] [163] [164] [170] [171] [172] [174] [175] [176] [178] [182] [183] [188] [189] [190] [191] [195] [196] [197] [200] [201] [202]	22
Older adults	[167] [168] [169] [185] [198] [199] [206] [207]	8
All	[159] [180]	2
Sample size		
0-50	[163] [169] [172] [179] [187] [185] [192] [197] [198] [199] [200] [205] [206] [207]	14
51-100	[159] [167] [168] [175] [177] [183] [193] [194]	8
101-500	[164] [165] [166] [170] [171] [173] [174] [176] [181] [195] [196]	11
501-1000	[112] [182] [184] [186] [189]	5
1001-5000	[180] [201] [202] [203] [204]	5
>5000	[178] [188] [190] [191]	4
Study design[a]		
Cross-sectional	[112] [159] [164] [165] [167] [168] [170] [171] [172] [173] [174] [175] [177] [178] [179] [180] [181] [182] [183] [184] [185] [186] [187] [188] [189] [190] [191] [192] [193] [194] [195] [196] [197] [198] [199] [200] [201] [202] [203] [204] [205] [206] [207]	43
Longitudinal	[163] [169] [176]	3
Both	[166]	1
Intervention	[171] [174] [175] [176]	4
Analysis		
Qualitative	[172] [185] [197] [198] [199] [200] [207]	7
Quantitative	[112] [159] [163] [165] [166] [167] [168] [170] [171] [173] [174] [175] [176] [177] [178] [179] [180] [181] [182] [183] [184] [186] [188] [189] [190] [191] [192] [193] [194] [195] [196] [201] [202] [203] [204]	35
Both	[164] [169] [187] [205] [206]	5

[a]Column totals to more than 47 as some studies listed in more than category and categories not mutually exclusive

3.3.3 Methods employed

3.3.3.1 Spatial extent of activity space

Activity spaces were derived from objective GPS data (n=30) and reported locational data (n=24). Seven studies used both. Regardless of the method used, all studies assessed the spatial extent of activity in one of three ways: (i) by using all movement, (ii) by focusing on key locations visited or (iii) by focusing on specific routes or activity types (Figure 3.3). There was a range of different methods employed within each broad grouping and sometimes within a single study.

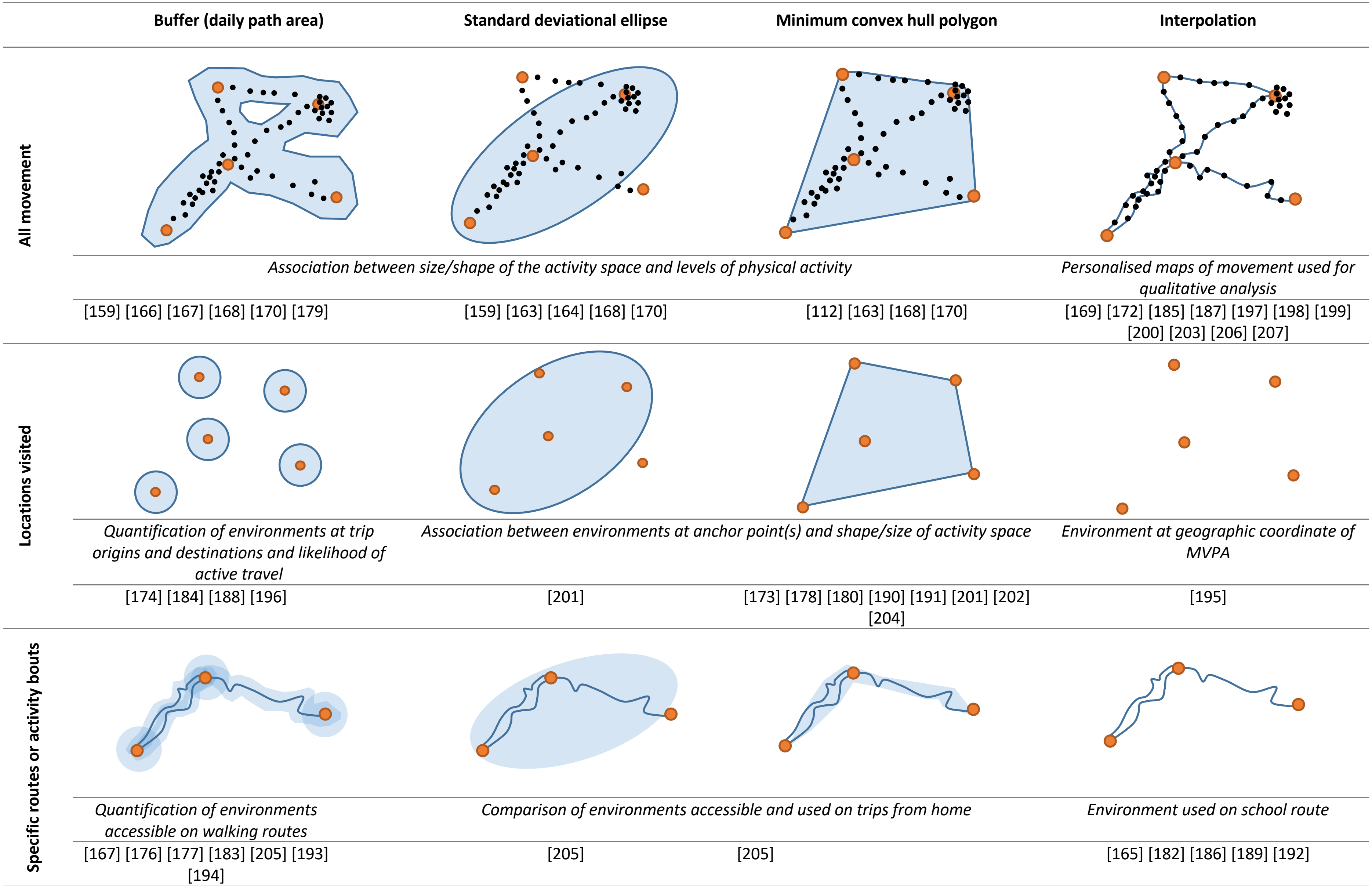


Figure 3.3: Methods used to delineate activity spaces with descriptions of example applications

Anchor point (for example: home/work/school/sports club location) • / Geo-located movement Activity space

(i) All movement (n=20 studies)

Methods included daily path areas (DPA) (buffer of all points or tracks) (n=6), standard deviational ellipses (SDE) (n=5), minimum convex hull polygons (MCP) (n=4), and personalised maps of plotted points or tracks of movements (n=11). Some examples of the latter indicated areas accessed using different travel modes [172], [206] or for physical activity [207]. Unique estimations of the activity space also used a maximum path distance to recorded points of movement [181] and a composite measure of distances travelled and frequency at locations [171]. There was little consistency across delineation methods, for example, DPA buffer sizes ranged from 50 m [166] to over 800 m [159], one study added an additional 20 m buffer to an MCP [112], and one [159], [163], [168] or two [164] standard deviations were used for SDEs.

(ii) Key locations (n=13 studies)

For these studies, key locations were used to define the activity space. Locations included trip origin and destinations [174], [188], [195], [196], destinations actively travelled to [204], locations for activities [173], [178], [180], [190], [191], [201], [202], and home and school addresses [184]. Measures did not capture movement between locations. MCP was commonly used to delineate activity spaces in these key location studies (n=7). Four studies used a buffer of point locations, one of which was radial [184] and three were network [174], [188], [196], one study used an SDE [201], and one interpolated from GPS coordinates [195].

(iii) Specific routes or activity types (n=12 studies)

These were typically assessed using a buffer (n=7). Buffers of active trips or routes to and from home or school were generally smaller than those employed for other movement limits, ranging from 50 m to 500 m. One study used MCP and SDE measures to summarise the space used to make trips to or from home [205] and five studies interpolated environmental characteristics directly from point or polyline locations, describing the locations used for physical activity or passed on route [165], [182], [186], [189], [192].

3.3.3.2 *Temporal extent of activity space*

Three key elements in relation to temporality were considered when deriving activity spaces.

<u>(i) Scale of data accumulation</u>

Most activity spaces were delineated using data accumulated at the trip (n=16), day (n=13), or multi-day (n=18) level. The majority of studies which assessed activity spaces at the day level used a separate measure per person per day but two studies used an average measure over a number of days to determine a mean daily activity space [178], [191]. Where specified, the minimum number of days required for participants to be included in analysis ranged from one to four. Six studies used reports of usual places visited or routes used, geo-located these and then derived activity spaces [184], [186], [201], [202], [204], [207]. Usual places were defined as those visited on a regular basis [184], [186], [204], were meaningful to the participant [207], were visited at least once a month [202], or had varying frequency depending on the type of destination (at least once a week, except for workplaces and supermarkets which were required to be visited for at least one third of the week or once a month respectively) [201].

The level of data used, whether a single trip or day, several days, or usual, appeared dependent on the research question and whether the activity space was used as an exposure or outcome. Example applications are detailed in Table 3.5. Often the level of aggregation was related to the temporality of other measured variables, for example, if step counts were investigated, data were often accumulated at the day level. However, authors did not always make clear the level at which data had been accumulated and justification was rarely provided. Three qualitative studies recognised that temporality may be important and investigated activity spaces for a trip, day and over several days [197], [199], [206]. These studies aimed to discuss specific spatial patterns of activity bouts and daily routines, how these contributed to general use of space for frequent and occasional activities over a number of days, and suggested different associations based on the temporal scale of the data.

Table 3.5: Scale of temporal data

Level of data accumulation for deriving activity space	Example application
Trip	Estimate environmental characteristics on the route to school and investigate their associations with travel mode used on route [193]
Day	Based on all activity location visited per day, estimate environmental characteristics within the activity space and their associations with the area of the activity space [180]
Multi-day	Based on all trips made over the course of several days, estimate walkability within activity space and investigate association with total weekly minutes of moderate physical activity [112]
Usual	Identify 'regular' locations visited by individuals and estimate the size of the activity space and its association with active trips made [201]

Multi-day level measures aggregated movement data collected over a number of days.

<u>(ii) Weekday and weekend</u>

Differences in behaviour and extent of activity for weekdays and weekends were accounted for by relatively few studies. This may be due to the limited number of days' worth of data collected, however, there is evidence of key differences in physical activity across these times [208], [209]. Only three studies quantitatively investigated the differences between weekday and weekend activity spaces [164], [189], [191] with one finding that utilitarian walking was more likely to occur on weekdays than recreational walking [189]. Two of the qualitative studies described how activity was patterned by day [197], [198] and found unique activities occurring at the weekends [197] and differences in times and geographies of older adults' mobilities over different days depending on factors such as weather and availability of family [198].

Seven studies limited their investigations to weekdays as a way to capture school or work-related behaviours and two studies acknowledged the day of the week as a potential confounding factor in their analysis [166], [183], however, most studies did not account for or investigate differences in activity by day of the week.

<u>(iii) Exposure weighting</u>

The extent of an individual's environmental exposure varies by the proximity to or amount of time spent in a location, and the type of activity being undertaken [68]. For example, the same space may be experienced for longer and more closely when walking compared to when driving.

Some studies accounted for this by weighting the exposure of an environmental characteristic within an activity space. Rudimentary examples of this included limiting analysis to only locations that are frequently visited [181], [183] and investigating environments experienced during a single behaviour such as walking [165], [176], [182], [189]. These studies assumed that the environments in which participants spent most time are the most important exposures and that environments are experienced in an equal way when undertaking the behaviour of interest. Some qualitative studies recognised clusters of activity on maps of individuals' movements and discussed reasons for 'lingering'; identifying functional and emotional connections to regular and unique places visited [197], [200]. More complex approaches used to weight exposures included cell [193], [194] and inverse distance weighting [177] whereby a distance decay effect between individuals' recorded locations and proximate environments was applied, and a kernel density approach which weights locations by the density or duration of activity taken place there [159], [171], [174], [180], [188], [196].

Those studies that attempted to apply weighting techniques provided a methodological step in accounting for temporal dimensions of an individual's environmental exposure. However, the majority of studies did not consider the duration of time in any area, with some averaging measures across bouts or routes so that exposures received equal weight, irrespective of time spent in them [165], [186], [189]. The frequency of visits to key locations was not always measured and no studies assigned different weights depending on the behaviour being undertaken. Although weighting methods were relatively uncommon and inconsistent, they highlight potential ways to capture the density of activities or identify which environmental characteristics are most strongly associated with physical activity.

3.3.4 Research questions answered

In describing the research questions answered by these studies, and considering my interest in eliciting how activity spaces have been used to strengthen causal inference, I categorised studies according to the research questions as they related to a possible hypothesised causal pathway (Figure 3.4). This pathway might work in the following way: characteristics of the environment might influence where people are active or spend time (captured through activity spaces) which might be related to levels of physical activity and subsequent health outcomes. Studies using activity spaces addressed different research questions which mapped on to different areas of the causal pathway: studies assessed either the extent of movement by assessing the *features or parameters of the activity space itself* or used the delineation of the activity space to measure environmental *features within the activity space*.

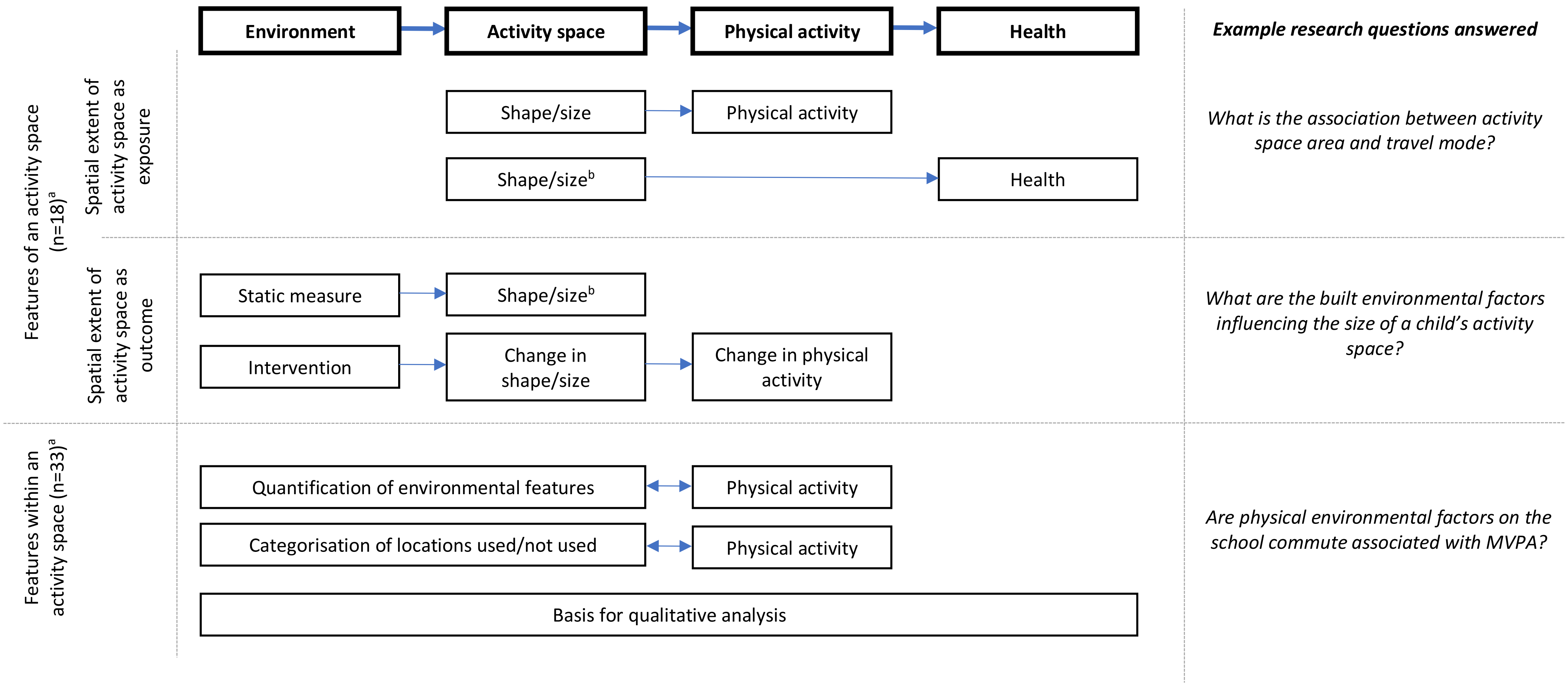


Figure 3.4: Conceptual framework of research questions answered

[a]Totals to more than 47 as some studies address both types of research question

[b]Some studies used shape and size as an indicator of physical activity

3.3.4.1 *Features of an activity space*

Eighteen studies assessed the shape and/or size of the activity space, measuring the area, perimeter or compactness as an independent or dependent variable, or moderator. Shape was measured using SDEs whilst size was derived from polygons of activity spaces. Perhaps the most straightforward application was the use of activity space parameters as an independent variable to assess the spatial extent of movement in relation to physical activity outcomes [179], [187], [190], [201]. Two of these studies found that a smaller and more compact activity space may be related to an increased likelihood of active travel [190], [201], however, when assessing MVPA in adolescents, Lee and colleagues found that travel mode to school may be an important source of physical activity irrespective of activity space size [179]. An additional study used the size of the activity space as an indicator of physical activity and reported a weak positive correlation with perceived health [202].

Twelve studies used the activity space as an outcome to understand if access to different environmental characteristics influenced the extent of mobility and space used. Some used measures of the built environment as an independent variable and typically found that access to denser characteristics of urban form, such as walkability and connectivity, was associated with smaller activity spaces [163], [168], [178], [180], [191]. Studies investigated and adjusted for sociodemographic factors and one considered the effect of weather [180].

Similar studies built on this relationship and investigated the role of physical activity by incorporating travel mode into their independent variables [164], [170], [181] or by investigating the moderating role of public transport services [164], [171]. In the latter studies, Kamruzzaman and colleagues reported that environment-activity space relationships are sensitive to the accessibility of public transport services, car ownership and day of the week, which may be indicative of fewer travel needs or more constraints at weekends. Developing these hypotheses further, one study reported an inverse association between the presence of utilitarian destinations and activity space size but no association between activity space size and steps in school children [204]. The lack of association observed is in contrast to other work [190], [201] but aligns with findings by Lee and colleagues that suggest the size of the activity space may not be important for increasing physical activity [179]. Despite the majority of these studies focusing on a narrow part of the causal pathway and differences in the strength of study design and questions answered, the general pattern of results when viewed together suggest that denser, more urban environments were associated with more contained activity spaces, and more contained activity spaces were associated with active travel.

One study assessed the area, shape, and overlap of walking-specific activity spaces with self-defined neighbourhoods before and after the development of a street design intervention designed to more safely accommodate all transport modes [176]. The authors found that walking activity spaces were significantly smaller than neighbourhoods, were included within but comprised a small proportion of the defined area, and became more compact following the intervention. Similar comparisons between activity spaces and neighbourhoods or 'potential' activity spaces (possible environments that could be used) are made elsewhere [112], [173], [204], [205] with reports of comparable findings that walking activity spaces are smaller than neighbourhood buffers used in walkability research [112].

3.3.4.2 Features within an activity space

The majority of studies used activity spaces as a way to quantify environmental characteristics that populations were exposed to and then investigated the relationship between these features and physical activity (n=33). The density or diversity of features as well as categorical descriptions of where activity had taken place (e.g inside/outside the residential neighbourhood) were used as independent variables. Measures of features were often derived from secondary and audit data, digitised in GIS, and quantified within activity space polygons or interpolated from points or routes.

The characteristics investigated varied across the studies although a number found that activity spaces with greater walkability, residential density, or utilitarian services were associated with higher MVPA and walking [112], [166], [174], [184], [195]. However, findings were not consistent and mixed observations were reported for greenspaces and street densities [183], [186], [188], [192]–[194]. Houston and colleagues found associations between environmental features and MVPA in adults were sensitive to the proximity of features to GPS locations [183] which suggests observations may be dependent on the method used, as well as heterogeneity in the populations investigated. One study investigated measures of the built environment within the defined neighbourhood and walking activity space and reported that cross-sectionally, the environments within the activity space were more strongly related to walking trips but that changes to environments within the defined neighbourhood were more important for explaining changes in the number of walking trips made [112].

Although not always made explicit in research aims, the nature of locational data used by a number of studies implied that they investigated the environments used for physical activity rather than those potentially accessible [166], [175], [182], [183], [186], [192], [193], [203]. For example, studies which characterised environments on a route described spaces that have been used for a specific behaviour or purpose. Quistberg and colleagues used the location of walking bouts to derive an activity space and estimated the risk of pedestrian collisions with motor vehicles with measures of walking [182]. The risk of pedestrian collision is defined as an outcome and findings suggest that participants walked for recreation in areas with lower risk. However, a more interesting research question from an epidemiological perspective may be to understand how exposure to collision risk affects levels of walking and health which could plausibly be deduced from the same methodology.

Some studies addressed differences between potential access and actual usage by capturing broader spaces experienced over a day or week to identify a range of spaces that may be accessible to an individual, or by comparing features within an activity space with those accessible from a home address [112], [159], [167], [196]. Ten studies used qualitative analysis to understand why particular environments were chosen for use [169], [172], [185], [187], [197]–[200], [206], [207].

3.3.5 Strengthening causal inference

The unit of analysis and method used to delineate an activity space gives rise to different strengths, limitations, and conceptual considerations and many studies used the most appropriate method to answer the specific research question addressed. In general, few studies discussed the implications for causal inference; however, many noted that the use of locational data beyond the residential neighbourhood was an important development in improving understanding of the causal relationships between the environment and physical activity. I used the most relevant aspects of Bradford Hill's principles of causation [160], to frame a synthesis of how issues were discussed and the strategies employed to deal with them. Here I focus on consistency, specificity, plausibility, temporality and experimentation.

3.3.5.1 Consistency

The broad pattern of results suggests that dense characteristics of urban form are associated with smaller activity spaces and higher levels of physical activity. However, there is a large degree of variation in the research questions answered, methods used to derive and summarise activity spaces, environmental features identified within activity spaces and associations with activity. Some studies assessed relationships between specific behaviours and micro-level features of the environment whilst others assessed more general patterns with regards to mobility. While some similarities in results were seen for studies that answered similar research questions using similar methods, as a whole there were mixed results across the entire body of literature identified.

The activity space can be used in a number of ways and applied within the same dataset to answer different but related questions. For example, Perchoux and colleagues investigated a range of questions by assessing features within the activity space and their association with the spatial dimensions of the activity space as well as assessing the features of the activity space with transport related outcomes [201]. Findings from the different questions were consistent; showing that higher levels of active transport were associated with smaller activity spaces.

3.3.5.2 Specificity

Delineations of the activity space typically drew on all movement or locations visited and provided little insight into how this relates to specific behaviours or whether spaces were used for physical activity. However, if the research question aims to understand how people use space, greater specificity of activity space measures might provide a stronger basis for causal inference. Daily path areas, particularly those with smaller buffer sizes, provide a more accurate estimation of space used than the SDE or MCP which can overgeneralise and lead to residual confounding [170]. Although these latter measures provide a useful measure of environments potentially accessible to the individual.

3.3.5.3 Plausibility and circularity

The use of the activity space reduces the spatial and temporal uncertainty relating to actual areas visited and time spent in locations compared to static measures of the environment, as described in the concept of the UGCP [73]. In most studies, the design limited the ability to understand whether spaces are used because they are supportive of a preferred activity or because they are accessible from an anchor point – the problem of selective daily mobility bias. Studies which interpolate environmental features from spatial data of a route or activity bout were at the greatest risk of selective daily mobility bias [165], [182], [186], [189], [192]. For example, McMinn and colleagues investigate what physical environmental characteristics are associated with MVPA on the school commute by assigning a land use category to GPS points [192]. Here, the direct environments used for travel are described, however, the environmental exposures are a direct result of individuals' travel choices which leads to an issue of circularity. Studies which use a summary measure of all locations visited provide a more plausible measure of environmental exposure, including characteristics which are both potentially accessible and used for physical activity. For example, one study characterised the percentage of parkland within a standard deviational ellipse and a buffered daily path area of all GPS trips made over the course of one week [159]. However, this measure is formulated around movement that has actually occurred and environments that individuals are exposed to as a result of their choices and does not provide any insight into where MVPA took place. Consequently, the basis for causal inference is low with respect to plausibility and circularity.

Nine studies highlighted selective daily mobility bias as an issue [159], [164], [167], [173], [177], [188], [195], [196], [201] and two tried to address this by comparing potential and actual routes taken [177], [196]. Where no significant differences were observed it was assumed that bias was minimised as route choices appeared not to be heavily based on preferences [177]. One study controlled for selective daily mobility bias by adjusting for residential and transport preferences, as well as modes used in previous trips taken as these were thought to influence characteristics of the place visited and mode used in present trip [196]. The authors characterised trip origin and destinations but not environmental conditions along routes. This reduces the issue of circularity and by considering all destinations, provides an advance on studies which investigate environments within a residential neighbourhood. Chaix and colleagues discuss the filtering of locational data to remove locations where physical activity occurs from measures of accessible environments to mitigate bias [152]. Although this could be achieved by combining different spatial and temporal methods that are present across the studies, none of the studies in the review have attempted this.

Activity spaces were rarely used to provide evidence of plausible *mechanisms* behind observed relationships, although a number of studies used qualitative data to understand why some spaces are used and others are not. For example, Hand and colleagues used go-along interviews and personalised GPS maps to shed light on person-place transactions and commented that quantitative data could be explored further to complement these findings [185].

3.3.5.4 *Temporality and experimentation*

Studies which used aggregated momentary measures of movement, such as GPS or travel diaries, captured all locations visited over multiple days (n=18). Conversely, some studies used self-reported measures of usual places visited explored those visited only on a regular basis (n=6). Whilst the momentary studies capture more specific locations visited, the shorter data collection period may mean that the general pattern of behaviours are not adequately represented. Both of these are valuable depending on the research question. The distinction between these methods and the behaviour of interest should be considered in future studies and the temporal dimension of activity spaces should be well-matched to exposures or outcomes and relevant for the research question.

Little consideration was given to different temporal scales and few studies weighted exposures by length of time or type of behaviour. Consequently, it is difficult to understand whether relationships are strengthened for more proximal or longer environmental exposure.

Only four longitudinal studies were included in the review [163], [166], [169], [176] which limits the causal inferences that can be made. Assessing changes in the environment, locations of activity, or anchor points over time may provide an understanding of whether this increases physical activity and could strengthen the basis for causality inference. One study used geo-referenced qualitative data to investigate why older adults chose to be active in different places [169] and another assessed temporal differences in associations between the built environment and MVPA in adolescent girls [166]. Both observed changes in physical activity and environmental interactions over time. However, neither considered displacement of activity due to a change in the environment.

There is an opportunity to use activity spaces in evaluative studies to complement assessments of physical activity. I identified four intervention studies which all examined built environment interventions despite the search strategy enabling individually delivered interventions to potentially be identified. A study to assess the effect of an intervention to promote activity could use activity spaces to understand if this has changed where activity takes place or if it has changed the types of activities undertaken or with whom. It might also provide validation that changes in physical activity were directly attributable to the intervention under study. This general approach was used by Kosaka and colleagues in their assessment of covered walkways [175] and by Kamruzzaman and colleagues who assessed if distance to a transport service affected the size of the activity space [171]. Although both examined an intervention, study designs were cross-sectional. One study assessed changes in activity spaces and walking in response to street safety developments [176] and it provided the strongest basis for causal inference due to its within-individual follow-up, controlled assessment of an intervention, and investigation of research questions relating to the features of and features within an activity space. Further evaluative studies of these types are required.

3.3.6 Recommendations for future work

My findings illustrate that the activity space can be used to characterise the environments which people are exposed to or engage with as a result of their activities. Both the features *of* and *within* the activity space have been shown to be associated with activity but more evidence is needed to establish the direction of the causal pathway and whether the relationship between potential accessibility to environmental features and physical activity behaviours are explained by use of space. Different but complementary research questions have been addressed and could be combined to advance the field. For example, separate methods to measure potential accessibility to environments and use of those environments could be used in the same study to answer research questions framed around understanding whether environments accessible to individuals are used for physical activity and what this means for overall activity levels.

There are a variety of spatial methods used to delineate the activity space as shown in Figure 3.3, but all studies within the review captured either all movement, key locations, or locations of specific routes or activity bouts. I recommend carefully considering the distinction between measuring environments that are potentially accessible to an individual from those which the individual is directly exposed to a result of their use and using methods appropriate for the specific research question. Some studies considered differences in access and use and go some way to reducing selective daily mobility bias by comparing the activity space to residential neighbourhoods or shortest routes [112], [173], [176], [177], [204], [205]. Further strategies to account for selective daily mobility bias may involve sensitivity analyses whereby separate analyses are performed for activity spaces including all behaviour and activity spaces where the behaviour or route of interest is filtered. Some authors commented on the need to understand why individuals may be active beyond their neighbourhood [176]. Future studies could improve the definition of accessibility and help unpack mechanisms to understand why some spaces are used and others are not by incorporating qualitative evidence or controlling for individuals' activity preferences.

Currently, there is little consistency in the application of temporal elements and more consideration could be given to weighting environments by their duration of use. Weights may be derived from a kernel density map of activity duration and types of activity and applied to measures of the activity space. The level of data accumulation used to derive the activity space should be appropriate for the outcome under investigation and it is important to analyse weekday and weekend relationships separately given observed differences in patterns of behaviour over these times. I identified relatively few longitudinal and intervention studies. Additional studies assessing the effects of environmental change are encouraged to strengthen casual inference and aid understanding of how interventions affect the spatial patterning of physical activity and whether levels of activity are increased, decreased, or displaced over time. I also recommend more studies in low and middle income countries to improve the generalisability of findings. This is important for understanding where physical activity occurs in different settings which could help to guide future interventions.

3.3.7 Strengths and limitations of the review

The strengths of this review include an extensive search strategy which was developed following an iterative process and applied to a range of specialised and interdisciplinary databases and having no restrictions on study type. The search process helped to develop and identify concepts of the activity space and the ways in which it has been used to answer questions about the relationship between environmental features, the use of those environments for activity and overall levels of physical activity. Although all included studies were in English, there was no language restriction and a number of full texts were translated to assess eligibility for inclusion. Studies published since the search may have been missed; however, the aim of the review was to describe general methods and conceptual issues which are prevalent in the literature, not to comprehensively search [210]. I discussed themes relating to study characteristics, methods and, conceptual issues in an emerging area of research that currently has little standardisation. In doing so, I highlight potential methods which could be used to answer important research questions to help researchers reduce issues of bias and strengthen causal inference, and ultimately guide future intervention research in the field.

The evidence reviewed here is complementary to evidence that describes which types of environments people are more active in, without producing a spatial summary measure. I excluded those types of studies to focus my review on the research questions addressed and methods of summarising spatial data. Some excluded studies might have strengthened the basis for causal inference in other ways, such as the use of more statistically complex models. I also focused on outcomes related to activity and excluded studies using activity spaces to examine associations between environmental exposures and outcomes not on the causal pathway, such as smoking or food-related behaviours [157], [177], and those relevant to other disciplines. In doing so, studies from other subject areas that might have important methodological contributions to add were not reviewed.

In ecology, telemetry data has widely been used to map spaces used for foraging and grazing by different species, as well as habitat preferences and the spatial limits of ranging behaviour [211]–[215]. Advanced spatiotemporal techniques including utilisation distribution (a probability function that maps an individual's relative use of space) and kernel density estimation (KDE) have widely been applied to identify the most common activities for species in specific locations [211], [213]. Similarly, techniques employed in social sciences extend those identified in this review. Space-time budget methodology has been incorporated into tourism, crime and urban planning research to analyse the timing, sequence, frequency and location of activity patterns [216], [217]. The role of the social environment and person-place interaction has also received attention in understanding exposure to criminogenic settings [218]–[220]. These methods show promise for advancing knowledge about the role of the environment in influencing behaviour. Although the potential of these methods was not explored in this review, as the amount and quality of geolocational data becomes increasingly available, it is important that health researchers draw on cross-disciplinary methods to more accurately measure exposure and activity patterns.

By focusing the review on physical activity outcomes, it was possible to outline progress and to identify gaps in the field. The review highlighted methodological limitations which suggests the field may lag behind other sciences with regards to the use of spatiotemporal methods. However, the review was not intended to solely advocate better measures or provide recommendations on methods for delineating activity spaces. Rather, it identified research questions which have been investigated and sought to understand how the activity space may be applied to strengthen the basis for causal inference in future research. The review explored key conceptual issues relating to spatiotemporal measures and causal inference – an area of relevance to other disciplines where understanding remains underdeveloped [213], [219].

3.4 Conclusions

The use of the activity space is an emerging methodology for advancing studies of environment-physical activity relationships which may also be relevant for outcomes from a variety of disciplines. A range of activity space types exist and the activity space used within studies was often subject to the availability of data and the research question which the authors aimed to answer. Activity spaces can be used as both exposures and outcomes on the hypothesised environment-physical activity causal pathway and questions may relate to either the features of an activity space or features within an activity space.

There is a need for greater consistency across study designs to enable comparisons and assessment of both potentially accessible spaces and spaces used for physical activity within the same study. Longitudinal data and evaluations of interventions enabled changes in the use of space and behaviours in response to changes in the environment to be investigated, and controlling for residential and travel preferences reduced selective daily mobility bias. Currently, the application of these strategies is limited which highlights the paucity in thinking about how activity space can be used to strengthen causal inference.

Chapter 4

Cleaning and preparing GPS data to derive activity spaces

4.1 Introduction

Activity spaces provide a dynamic measure of the locations and spaces an individual interacts with. As highlighted in the systematic review in the previous chapter, the activity space can help to improve understanding of environmental exposures and the types of places people spend time. Assessing this use of space and activity following a specific change in the environment may help us to understand whether there is an uptake of additional activity in new locations as a response to the change, or whether activity is being substituted or displaced from elsewhere. Chapters 4 to 6 utilise GPS data to address some of the methodological and conceptual limitations identified in the systematic review.

4.1.1 Chapter overview

In order to utilise GPS data effectively and to derive meaningful activity spaces, GPS data should be cleaned to remove points that have been erroneously populated or positioned, and prepared for analysis. This chapter outlines the development of methods to clean GPS data and to derive activity spaces. The prepared data will be used in subsequent chapters to analyse changes in spatial patterns of movement and physical activity in response to a change in the built environment.

4.1.2 Background: Using GPS data in health research

In order to apply the activity space concept, a measure of individuals' daily spatial behaviour is required. Whilst some studies have used self-reported methods such as map-based questionnaires to locate daily mobility patterns [221], an increasing number of studies rely on objective location sensing methods with the majority of studies in my systematic review (30/47) using Global Positioning System (GPS) data to derive activity spaces [142]. GPS sensors record spatial location data using signals transmitted from a network of orbiting satellites. Improvements in data storage and portability of GPS technology mean that sensors embedded in custom devices or smartphones provide an unobtrusive and convenient means of continuously tracking an individual's mobility patterns. GPS data have been used as the basis of a range of activity spaces which vary in design and scale, and to identify routes travelled, travel modes, and to provide environmental context for a number of health-related behaviours.

GPS data provide an advance on static measures of environmental exposures that focus, for example, on the residential neighbourhood. However, the review in Chapter 3 highlighted some conceptual limitations for its use with respect to determining causality. The use of GPS data therefore requires careful consideration for the methods used to delineate activity spaces in order to answer research questions effectively. For example, GPS data could be used to locate and compare users and non-users of an environmental intervention pre and post or combined with qualitative data to help understand why accessible environments are or are not used directly for activity. This helps to strengthen the basis for causal inference which is particularly important in natural experimental studies of environmental or place-based interventions where randomisation is not possible or ethical.

A review by Krenn and colleagues [151] identified studies that assessed the association between the environment and physical activity but focused solely on studies which used GPS. The findings corroborated with my review, indicating the capabilities of GPS as a tool for improving understanding of the spatial context of physical activity. However, the Krenn review appraised factors that influence the quality of GPS data and highlighted key technical limitations relating to data accuracy and data loss. The causes of these limitations, and ways in which they have previously been addressed are detailed in the following sections (4.1.2.1 to 4.1.2.3).

4.1.2.1 Data accuracy

GPS receivers require a direct line of sight with at least four satellites to determine spatial position. The reflection of GPS signal off nearby buildings, interruption of signal due to tree canopies or indoor environments, or limited satellite visibility can limit signal strength. Topography including urban canyons or dense foliage therefore cause a degradation of positional accuracy and data quality [222]. Comparing GPS estimated positions to known geodetics points under various environmental conditions, one study measured a mean error of 7.3±27.7 m under open sky and a mean error of 59.2±99.2 m between high rise buildings [223]. The same study compared seven portable GPS sensors and found the degree of positional error to vary across different sensors [223], ranging from 12.1±19.6 m to

58.8±393.2 m. This can lead to invalid positioning of points which I herein refer to as *signal stray* (Figure 4.1).

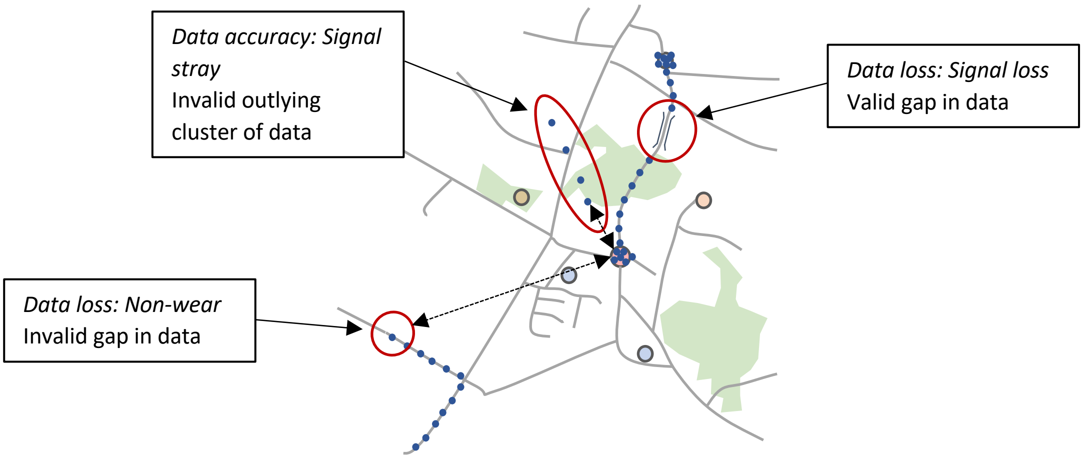


Figure 4.1: Example issues with GPS data due to technical limitations

The initialisation period required for a receiver to acquire satellite signals after it is switched on also varies across different sensors. Acquisition periods are typically referred to as hot, warm, or cold start times, depending how recently the device was last used [224]. Duncan and colleagues found hot start times ranged from 1 to 6 seconds, warm ranged from 15 to 38 seconds and cold from 35 to 45 seconds [223]. During these times, data relating to location, time, speed, and satellite coverage may not be recorded correctly. This differs to signal stray where attributes are populated correctly but positional errors exist.

Consequently, cleaning processes are required to identify and remove raw data points that have been positioned or populated erroneously. Leaving such data points unaccounted for may lead to a misrepresentation of spaces used.

4.1.2.2 Data loss

The obstruction of signal between satellites and receivers in covered locations such as tunnels and subway systems can lead to periods of *signal loss* [222], whereby the location of a GPS device is not recorded. The loss of reception can create a gap of up to several minutes in the data which affects the quality of data and ability to monitor interim movement (Figure 4.1) [224]. Simulating movement or routes taken during periods of signal loss is difficult, particularly in dense urban areas where many route options are available and the risk of signal loss is greatest.

Depletion of battery power and periods of non-wear by participants may also contribute to times when location is not being recorded. Batteries which last for a day or longer are therefore useful for capturing free-living activity and reducing the burden on participants [223]. However, battery life is compromised by the epoch used to collect data points with shorter epochs contributing to richer locational information at the cost of reduced battery life.

The review by Krenn and colleagues found that the longer the measurement period, the greater the proportion of data lost due to missing or unusable data [151]. Where the intended measurement period was one day, up to 40% of data was lost compared with up to 90% of data lost for studies with a 7-day measurement period. While the use of GPS provides a wealth of spatial information within and beyond the residential neighbourhood, it may therefore be important to account for missing data to provide a representative picture of where individuals spend time. Studies should consider battery life of devices and researchers should visually inspect raw data for potential errors prior to analysis [222]. In line with other objective measures of behaviour, such as accelerometers, sufficient wear time is also required to capture representative coverage of an individual's activity.

4.1.2.3 Existing GPS data cleaning methods

Drawing on relevant studies from Chapter 3, the review by Krenn and colleagues, and GPS studies centred around health and place, I describe methods implemented to deal with the issues of data accuracy and data loss and to prepare data for analysis. This section is not intended to be a comprehensive review of literature on the application of GPS, instead it aims to provide insight into key processes previously employed. In general, methods prevalent across the studies are used to i) remove erroneous points such as locations recorded outside of the study area, ii) remove irrelevant points that do not pertain to the research question, or iii) remove points that do not accumulate to valid times. Whilst some criteria focus on removing erroneous data and others focus on tailoring the data for the research question, for simplicity, I refer to (i), (ii) and (iii) all as *data cleaning methods* in this chapter.

The data cleaning methods and criteria used to exclude GPS points in existing studies are detailed in Table 4.1. I group criteria which relate to attribute values collected from devices under 'macro-level' cleaning methods, and criteria based on the spatial or temporal distribution of points under 'micro-level' cleaning methods.

Macro-level cleaning in previous studies

Macro-level cleaning primarily related to issues of data accuracy and typically involved the identification and removal of points with systematic errors based on a range of variables collected from devices. Variables included the number of satellites in view, a measure of GPS accuracy (the horizontal dilution of precision (HDOP)), speed, and directional heading.

Few studies filtered points based on satellite and HDOP values. This may be because the variables available for formulating exclusion criteria are dependent on the device used and its settings. For example, recording information relating to satellites affects battery life and so may not be available in datasets which favour shorter epochs or longer data collection periods.

The most prevalent variable used for filtering point based on attribute values was speed and although thresholds used varied across studies, all were designed to capture unrealistic values. Excessive speeds are often caused by multipath reflections, whereby GPS signals are received directly from satellites but also reflected from local buildings or objects [225], [226]. Maximum speed thresholds ranging from 100 km/h to 200 km/h were therefore used by a number of studies [163], [170], [194], [227]–[229] to remove erroneous data points and capture free-living movement. Other studies used lower thresholds to identify behaviours of interest and remove points irrelevant to the research question. To capture walking, maximum speed thresholds of 6 km/h and 8 km/h were used [182], [230] and studies interested in travel or activity bouts applied a minimum threshold of 0 km/h to remove stationary points [231], [232].

Table 4.1: GPS data cleaning methods and criteria in key literature

Lead author, Year	Purpose of GPS data	**Macro-level cleaning** *Exclusion criteria based on attribute values of points*					**Micro-level cleaning** *Exclusion criteria based on spatial/temporal distribution of points*				
		Satellites	**HDOP**	**Speed [km/h]**	**Date/ Time**	**Additional**	**Signal stray**	**Signal loss [minutes]**	**Non-wear [minutes]**	**Invalid data [measure per day]**	**Additional**
Tsui, 2006 [231]	Identify travel mode	< 3	> 5	0		0 directional heading	Points that jump significantly from original traces *np*	> 2			
Auld, 2009 [228]	Identify activity and travel locations	*np*	*np*	> 160			Clusters of up to 9 points recorded between two jumps of > 15 seconds	> 0.25			
Wheeler, 2010 [203]	Quantify after-school activity in greenspace, non-greenspace, and indoors			> 15	< 3pm > 7pm				> 60	< 60 seconds	
Cho, 2011 [230]	Identify walking trips			< 2 > 8						< 5 minutes	
Lin, 2013 [229]	Identify travel mode			> 144			Zig-zag traces which deviate from road network smoothed using Kalman filter				
McMinn, 2014 [192]	Identify environmental features on active school commute								> 90		
Houston, 2014 [183]	Identify walking trips and measure exposure to built environment								> 60	< 8 hours	Points near anchor locations such as home, work, or school
Kosaka, 2014 [175]	Identify activity locations and measure use of an environmental intervention				< 7am > 8pm					> 4 hours of non-wear	
Wolf, 2014 [232]	Identify travel mode [233] (Three data cleaning [234] methods compared) [235]	< 3 < 3	> 4 > 5	0 < 10.8		Speed and acceleration smoothed and filtered Altitude outside study area	*np*				Points with < 15 m movement
Cetateanu, 2016 [227]	Identify travel mode and measure exposure to food environments			> 100	< 8am > 10pm		Isolated point > 500 m from neighbouring point			< 60 seconds	

Helbich, 2016 [194]	Identify trips and measure exposure to natural and built environments	> 150				Trips identified using cluster detection algorithm
Lee, 2016 [179]	Identify trips and derive activity space				< 1 trip	GPS trips identified using the tracking analyst tool in ArcGIS
Sanchez, 2017 [163]	Derive activity space	> 120	< 6am > 10pm	Jumps > 1km	< 13.9 hours	Cold start points > 50 m from house
Quistberg, 2017 [182]	Identify walking trips and measure exposure to risk of pedestrian-vehicle collision	> 6			< 60 seconds	
Babb, 2017 [205]	Identify trips and derive activity spaces					Trips identified manually in GIS using aerial data, topographic maps, and travel diaries
Vich, 2017 [170]	Derive activity spaces	> 140		Spatial accuracy < 50 m	< 12 hours for at least 2 days	
Chaix, 2017 [174]	Derive activity spaces					Places visited identified from Kernel density surface

HDOP = Horizontal dilution of precision (measure of data accuracy)

np = variable used in cleaning criteria but no parameters given

Micro-level cleaning in previous studies

Micro-level cleaning involved the investigation of GPS points, typically in consecutive order, to deal with issues of both data accuracy and data loss.

Some studies acknowledged the importance of jumps in the data and removed points based on their spatial or temporal proximity with neighbouring points. This method appeared useful for removing isolated points offset from the main cloud of GPS data. A novel approach used by one study aimed to identify small clusters of outlying data by cycling through previous and subsequent points to identify groups of points recorded between two temporal jumps of more than 15 seconds (using data collected at 5 seconds epochs) [228]. Another study assessing travel modes identified and smoothed data located away from road networks [229]. These approaches build on macro-level cleaning methods by locating artefacts in the spatial distribution of GPS points as signal stray or data irrelevant to the research question.

Studies also used temporal jumps to recognise times where devices stopped recording and classified periods over 15 seconds [228] and 2 minutes [231] as signal loss and periods longer than 60 minutes [183], [203] or 90 minutes [192] as non-wear. Some studies have attempted to interpolate between points where signal has been lost by using smoothing approaches such as Kalman and Gaussian filters, to predict where individuals spend time [193], [229]. However, these approaches are uncommon in health research, especially outside of transport research where the aim is to identify travel modes from GPS data rather than derive exposure measures.

Most studies applied an exclusion criteria based on minimum wear time to ensure data were valid for identifying behavioural patterns. Wear time related to total wear for the day with minimum thresholds ranging from 8 hours [183] to 14 hours [163]. Periods of consecutive wear were also stipulated with three studies excluding days where participants did not record trips that summed to at least 1 minute [182], [203], [227].

4.1.2.4 Reflections on existing cleaning methods

Various methods were employed across studies with some focusing on either macro or micro-level cleaning. There was little consistency in the variables used and the threshold values chosen to exclude points. This may be due to a lack of systematic reviews of methods and recommended guidelines to date, as well as differences in data availability and devices used across studies. Variability in cleaning methods may also be driven by the aims of the studies as different research questions appear to have different requirements for the variables and thresholds used.

Although spatial and temporal jumps between consecutive points were calculated for some studies, few acknowledged the importance of dealing with both signal stray and signal loss. Often, cleaning processes were not transparent in the literature and thresholds chosen to remove GPS points were not justified. This may be because studies that use cluster analysis, specialised software to identify trips, or qualitative data to interpret or confirm GPS locations assume that errors in the data are accounted for in these additional processing steps. However, the errors remain present in the data, highlighting the need for a tool to help deal with data cleaning efficiently.

Despite the lack of coherence in approaches to cleaning GPS data across the studies, it is possible to draw on the most applicable methods previously used. By adapting and combining them, activity spaces can be generated from cleaned GPS data to answer my research questions in subsequent chapters relating to changes in the spatial distribution of activity in response to an environmental intervention. For example, the thresholds used for data cleaning in transport studies to identify trips may be too sensitive for the purpose of my study as I aim to capture where people spend time and how space is used. However, lessons can be learnt from these methods regarding thresholds and identifying erroneous data based on their spatial and temporal distribution which can be adjusted to suit my aims.

4.1.3 Aims and scope

The aim of the methodological work is to develop an automated process to prepare and clean GPS data which can be replicated in alternative datasets. The intention is to develop a process that allows activity spaces, as well as outcomes relating to change in spatial patterns and locations of activity, to be derived in order to answer research questions proposed in Chapter 5 and Chapter 6.

4.2 Methods

4.2.1 Dataset

GPS data were obtained from the Commuting and Health in Cambridge study; a natural experimental cohort study conducted in four waves between 2009 and 2012 [89]. The study was set up to investigate changes in travel behaviour and associated health impacts in response to the opening of the Cambridgeshire Guided Busway; a major transport infrastructure project comprising a bus network and adjacent traffic-free walking and cycling route.

The dataset was considered appropriate for the study given its small geographical area and heterogeneous sample of working adults. Participants were recruited from workplaces within Cambridge and were required to live within 30 km of Cambridge City Centre and commute regularly to and from work, irrespective of their employer, occupation type, working hours, and number of work locations. Consequently, the majority of GPS traces for the sample were relatively contained which, combined with a small sample size, made management and processing of the data practicable. The sample study population of working adults was also suitable as travel behaviours within this age group are largely autonomous and flexible, compared with children or older adults for example. The main arguments used by policy makers to support the busway was that it would provide an alternative commuting route and reduce traffic on the busy A14 trunk road. The opening of the busway is likely to have effects on habitual commuting patterns for this group which has implications for changes in travel and physical activity behaviour at the population level. The collection of data before and after the busway therefore allows for a longitudinal assessment of changes in spatial behaviour and physical activity in response to an intervention which has rarely been attempted in previous research.

4.2.2 Potential participants for analysis

After baseline, participants were re-surveyed each year from 2010 to 2012 and new cohort members were included in each phase of the study to account for attrition. All participants completed a questionnaire at each phase and a sub-sample were invited to participate in objective physical activity monitoring at baseline, but no GPS data were collected at baseline. From phase 2 of the study onwards, a sub-sample were invited to wear an accelerometer and GPS device simultaneously for 7 days.

Table 4.2 shows the number of participants with each data type at all of the study phases. Due to the introduction of new cohort members throughout the study, repeat measures were only available for a sub-group of participants. The busway opened in August 2011, effectively midway through phase 3. To assess behaviours before and after its opening, data from phases 2 and 4 of the study collected in 2010 and 2012 respectively were used. For comparative purposes, the full sample refers to participants with data at both phase 2 and 4 of the study. I therefore included participants from the full sample with questionnaire, GPS, and accelerometer data at both phases which created a potential sample of 78. GPS data were cleaned for all of the potential sample and participants with sufficient data to derive activity spaces were included in the final analytic samples.

Table 4.2: Participant numbers for each phase of the cohort study

Element of study	Phase 1 2009	Phase 2 2010	Phase 3 2011	Phase 4 2012
Core questionnaire	1168	774	770	665
Household travel diary	n/a	491	365	n/a
ActiGraph	501	142	120	73
ActiHeart	n/a	201	141	131
GPS	**n/a**	**196**	**132**	**131**

4.2.3 GPS devices

GPS data were collected using QStarz BT-1000X receivers (QStarz International Co. Ltd, Taipei, Taiwan). The receivers are small portable devices attached to an elastic belt and worn on the hip during waking hours. At both phases participants were asked to wear the device for 7 days and to recharge the battery each night using a charger provided to them. GPS locations were collected at 5 second epochs at phase 2 of the study. However, the short epoch affected the battery life of the device, requiring participants to recharge devices every night which resulted in non-compliance and a loss of data. To reduce the burden on the participant and frequency at which devices needed to be recharged, data were collected at 10 second epochs at phase 4.

Duncan and colleagues compared the performance of the QStarz device with six other receivers and found the QStarz device to have the longest battery life with an average of 42 hours based on a 1 second epoch [223]. The receiver also had one of the shortest signal acquisition times as well as the greatest positional accuracy under six diverse environmental conditions. However, issues of signal reflection were noted in urban areas between high rise buildings [223]. An initial exploration of the GPS dataset showed some points offset from expected locations such as the road network in the centre of Cambridge and in areas of high building density such as the Biomedical campus. This highlighted the need for a data cleaning process, despite data having been collected from a relatively strong device with regards to data accuracy.

4.2.4 Data cleaning

Data were cleaned in a two-step process to firstly remove points with systematic errors based on attribute values that had not been correctly populated (macro-level cleaning). Removing these points using a primary filter ensured that models created for the latter stages of the cleaning process would run more efficiently. The aim of the second step was to identify significant jumps in distance and time between consecutive points in order to detect signal stray where points had been incorrectly positioned, signal loss, and periods of non-wear (micro-level cleaning).

The process was developed using an investigative approach whereby methods were refined on a random 10% of participants in the potential sample (test sample). Wolf and colleagues highlight the difficulties in identifying issues specific to datasets at the initial stages of development and suggest the need for iterative investigation to finely tune the cleaning process [232]. A series of approaches was therefore tested before compiling the final data cleaning process. The aim of this, in part, was to make running the process on the whole sample as efficient as possible. Geographic information system (GIS) software, ArcGIS, was used to develop the process. Although a range of GIS software and methods to manage geographical data are available, I chose to use ArcGIS for the remainder of the projects presented in the PhD as I have extensive experience of using the software to analyse data and produce maps. Drawing on this experience, I used the Model Builder function within the ArcGIS suite to streamline processes by collating scripts and tools which iterate through files of GPS data. Although I have experience writing applications in alternative programming languages, self-directed, I developed new skills writing Python in order to manipulate the data within ArcGIS effectively. All data processing and development of functions in Python was performed by myself. To maintain patient confidentiality, the processed GPS data were matched with questionnaire data by the Data Management team in the MRC Epidemiology Unit.

The final workflow is detailed in Figure 4.2 and the decisions behind each stage, justification and explanation of each process are described in the remainder of Section 4.2.4. Examples of participant data are provided in Figure 4.3 and detailed examples of code developed in Python to process the data are included in Appendix C1.

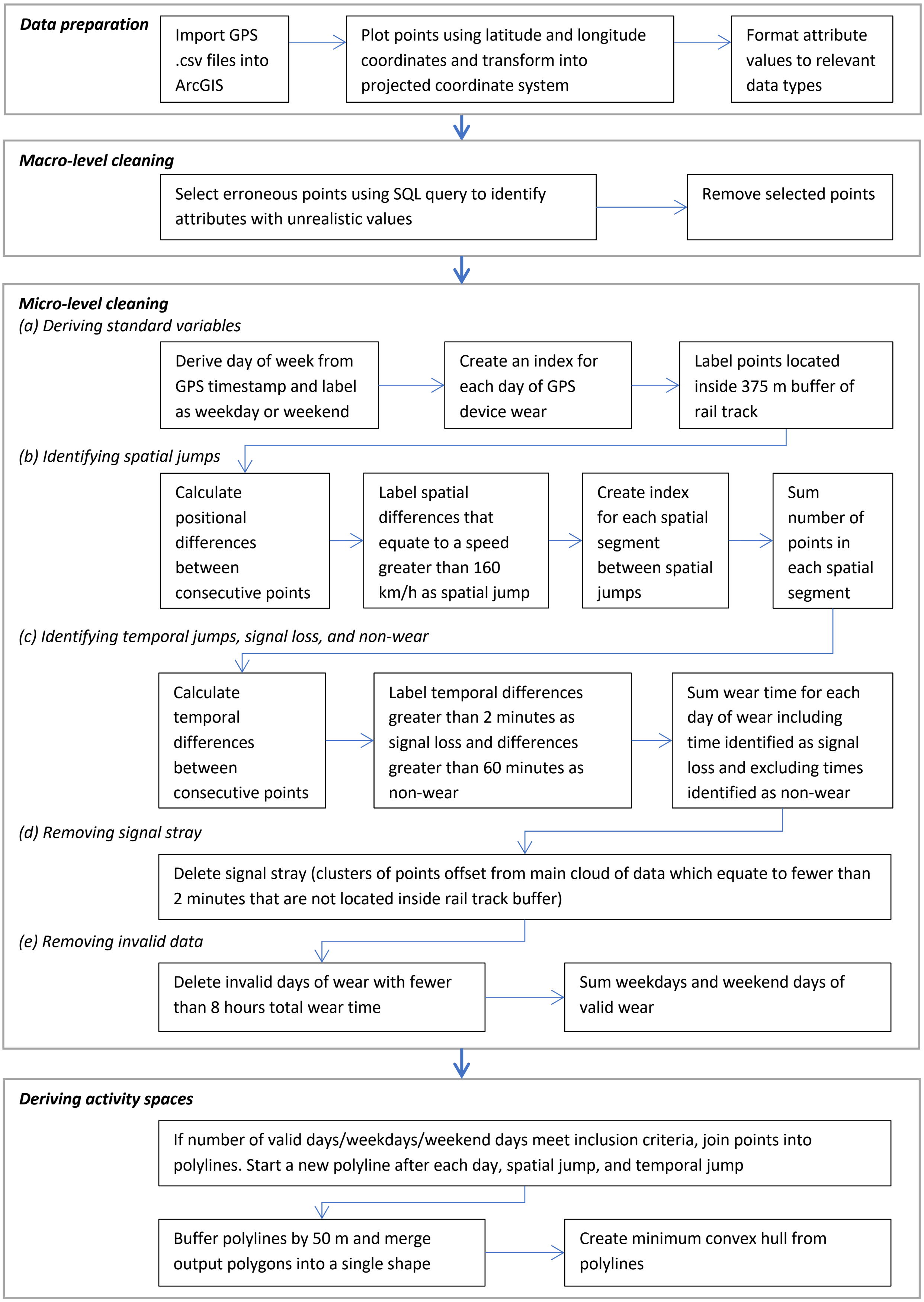


Figure 4.2: Workflow of final data cleaning methods applied to Commuting and Health in Cambridge GPS data

4.2.4.1 *Data preparation*

ArcGIS was used to plot the coordinates of the GPS data for each individual. Data points were plotted in geographic coordinate system WGS 1984, the standard for use in satellite navigation and GPS, and subsequently projected to British National Grid using the OSTN02 transformation. Working with the data in a projected coordinate system allowed for distances between points, as well as features of the activity space to be calculated in metres. The positioning of points was checked against Ordnance Survey (OS) data to ensure the data correctly overlaid route networks.

Attributes for each variable stored within each file were formatted to relevant data types. For example, dates and times were merged and converted from text fields into datetime formats, and speeds and heights were converted from text fields into numeric formats. Outlying points in the data were investigated here to identify any consistency in their attribute values; this information was used to inform the threshold values used in the macro-level filter.

4.2.4.2 *Macro-level filter*

The variables and thresholds for exclusion (Table 4.3) were informed by the literature and pilot testing criteria on the raw data of the test sample.

Table 4.3: Macro-level exclusion criteria

Field	Exclusion values
Longitude [decimal degrees]	< -7 > 15
Latitude [decimal degrees]	< 45 > 57
Date	NULL < 04/05/2010 > 7/11/2012
Speed [km/h]	< 0

Outlying points were identified and a minimum bounding box was drawn to capture GPS points located within the UK from all participants. The coordinates of the rectangle were used as the threshold for latitudinal and longitudinal values and GPS points with coordinates outside of this range were excluded. Next, the range of date values recorded by the GPS devices were investigated and NULL values or dates outside of the study range were removed. The distribution of speeds recorded for each GPS point was also explored. Maximum speeds were initially considered when developing the criteria but some high speeds appeared to be in locations of interest. Instead, all speeds were retained to include sedentary and high speed points, except those with negative values. An example of the process used to select and remove data using the macro level filter is shown in Figure 4.3.

Lone points within the minimum bounding coordinates but offset from the main data were initially considered when developing the criteria but a more advanced approach to deal with outlying clusters was applied during the micro-level data cleaning phase. Height was also investigated against OS terrain data but there appeared to be substantial mismatches in clusters or built up areas. Excluding points with extreme speed or height values would have removed a lot of data which may be of spatial interest for analyses, these attributes were therefore not used in the data cleaning process.

Although data on the visible number of satellites and HDOP values were collected at phase 4, none were collected at phase 2. Macro-level cleaning was therefore based only on location and date. A sensitivity analysis was performed by comparing the number of points removed from phase 4 data when incorporating HDOP and satellite threshold values in the exclusion criteria. The criteria were based on the literature (Table 4.1) and used HDOP values greater than five, or fewer than three visible satellites recorded.

Macro-level filter

Query to identify points with erroneous attributes:

Select * if

```
Longitude<-7 OR Longitude>15
OR Latitude<45 OR Latitude>57
OR Date IS NULL OR Date < date '2010-05-04
00:00:00' OR Date > date '2012-11-07 00:00:00'
OR Speed_kmh <0
```

Point ID	Date	Time	Latitude	Longitude	Speed	Height
1	10/05/2010	11:11:05	90	0	0	150
2	10/05/2010	11:11:10	52.17	0.14	0	88.2
3	10/05/2010	11:11:15	52.17	0.14	-3.2	89.0
4	10/05/2010	11:11:20	52.18	0.15	2.1	75.1
5	10/05/2010	11:11:25	52.19	0.15	2.5	70.5

Highlighted points selected and removed based on Latitude and speed criteria

Variables derived for micro-level cleaning

(b) Identifying spatial jumps
(d) Removing signal stray

(c) Identifying temporal jumps, signal loss, and non-wear
(d) Removing invalid data

VB code in Field Calculator[a] `WeekdayName(weekday([Timestamp]))`

CreateIndex function[b]

Add XY Coordinates[a]

VB code in Field Calculator[a] `Sqr(([DiffPrevE]^2)+([DiffPrevN]^2))`

SegmentTotal function[b]

PopulateIf function[b]

Timestamp	Day	Weekend	DayIndex	Railtrack	Easting	Northing	Difference with previous E	Difference with previous N	Distance from previous point	Spatial Jump	Spatial Segment ID	Points in Segment	Difference with previous time	Time Jump	Day Wear Time
07/05/2010 23:59:30	Friday	0	1	0	561725.8	265184.2	8.8	17.4	19.6	0	1	145	00:00:05	0	10:50:16
07/05/2010 23:59:50	Friday	0	1	1	561732.4	265173.8	6.6	10.4	12.3	0	1	145	00:00:20	1	10:50:16
07/05/2010 23:59:55	Friday	0	1	1	561136.4	265306.9	596	133.1	610.7	1	2	2	00:00:05	0	10:50:16
08/05/2010 00:00:00	Saturday	1	2	0	561123.2	265313.3	13.2	6.4	14.7	0	2	2	00:00:05	0	08:10:15
08/05/2010 01:32:00	Saturday	1	2	0	561724.3	265186.2	601.1	127.1	614.4	1	3	320	01:32:00	2	08:10:15

PopulateIf function[b]

Select Layer by Location[a]

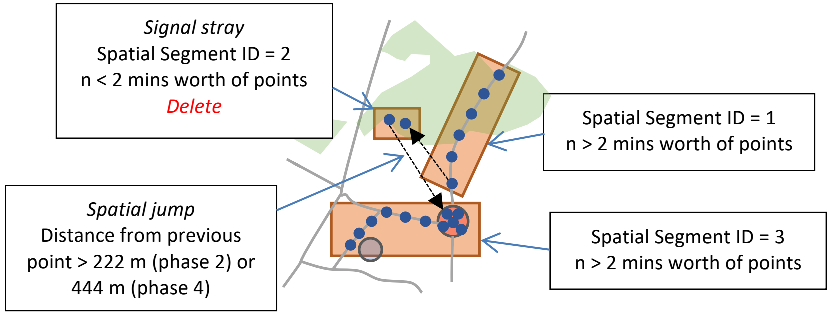
PreviousRow function[b]

PopulateIf function[b]

CreateIndex function[b]

PreviousRow function[b]

SegmentTotal function[b]



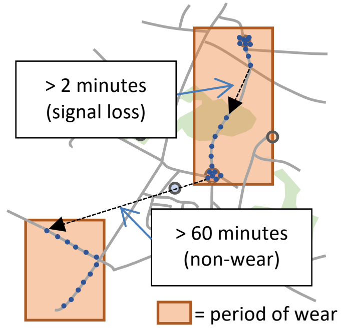


Figure 4.3: Cleaning processes applied to sample of participant's GPS data points *[a]Tool within ArcGIS Suite [b]Function written in Python (Appendix C1)*

4.2.4.3 Micro-level filter

(a) Deriving standard variables

The standard variables and methods used to derive them are detailed in Figure 4.3. Using the date recorded by the GPS receiver, days of the week and an index for each day of wear were derived. The day of the week allowed for a weekend variable to be populated whilst the purpose of the index is to allow for the identification of non-wear time. The index was populated using an iterative function written in Python which increments by one each time a new date is passed as a parameter into the code (Appendix C1).

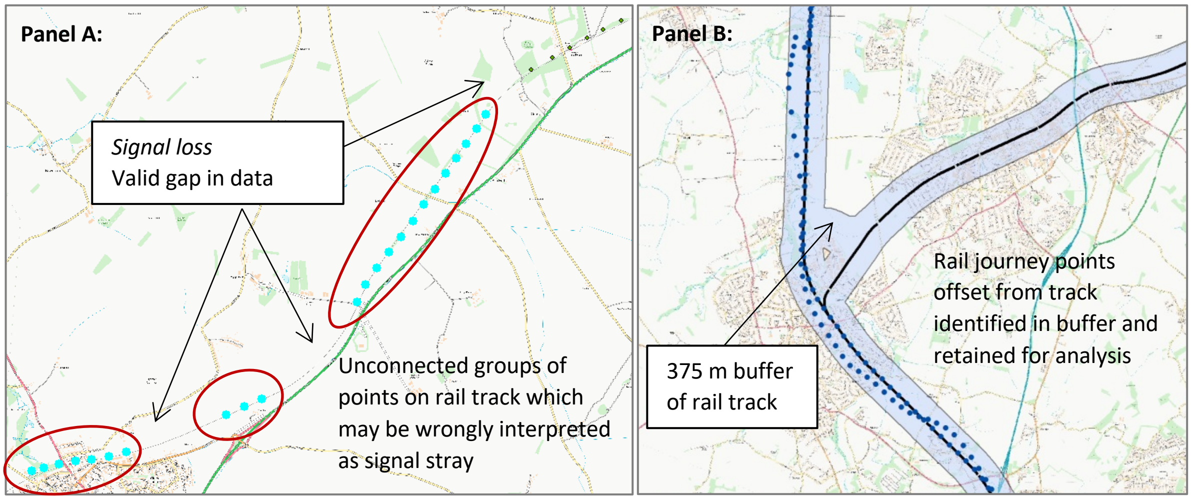


Figure 4.4: Examples of problematic points located on or near rail track

When investigating the spatial distribution of points of the test sample it was noted that single trips via train were often represented as unconnected groups of points due to tunnels and periods of signal loss (Figure 4.4, Panel A). To avoid these groups being identified as signal stray, a 375 m buffer around rail track was generated. The size of the buffer was based on observed distances between GPS points of train journeys and mapped rail tracks (Figure 4.4, panel B). Points located within this buffer were labelled, so that they could be retained for future processing.

(b) Identifying spatial jumps

Using Pythagoras theorem and Python code in the ArcGIS field Calculator tool, the spatial distances between each point were first calculated. To enable this, eastings and northings were generated for each point and a function to store the value from the previous row was developed.

To determine a reasonable threshold for spatial jumps and to enable the identification of signal stray in the subsequent phase of data processing, the distribution of distances between points were investigated for the test sample. As all points close to the rail network were not entered into the exclusion process, a threshold of 160 km/h, based on maximum car speed, was chosen. The calculation below shows the distance possible to travel at this speed. Spatial distances between points that exceeded 222 m at phase 2 and 444 m at phase 4 were labelled as a spatial jump.

3600 second per hour: *160/3600 = 0.044 km/s*

5 second epoch (phase 2):	*0.044 x 5 = 0.2222 km*	*= 222 m*
10 second epoch (phase 4):	*0.044 x 10 = 0.444 km*	*= 444 m*

The number of points were summed for each segment between spatial jumps using a function which counts points with the same segment ID. This value is used in subsequent data processing to remove signal stray.

<u>(c) Identifying temporal jumps, signal loss, and non-wear</u>

Time differences between consecutive points were calculated using a variation of the previous row function. Differences greater than 2 minutes were identified and labelled as signal loss and time differences greater than 60 minutes were identified as periods of non-wear (Figure 4.3).

Thresholds of 15 seconds [228] and 2 minutes [231] have been previously used in the literature for signal loss. As the focus of this study is to generate general activity spaces using multiple days of data, rather than travel or activity-specific locations [228], [231], the threshold for signal loss does not have to be so sensitive. Previous studies used thresholds of 60 minutes [183], [203] and 90 minutes [192] when identifying non-wear. I compared both these thresholds on the test sample and found no additional points were included when using the more conservative time.

Using the day index derived in part (a), daily wear time was calculated, including periods of signal loss but excluding periods of non-wear.

<u>(d) Removing signal stray</u>

Based on the sum of points in each spatial segment, derived in part (b), segments with fewer than 25 points (phase 2) or 13 points (phase 4) were removed (Figure 4.3). Eight points of signal stray collected using a 5 second epoch have been used by Auld and colleagues in their study on travel locations [228], but when investigating this method in the test files, this appeared to be too sensitive. After testing different thresholds, around 20 points appeared to detect most clusters of signal stray for phase 2 (at a 5 second epoch) and 24 points equated to 2 minutes which has previously been used as a temporal threshold for signal loss [170], [231].

Clusters of signal stray were removed after wear time was calculated. Deleting points earlier in the workflow would create gaps in the data which may have been incorrectly labelled as signal loss or non-wear.

<u>(e) Removing invalid data</u>

Using the daily wear time variable calculated in part (c), days with fewer than 8 hours of wear were excluded in line with wear time limits typically used in accelerometer studies [236]–[238]. Both 8 hours of total wear and 8 hours of consecutive wear were investigated for the test sample. Applying limits for consecutive wear meant that over 50% of the data was removed, so this was not taken forwards and a total of 8 hours wear time was used.

4.2.5 Preparing data for analysis

4.2.5.1 Deriving activity spaces

After cleaning, data from participants with sufficient wear time at phases 2 and 4 of the study were used to create activity spaces for the three temporal scales detailed in Table 4.4. Findings from the systematic review (Chapter 3) indicate that studies typically use a minimum of 3 or 4 days of valid data to delineate activity spaces for multiple days, in line with studies that use objective measures of physical activity. For a weekly activity space, participants with a minimum of 4 valid days of data, including 1 weekend day were therefore included. For weekday activity spaces, a minimum of 3 weekdays was required, and for weekend activity spaces, participants must have recorded at least 1 weekend day of data. Participants were included for analysis if they had data for one or more temporal scales.

Table 4.4: Valid wear time for deriving activity spaces at different temporal scales

Temporal scale	Data requirement
Week	Minimum of 4 days of wear, including 1 weekend day (could include 2 weekdays and 2 weekend days of wear)
Weekday	Minimum of 3 weekdays of wear
Weekend	Minimum of 1 weekend day of wear

As demonstrated in the activity space review in Chapter 3, daily path areas provide an accurate measure of spaces visited. In the subsequent analysis in Chapter 5, I aim to describe spaces used and detect changes in spatial behaviour over time, in the context of an environmental intervention. I therefore chose to use the daily path area to estimate individuals' activity spaces. Valid points were used to derive polylines that represent traces of GPS data, with new lines being started after each spatial or temporal jump. Polylines were grouped for each temporal scale of interest and buffered by 50 m to create a daily path area. The buffer size was kept relatively small compared to other measures used in the literature which range from 50 m to over 800 m, to capture relevant spaces actually used (such as the intervention) and to be sensitive enough to detect change.

4.2.6 Outcome measures

In order to analyse how participants' movement and use of space changed in response to the intervention, metrics relating to features of the activity space (shape and size) were investigated and generated for data at both phases 2 and 4 of the study.

4.2.6.1 Activity space size

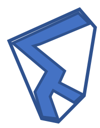
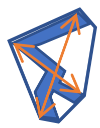
The area of each daily path area was derived for each temporal scale using geometry tools within the ArcGIS suite. An absolute area was measured, as well as an average area by dividing the absolute area by the number of valid days of wear for each temporal scale.

4.2.6.2 Activity space shape

Four different compactness scores were considered to measure the shape of the daily path areas (Table 4.5), based on the published literature. The measures of compactness have been used in political sciences to assess administrative boundaries [239], [240] and typically compared the daily path area to an optimum compact shape such as a circle or minimum convex hull [190].

All scores fall with the range of 0 and 1 with 1 indicating a more compact score. The range and distribution of different scores from each measure of compactness were explored for weekly activity spaces. Low scores within a narrow range were returned for the Polsby-Popper (0.0005 to 0.04), Reock (0.0002 to 0.2), and Convex Hull scores (0.001 to 0.04). As the sample from the study is made up of commuters, their activity spaces are largely constrained by road networks which meant their use of space rarely resembled a circle or square. It therefore made sense conceptually to look at the broader space potentially covered by participants. The ratio of length by width of the minimum convex hull polygon, with scores ranging from 0.1 to 0.8, was therefore chosen as the final measure of compactness which was also identified as the most common measure used in the review in Chapter 3.

Table 4.5: Potential measures of activity space compactness

Score	Description	Equation	Illustration	Assumption
Polsby-Popper	Ratio of the area of DPA to the area of a circle whose circumference is equal to the perimeter of the DPA	$PP = 4\pi\ x\ \frac{area_{DPA}}{Perimeter_{DPA}{}^{2}}$		Most compact district is a circle (=1)
Reock	Ratio of the area of DPA to the area of the minimum bounding circle that encloses the DPA	$R = \frac{area_{DPA}}{area_{MBC}}$		Most compact district is a circle (=1)
Convex hull score	Ratio of the area of DPA to the area of the minimum convex polygon that encloses the DPA	$CH = \frac{area_{DPA}}{area_{MCP}}$		Most compact district is convex (=1)
Convex Hull length:width score	Ratio of the area of DPA to the area of the minimum convex polygon that encloses the DPA	$LW = \frac{width_{MCP}}{length_{MCP}}$		Most compact district is where length and width of convex polygon are equal (=1)

4.3 Results and discussion

GPS data were cleaned and processed to create and measure attributes of activity spaces for participants with sufficient data at phases 2 and 4 of the Commuting and Health in Cambridge Study. The process was planned, developed and refined in stages using a random test sample of participants' GPS datasets from February 2019 to May 2019.

4.3.1 Exclusion of GPS data points and invalid wear time for test sample

The final macro-level filter detected and removed a small number of points (maximum of 2 per participant), typically with erroneous latitude, longitude, or date values. The total number of points removed after running the whole data cleaning process for each file in the test sample was no greater than 0.05% of the raw data points for each participant (Table 4.6). In contrast, a sensitivity analysis of phase 4 test files using HDOP and satellite values for exclusion criteria removed up to 6% of raw data points for each participant. Inspecting the points identified for removal in the sensitivity analysis, nearly all were located inside buildings and had valid attribute values. Of the points removed by the data cleaning process, 67-100% were also identified and removed in the sensitivity analysis for 78% of the test sample. The points removed for participants in the test sample which were not captured by the sensitivity analysis, were visually inspected and identified as signal stray which would create issues if used to derive activity spaces. This highlights the benefit of cleaning data based on its spatial and temporal distribution, not just attribute data.

Table 4.6: Details of GPS points removed and valid wear for test sample

			Number of points removed				Valid number of days of data			
	ID	No. raw data points	Macro filter	HDOP/ satellite	Micro filter	Invalid days	All days	Week days	Weekend days	Activity space temporal scale
Phase 2 data	1	55167	0	-	14	2	6	4	2	All
	2	44199	0	-	0	2	6	4	2	All
	3	93623	1	-	0	1	7	5	2	All
	4	80331	0	-	22	0	8	6	2	All
	5	64398	0	-	3	0	7	5	2	All
	6	67874	1	-	19	1	7	5	2	All
	7	75157	0	-	28	1	8	6	2	All
	8	98623	0	-	12	1	8	6	2	All
	9	88326	0	-	26	0	7	5	2	All
Phase 4 data	1	23689	0	101	0	1	3	1	2	W/E
	2	22244	0	1035	0	1	6	4	2	All
	3	26565	0	159	4	4	4	2	2	W + W/E
	4	68496	0	858	8	0	6	4	2	All
	5	12752	0	346	2	7	0	0	0	None
	6	32112	0	278	0	1	6	4	2	All
	7	27398	0	555	15	4	4	2	2	W + W/E
	8	17012	0	957	9	6	2	0	2	W/E
	9	34260	0	381	0	1	6	4	2	All

Temporal scale of activity space: W = week, W/E = weekend

One participant was excluded from the test sample (11%) due to insufficient wear time. All others had data to derive activity spaces for at least one temporal scale. Phase 4 data appeared to be more sensitive to the data cleaning process with typically fewer valid days recorded for each participant, despite the longer epoch of data collection and less demand for recharging the GPS device's battery.

4.3.2 Derivation of activity space metrics

Seventy-eight participants (17.6% of the full sample) had GPS data at phases 2 and 4 of the study (Table 4.7). The final data cleaning process was performed on this potential sample in June 2019. Following the cleaning process, 85.9% of the potential sample had sufficient data for the derivation of week activity spaces, 80.8% for weekday activity spaces, and 91% for weekend activity spaces.

Briefly, all samples had a higher proportion of females than males, and a mean age of 45-46 years. The mean activity space size varied across temporal scales, with weekend activity spaces being the smallest. Mean week activity spaces, which include data from weekday and weekend days, were the largest suggesting movement typically occurs in different locations on weekdays and weekends. Some differences in mean activity space size were shown between study phases for each analytic sample which may be due to a change in spatial patterning of behaviour. Little variation was shown for measures of activity space shape across temporal scales and study phases.

Descriptive results relating to sample characteristics and features of activity spaces are explored further in Chapter 5. For the purposes of this chapter, high retention of the sample for analysis after the cleaning process and comparable distribution of characteristics to the potential sample, as well as the successful derivation of outcome measures, highlight the data's operational capacity and suitability for analysis.

Table 4.7: Sample characteristics and activity space metrics

	Full sample		**Potential sample**		**Analytic samples**					
	Data collected		GPS data collected at phases 2 and 4		Included for week analysis		Included for weekday analysis		Included for weekend analysis	
	n	**%**	**n**	**%**	**n**	**%**	**n**	**%**	**n**	**%**
n	**444**	-	**78**	17.6	**67**	85.9	**63**	80.8	**71**	91.0
Characteristics										
Sex										
Male	**133**	30.0	**35**	44.9	**28**	41.8	**29**	46.0	**31**	43.7
Female	**277**	62.4	**43**	55.1	**39**	58.2	**34**	54.0	**40**	56.3
Age [Years]										
Mean (SD)	***45.6 (11.2)***		***45 (10.1)***		***45 (9.8)***		***45.9 (9.5)***		***44.8 (9.9)***	
					Study phase					
Features of activity space					**2**	**4**	**2**	**4**	**2**	**4**
Size [km²]										
Mean (SD)					14.8 (12.5)	13.2 (13.2)	8.8 (8.1)	9.1 (11.4)	7.5 (7.8)	6.1 (7.5)
Shape [CVH L:W ratio]										
Mean (SD)					0.34 (0.16)	0.35 (0.15)	0.34 (0.18)	0.34 (0.16)	0.35 (0.19)	0.35 (0.17)

4.3.3 Strengths and limitations

The application of the data cleaning process allowed for the removal of GPS points to be automated for a large number of files. Due to the limited requirements of attribute values, and focus on spatial and temporal distribution of points in the cleaning process, this automation can be applied to alternative datasets and assist in future analyses by other researchers. The present lack of automated cleaning processes to deal with both signal loss and stray, and reliance on attribute values which vary across GPS devices, highlights the novelty of the processes and workflow developed and may explain the lack of attention focused on data cleaning in the literature to date.

Previous studies have used criteria related to positional accuracy of GPS points, such as the HDOP value and number of satellites in view, to remove erroneous data points. However, these attributes were only available for data collected at phase 4 of the study due to a decision made by the study coordinators after phase 2 to change the GPS device's settings to collect more data at less frequent epochs to reduce the rate of battery loss. The results from the sensitivity analysis showed that applying criteria based on HDOP and satellite values would have resulted in a greater loss of wear time. Furthermore, there was little consistency in other values attributed to the points removed in the sensitivity analysis, such as speed, and the points were typically located in indoor environments or along routes. This would have made it difficult to replicate the filter using attributes available for the phase 2 data. Relying on a filter which assessed HDOP and satellite values may also have removed points collected in locations which may be useful for the analysis of time spent indoors or travelling.

Due to the iterative development of the process using a test sample of files, limitations of the dataset could be investigated in detail. An important stage in the iterative development of the micro-level filter was the identification of signal stray in the raw data. This allowed for a clause in the process to be developed which detected and removed small clusters of GPS points offset from the main data. However, a limitation of this feature is that the clause cannot recognise signal stray if the cluster is located close to valid data points. This is accounted for to some extent by starting a new polyline after each spatial and temporal jump in the data before generating activity spaces.

Similarly, the development process allowed for issues of signal stray or loss close to rail tracks to be identified (Figure 4.3). Although a manual approach was applied to clean and join points located on or close to rail tracks which was relatively time consuming, this ensured that points of potential interest were maintained. If left unaccounted for, activity spaces derived for rail users would have been smaller due to signal stray from high speeds and periods of signal loss due to tunnels, which would bias analysis for users of that travel mode.

Lastly, data lost during warm-up periods of devices was not accounted for. However, the same devices were used across the sample and only participants with valid wear time were included in the analytic samples. Any issues with warm-up periods would have been the same for all participants in different contexts and would have less impact given the required 8 hours of wear time each day for included participants. Erroneous values recorded by the devices during this time would also have been captured and removed in the macro-level filter.

4.4 Conclusions and implications

Despite the limitations noted, the data cleaning process allowed for errors relating to data accuracy and data loss to be confidently removed from the potential sample and for the derivation of activity space polygons. This enables study aims and analyses in subsequent chapters to be achieved.

Due to the focus on relative positioning and little reliance on attribute data, specifically satellite information, the code and model developed in the ArcGIS suite can be applied to a range of GPS datasets. I would therefore like to encourage other researchers to use the methods developed in this chapter and adapt them for their research questions. My intention is to combine findings from Chapter 4 and Chapter 5 of the thesis into a paper for publication. The code and processing will be made available via an opensource hub and researchers may contact me to access the code. It is acknowledged that not all erroneous points may be removed using this process but it provides an important starting point. Future research may combine the cleaning process with smoothing techniques or raster analysis to account for signal stray that may not have been detected or for missing data due to device warm-up periods and signal loss.

Chapter 5

Using GPS data to assess changes in use of space in response to new infrastructure: the case of the Cambridgeshire Guided Busway, UK

5.1 Introduction

As described in the systematic review in Chapter 3, GPS data can provide information about locations directly visited by individuals to better describe mobility and provide a measure of spaces used. It can be used to understand the effects of changes to the built environment on use of environments and health. However, studies which describe environments used without exploring changes in behaviour, differences in access to environments, or the mechanisms behind their use limit the basis for causal inference.

It is hypothesised that activity spaces vary in size and shape across different populations, and according to different modes of travel [159], [190], [241]. The review highlighted differences in spatial patterning of behaviour and physical activity for utilitarian or leisure purposes by weekdays and weekends [142]. However, few longitudinal and intervention studies have used the concept of the activity space to investigate how spatial and temporal patterns of behaviour change in response to a specific change in the environment [142]. In Chapter 4, I derived activity spaces from GPS data collected at two different time points before and after the opening of new transport infrastructure. Using such measures, it is possible to objectively capture use of an intervention and to compare the size and shape of (*features of*) activity spaces over time. This allows for changes in spatial habits to be identified which may reveal information on the role of the intervention in travel behaviours and mobility. Changes in locations and modes of travel have implications for the direct environments experienced by individuals, and consequently their routine behaviours and health.

Linking access to and use of an environmental intervention to changes in activity spaces provides an important first step in understanding *whether* an intervention may bring about changes in behavioural outcomes [152]. For example, changes in features of activity spaces may indicate the spatial displacement of behaviour or the uptake of additional behaviour in new locations. By exploring these associations alongside qualitative data and individual maps of movement over weekdays and weekends, it is possible to understand *why and when* interventions are or are not used and to attribute *how* changes in behaviour may come about. This provides insight into the mechanisms behind the use and impact of specific environmental changes.

Evidence for how an intervention affects behaviour (causal explanation) complements evidence of associations between the intervention and behavioural outcomes (causal estimation) to strengthen the basis for causal inference [242]. Findings from such study designs are therefore useful for informing the design and delivery of future public health interventions [243]; an area where methods for evaluation and identification of plausible pathways to behavioural change remain underdeveloped [244].

5.1.1 Chapter overview

This chapter of the thesis uses GPS-derived activity spaces, described in detail in the previous chapter, to evaluate the impact of the opening of new transport infrastructure, the Cambridgeshire Guided Busway, on the spatial patterning of behaviour over time.

5.1.2 The intervention: the Cambridgeshire Guided Busway

The Cambridgeshire Guided Busway (hereafter referred to as the busway) is a major transport infrastructure project comprising a new bus network and an adjacent 22 km traffic-free walking and cycling route opened in 2011 in Cambridge, UK.

Cambridge has a number of major scientific and technology employers and the highest prevalence of cycling in the UK [90], [245]. The bus route was designed as an alternative to driving for commuters travelling into Cambridge and connects towns and villages to the north of Cambridge with the city centre and employment centres such as the Cambridge Science Park and the Cambridge Biomedical Campus. Buses run on a track segregated from traffic for the majority of the route, designed to improve journey times compared to routes taken along the road network. A path for pedestrians and cyclists runs alongside the busway; creating a new space for active commuting and physical activity (Figure 5.1). Bus stops located close to park and ride facilities near St Ives and Longstanton also allow for individuals to incorporate physical activity into their route by parking at the facilities and walking or cycling the remainder of their journey (Figure 5.2).

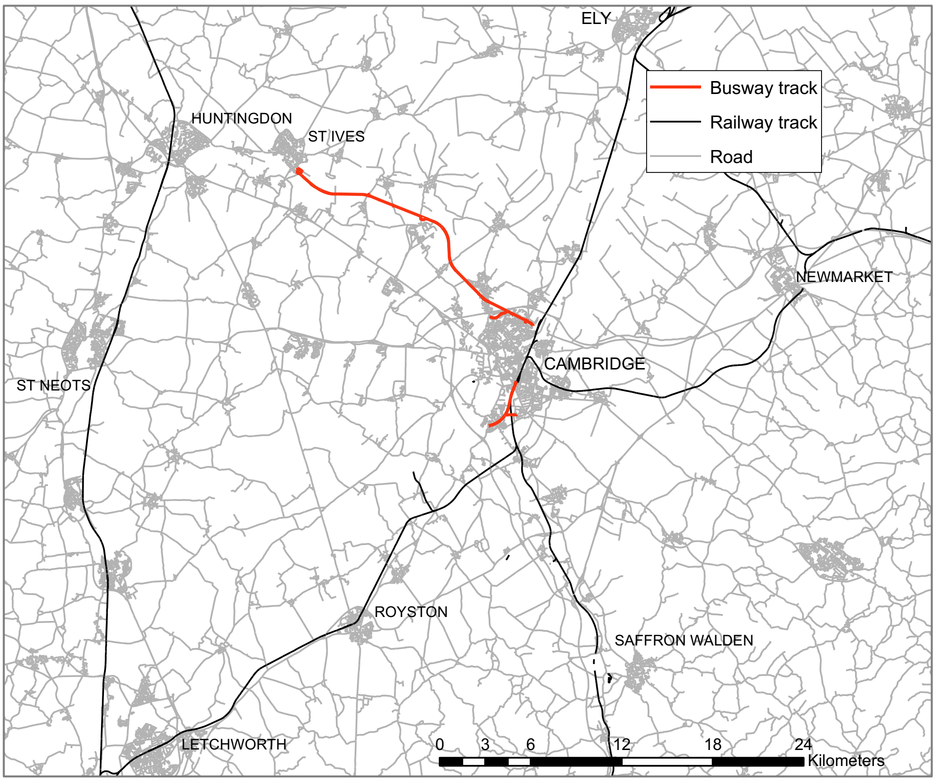


Figure 5.1: Image of Cambridgeshire Guided Busway and map of the off-road sections where path is located

Image source: https://www.flickr.com/groups/guidedbusway/pool/

Contains OS data © Crown copyright and database right 2019

Figure 5.2: Wider bus service route including on-road and off-road sections
Image source: https://thebusway.info/routes-times.shtml

5.1.2.1 Findings from the Cambridgeshire Guided Busway study

The Commuting and Health in Cambridge study was set up to evaluate the impacts of the busway. It was a 4 year mixed method study and the final report gives details of the project, including summaries of over 30 academic papers that were published [90]. The key evaluative papers and the research questions answered are detailed in Table 5.1.

Table 5.1: Key evaluative studies of the Cambridgeshire Guided Busway using data from the Commuting and Health in Cambridge Study

Lead author, Year	**Research question/description**	**Method**	**Data used**
Ogilvie, 2010 [89]	Protocol	-	-
Panter, 2011 [246]	What individual, workplace and environmental characteristics are associated with integrating walking and cycling into car commuting journeys?	Quantitative	Core questionnaire Household travel diary
Jones, 2013 [247]	How was the busway used and experienced in the weeks following its opening?	Qualitative	Interviews
Kesten, 2015 [248]	What were the longer term experiences of new transport infrastructure and its impacts on active travel behaviours?	Qualitative	Interviews
Heinen, 2015 [249]	What are the predictors of use of the busway for walking, cycling and public transport?	Quantitative - observational	Core questionnaire
Dalton, 2015 [250]	Are GIS-modelled routes a useful proxy for the actual routes followed by commuters?	Quantitative - observational	Core questionnaire Household travel diary GPS
Heinen, 2015 [251]	What was the impact of the busway on commuters' mode of travel, trip frequency and distance travelled to work?	Quantitative natural experiment	Core questionnaire Household travel diary
Panter, 2016 [252]	What is the impact of the busway on walking, cycling and physical activity?	Quantitative natural experiment	Core questionnaire Household travel diary
Foley, 2015 [253]	Are changes in active commuting associated with changes in physical activity?	Quantitative-cohort analysis	Core questionnaire Household travel diary ActiHeart
Costa, 2015 [254]	How much physical activity do commuters accumulate on the commute?	Quantitative - observational	Core questionnaire Household travel diary ActiHeart GPS
Prins, 2016 [255]	What are the causal pathways linking exposure to new transport infrastructure with changes in cycling to work?	Quantitative-mediation analysis	Core questionnaire

The evaluative studies looked at the associations between exposure to the busway and active commuting, changes in active commuting behaviour and its contribution to overall physical activity, and reasons for busway use. Over half of the sample of relatively affluent commuters reported walking or cycling to work at baseline [90]. Differences were observed for different sample and environmental characteristics with those without access to car parking at work and who reported environments supportive of active commuting on their route to work more likely to incorporate walking and cycling into their car journeys [246]. Following the opening of the busway, those living closest to the busway reported an increase in proportion of trips made involving any active travel and a decrease in the proportion of trips made entirely by car [251]. Proximity to the busway was also associated with an increase in time spent cycling over the week, which was most effective for those whose levels of active commuting were lower at baseline [252], as well as an increase in odds of busway use for cycling, bus travel and walking [249].

Again, some differences were observed for sample characteristics with the effect of exposure strengthened in towns for bus use, and in towns and villages for walking, compared with urban areas. Men were more likely than women to have cycled on the busway, whereas individual socioeconomic characteristics did not predict bus use or walking [249]. Using a combination of GPS and combined heart rate and movement sensors, one study found that 20% of the duration of journeys incorporating active travel was spent in MVPA; providing over half of the weekly recommended activity levels [254]. Despite exposure to the busway appearing to be associated with increases in active commuting, there was no direct evidence of an effect on overall physical activity but the study lacked statistical power to detect such an effect [90]. An increase in active commuting in the sample was shown to be associated with a borderline greater likelihood of an increase in total physical activity [253]. However, there was no evidence of associations between time spent active commuting and changes in recreational or overall physical activity [252], [253]. These findings together suggest people may have changed where they are active or taken different commuting routes.

Some studies used qualitative evidence to investigate participants' motivations for and experiences using the busway, as well as the role of the busway on the pathway to behaviour change [247], [248]. Some participants rarely considered the new transport infrastructure or described it as unappealing because of its inaccessibility or inconvenient routing. Others located more conveniently for access points experienced the new infrastructure as an attractive travel option. Likewise, the bus and pathway presented ambiguous spaces which were received in different ways; the availability of off-road cycling was appreciated but crowded buses and a lack of lighting were noted as barriers to use [248]. Previous users of regular bus services were also frustrated the new service was not coherent with the existing system whereas those who had previously travelled by car appraised the busway and perceived it to be a superior form of travel [247].

Use of the busway path appeared to explain the association between busway proximity and an increase in cycling but also explained a decrease in cycling for more active commuters [255]. This suggests some users started to cycle after the intervention whilst the busway may provide a quicker or more direct route for others. Currently, there is a gap in the evidence regarding the spatial and temporal elements of physical activity and changes in the location of behaviours and experienced environments in response to the opening of the busway. One study showed how GIS-estimated routes are acceptable for distance estimation [250] but to obtain accurate measures of environmental contexts in which physical activity occurs and how this changes over time, GPS measures of routes and spaces visited should be used.

5.1.3 Aims and scope

The aim of this study is to use the activity spaces derived in Chapter 4 to investigate how the shape and size of the activity space of study participants changes after the opening of the busway. I investigate whether sociodemographic and geographic characteristics, and travel options and behaviours (including proximity to and use of the busway) are associated with these changes. I also investigate whether this is different for weekday and weekend mobility patterns and use qualitative data to understand how and why certain individuals change their spatial habits in response to the busway.

The analysis in this chapter is exploratory and descriptive to understand how activity spaces change and for whom. It is designed to provide a foundation for the subsequent chapter which focuses on methodological approaches for understanding how and where changes in physical activity occur. To my knowledge, this body of work is one of the first studies in the field to investigate whether spatial patterning of movement and physical activity changes in response to an intervention and the potential pathways to change. An exploratory investigation is therefore considered appropriate to describe the sample and to inform the subsequent analysis, given the novelty of approaches to be used.

5.2 Methods

5.2.1 Data

Data from phases 2 and 4 of the Commuting and Health in Cambridge Study, as detailed in Section 4.2.1, were used for analysis. Participants were required to have questionnaire and GPS data at both phases, which created a potential sample of 78. The home and work locations of the potential sample in relation to the busway are illustrated in Figure 5.3.

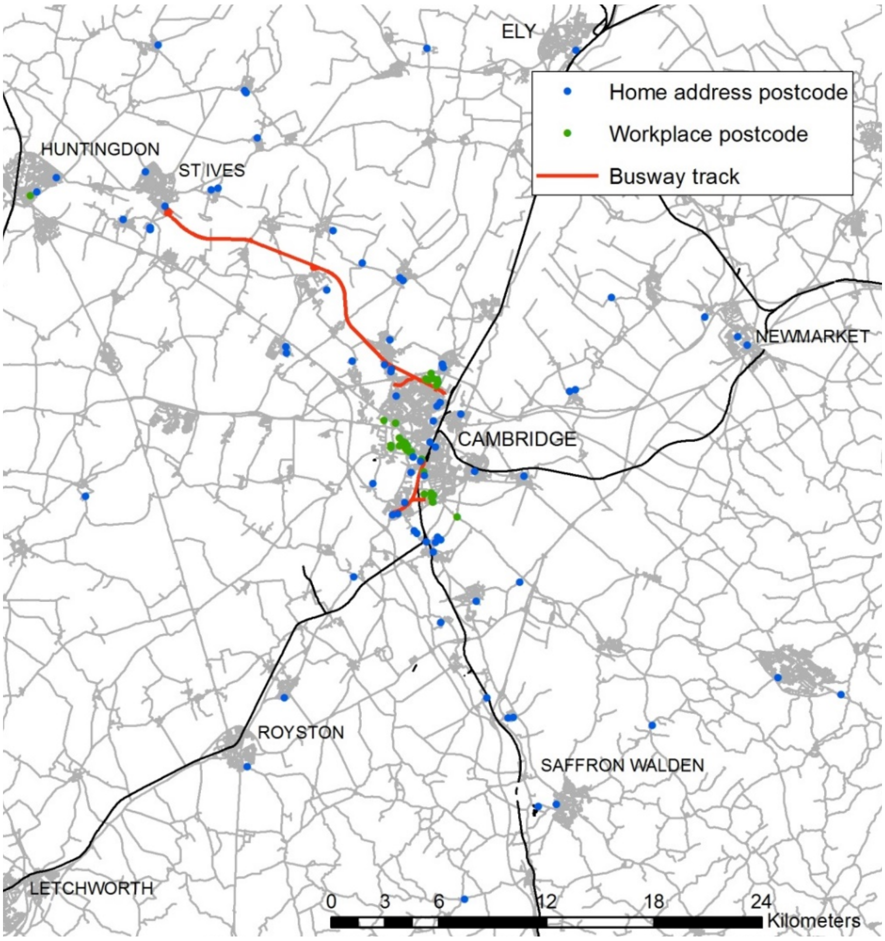


Figure 5.3: Home and work locations of potential sample at study baseline
Contains OS data © Crown copyright and database right 2019

Qualitative interview data collected post-intervention between February and June 2013 [89] were available for a sub-sample of participants and used to support case studies and maps of selected individuals. Interviews were semi-structured and questions asked related to experiences of using different modes of transport, and the facilitators of and barriers to travel behaviour change. Specific questions exploring the perceived impact of the busway on participants' travel behaviour were raised by the interviewer if not already discussed by the participant. Interview participants were recruited from the main study cohort, six of whom had valid GPS and questionnaire data.

I purposively selected four of the six participants' interview transcripts to provide contextual information for the descriptive analysis. Participants were selected based on heterogeneous data, and differences in activity space outcomes, reported and recorded busway use.

5.2.2 Outcomes: change in features of activity spaces

The activity space metrics derived in Chapter 4 were used to calculate changes in activity space and size and shape between phases 2 and 4 of the study. Data from valid days of more than 8 hours of GPS wear time were aggregated by week, weekday, and weekend days to create three temporal scales of activity space (Table 4.4).

5.2.2.1 Change in activity space size

To measure whether individuals' movements covered a larger or smaller space post-intervention, the area in km² of daily path areas measured at phase 2 of the study were subtracted from those at phase 4 for each temporal scale (week, weekday, and weekend). A negative outcome showed a decrease in activity space size and a positive related to an increase. Absolute changes were retained and percentages of change from phase 2 activity spaces were also derived to create a comparable measure across individuals.

The frequency and distribution of percentage changes in activity spaces showed the data were positively skewed with the top 25 percentile of participants recording a large increase in activity space size of 270% to 1600%. As the distribution of change was not normally distributed, it was considered appropriate to use categories rather than absolute change scores. It is assumed that there will be some inevitable change between the absolute size of activity spaces recorded at phases 2 and 4 for each participant. This is mainly because the number of journeys, time spent and destinations and locations visited is unlikely to be exactly the same for the two 7-day periods which are 2 years apart. Three categories of change ('increase', 'decrease' or 'no substantial change in activity space size') sensitive to some assumed variability were therefore used for analysis.

A range of definitions of change were explored based on the mean and median of change at each temporal scale including 25 per cent and 30 per cent cut-points, and tertiles. Little difference was observed across the different categories. Tertiles of percentage changes were therefore chosen for final outcome measure in line with a previous evaluative study of the busway on changes in physical activity outcomes [252]. This also ensured an even number of participants in each group of change which helped to maximise power.

5.2.2.2 Change in activity space shape

A measure of compactness based on the ratio of length by width of minimum convex hull polygons was used (Section 4.2.6.2) to assess activity space shape. To measure within-participant change, compactness measured at phase 2 was subtracted from that at phase 4 for each temporal scale. As compactness is measured on a scale of 0 to 1, with 1 indicating a more compact score, a negative outcome relates to an activity space becoming less compact and a positive more compact over time.

Absolute changes in activity space shape were recorded. As compactness measures were already on a scale of 0 to 1 for each participant there was no need to create a measure of percentage change. The frequency and distribution of change showed a slight skew, which varied in direction across different temporal scales. Tertiles were therefore chosen to represent activity spaces becoming less compact, having little or no change in shape, or becoming less compact post-intervention.

5.2.3 Exposure to the busway

Two different measures of exposure to the busway were used for analyses: proximity to the busway from each participant's home address and use of the busway.

5.2.3.1 Proximity to busway

The distance to the nearest busway stop or path access point from each individuals' home address, as reported in the core questionnaire, had been calculated for phases 2 and 3 of the study as part of a previous evaluation [249]. In line with that analysis, it was assumed that there will be some distance decay whereby a given increment in distance will have a smaller effect on use the further away from the busway. A linear relationship with distance was therefore not appropriate, instead proximity was defined as the square root transformation of distance.

Distance to the busway was calculated for nine participants who moved home address between phases 3 and 4, as this was not available in the existing dataset. The method used by Heinen and colleagues was replicated by georeferencing the centroid of participants' most recent postcode in ArcGIS [249]. Using the Network Analyst Closest facility tool, the shortest distance between home and the nearest access point to the busway were calculated along post-intervention route networks developed by Heinen and colleagues.

5.2.3.2 Use of busway

As there was a long delay between construction and commencement of the guided bus service in 2011, and the pathway alongside the busway was accessible to the public before its opening, users of the busway were grouped into the four categories shown in Table 5.2. Use of the busway was determined separately from self-report data, and through examination of recorded GPS data. Generating two measures of busway use from different datasets allowed for different time frames of behaviour to be investigated. The GPS data captured busway use from the past 7 days of device wear which is temporally relevant to activity space measures. However, habitual use may not be represented due to the short data collection period and irregular behaviours might be described. Self-reported measures therefore provide complementary information; providing a more general measure of use by recording whether the busway has ever been used and how.

Table 5.2: Categories of busway use as applied to self-report and GPS data

Use of busway	Use of busway path (phase 2)	Use of bus or busway path (phase 4)
Never	x	x
Former	✓	x
Continued	✓	✓
New	x	✓

x = no report of use or positioning of GPS data along busway
✓ = use of busway reported or GPS data positioned along busway

To measure whether the busway was used at each phase (x/✓as shown in Table 5.2) using GPS data, individuals' activity spaces for each temporal scale and a linear feature of the busway were overlaid within ArcGIS. Individuals whose activity spaces intersected the busway polyline were identified and their GPS points cross-referenced with OS transport network data to triangulate busway use. For example, if a series of points followed the busway, as exemplified by the individual's data shown in Figure 5.4, participants were categorised as busway users. Intersections between activity spaces, GPS points and the busway were inspected visually to ensure users followed the busway route, rather than passing under it on another road. No minimum distance or time spent on the busway was applied to classify use.

Figure 5.4: Busway user as identified by GPS data
Contains OS data © Crown copyright and database right 2019

Participants were asked to self-report: (i) use of the guided bus and (ii) whether the busway path had been used for walking and cycling. The following question wording was used:

"Have you travelled on a guided bus in Cambridgeshire?"

"Have you walked or cycled along any part of the footpath or cycle path beside the guided busway?"

For the first question, binary yes/no categorical responses were recorded for both study phases and for the second question responses could be "no", "walked" and "cycled" and respondents could tick more than one box, if appropriate. The responses were used to group users based on the categories in Table 5.2. A measure of any busway use was chosen for this analysis as it is comparable with the definition of use derived from the GPS data. Measures of walking or cycling and use of the guided bus were also chosen to provide complementary information on ways the busway was used.

5.2.4 Sociodemographic characteristics, travel options and other covariates

Key characteristics of the sample including age, sex, education, and urban-rural status of home address have been recognised as important correlates of health, activity spaces, and physical activity behaviours [170] and were captured in the questionnaire. Socioeconomic status has also been associated with physical activity and travel behaviours. However, given the small sample size, limited information on income at all study phases, and findings from previous studies [246], [249], I focused on highest educational attainment as a marker of socioeconomic status. As the sample were highly educated, the variable was dichotomised as at least having obtained a degree (or equivalent), or less.

Information relating to travel behaviours and travel options of the sample in the questionnaire was also investigated including car ownership, self-reported distance to work, and whether participants usually actively commute or not. Rather than excluding movers, a binary variable indicating whether participants had moved work or home or not was derived to explore the association between moving and changes in activity space and to maintain as large a sample size as possible. I did not adjust for season which may affect commuting behaviours as data were seasonally matched (i.e. collected in the same month at both phases) for each participant. A measure of BMI was not investigated due to missing data and a small sample size which would have resulted in small cell sizes.

5.2.5 Data analysis

5.2.5.1 Exploratory analyses

Baseline sample characteristics were compared between the sample with data at phases 2 and 4 (full sample), the sample with GPS data at phase 2 and 4 (potential sample), and the samples included for analysis (analytic samples). Differences were tested using chi-squared tests for categorical data and t-tests for continuous data.

Descriptive analyses were undertaken to assess the mean activity space size and shape for groups of sociodemographic characteristics at each study phase. Two-sample t-tests and ANOVA tests were used to test differences between outcomes at phase 2 and 4 and sample characteristics. Paired t-tests were used to test for within-person changes for each group.

Data were also explored using descriptive techniques including bar graphs to identify the prevalence and distribution of activity space changes by travel behaviours and options.

5.2.5.2 Regression models

Multinomial logistic regression models were used to assess the relationships between sociodemographic variables, exposure to the busway, and categorical changes in activity space size and shape.

Baseline values from phase 2 of the study were used for urban-rural status and mean distance to work and mean proximity to the busway were used to account for changes in work and home address. Univariate and adjusted regression analyses were performed. Age, sex, and proximity to the busway, as well as any additional variables significantly associated (p<0.05) with change activity space size or shape from the univariate regression were included as explanatory variables in a single adjusted model.

5.2.5.3 *Sensitivity analysis*

As the daily path measure incorporates all movement made over the course of multiple days, if participants moved home or work address between phases, or made an irregular long trip at one study phase, a measure of change may be recorded that was not reflective of habitual activities. Distributions of change were therefore compared for the whole analytic sample, including movers, and for participants who had the same home and work location at both phases. A variable measuring whether participants moved or not was tested in univariate regression and multivariate regression was repeated on the sample of non-movers.

To account for irregular long trips, variables were created to reflect changes in mean activity space areas for each temporal scale, calculated by dividing the total daily path area by the number of valid days. However, these were not used in the final analyses as it was more intuitive to use total movement, particularly when measuring shape, and when assessing individual maps and profiles.

5.2.5.4 *Individual profiles*

A combination of qualitative and quantitative data was used to evaluate individual profiles of four participants. Data included maps of GPS-measured mobility, self-reported information on travel and busway use, findings from the exploratory and regression analyses, and detailed information from qualitative interview transcripts.

I mapped individuals' activity spaces and GPS traces alongside a linear feature of the busway and OS data for each study phase and temporal scale. Activity spaces and GPS traces were visually inspected to identify how use of space had changed and which sections of the busway had been accessed. Interview transcripts were also reviewed for each participant. The use of interview data was not intended as a formal qualitative analysis, rather it was used to provide context for quantitative findings and to identify potential ways the busway was used and why, as well as possible mechanisms for changes in use of space [256], [257]. In line with this aim, three topics were outlined a priori: i) how the busway was used, ii) its reasons for use and non-use, and iii) how its use may relate to levels of physical activity. I extracted all relevant quotes from the transcripts and grouped them by topic. Quotes were also used to annotate individuals' maps and illustrate potential explanatory factors for changes in spatial patterns of behaviour. Findings from the maps and interview quotes were discussed narratively by topic.

To preserve confidentiality and for the purposes of reporting, participants were given pseudo-identification numbers.

5.3 Results

5.3.1 Sample characteristics

Data were collected at both phase 2 and phase 4 of the study for 444 participants (full sample), of whom 78 (17.6%) had GPS data at both phases (potential sample). Eleven participants (14.1% of potential sample) were excluded for week level analysis, 15 (19.2%) for weekday, and 7 (9%) for weekend due to insufficient number of days of valid data.

The characteristics of the full sample, potential sample, and samples included for analysis are detailed in Table 5.3. When compared with the full sample, smaller proportions of females, urban dwellers, and car owners were observed for the potential sample (those participants with GPS data). The distribution of characteristics was similar across the potential and analytic samples. The majority of participants included in the analytic samples were female (54%-58.2%) with mean ages of 44.8 to 45.9 years at phase 2 of the study. Most of the included participants were educated to at least degree level, lived in urban areas, did not change home or work address between the phases, and only a small percentage did not own a car.

Information on education was missing for 7.7% of the full sample but was complete for all samples with GPS data.

Table 5.3: Baseline characteristics of participants with data collected at both phase 2 and phase 4 of the Commuting and Health in Cambridge study

	Full sample		**Potential sample**		**Analytic samples**					
	Data collected (n=444)		GPS data collected (n = 78)		Included for week analysis (n=67)		Included for weekday analysis (n=63)		Included for weekend analysis (n = 71)	
	n	%	**n**	%	**n**	%	**n**	%	**n**	%
Sex			*							
Male	**133**	30.0	**35**	44.9	**28**	41.8	**29**	46.0	**31**	43.7
Female	**277**	62.4	**43**	55.1	**39**	58.2	**34**	54.0	**40**	56.3
Age [Years]										
Mean (SD)	***45.6 (11.2)***		***45 (10.1)***		***45 (9.8)***		***45.9 (9.5)***		***44.8 (9.9)***	
<40	**141**	31.8	**24**	30.8	**20**	29.9	**17**	27.0	**22**	31.0
40-50	**125**	28.2	**26**	33.3	**23**	34.3	**23**	36.5	**24**	33.8
>50	**178**	40.1	**28**	35.9	**24**	35.8	**23**	36.5	**25**	35.2
Education										
Less than degree	**107**	24.1	**15**	19.2	**14**	20.9	**10**	15.9	**15**	21.1
Degree or higher	**303**	68.2	**63**	80.8	**53**	79.1	**53**	84.1	**56**	78.9
Moved work										
No	**361**	81.3	**68**	87.2	**58**	86.6	**56**	88.9	**62**	87.3
Yes	**76**	17.1	**10**	12.8	**9**	13.4	**7**	11.1	**9**	12.7
Moved home										
No	**371**	83.6	**63**	80.8	**55**	82.1	**52**	82.5	**59**	83.1
Yes	**69**	15.5	**15**	19.2	**12**	17.9	**11**	17.5	**12**	16.9
Urbanicity			*							
Urban	**301**	67.8	**43**	55.1	**39**	58.2	**35**	55.6	**40**	56.3
Rural	**143**	32.2	**35**	44.9	**28**	41.8	**28**	44.4	**31**	43.7
Car ownership			†							
None	**59**	13.3	**3**	3.8	**3**	4.5	**3**	4.8	**3**	4.2
One	**204**	45.9	**39**	50.0	**34**	50.7	**30**	47.6	**37**	52.1
More than one	**181**	40.8	**36**	46.2	**30**	44.8	**30**	47.6	**31**	43.7

**p*<0.01 †*p*<0.05 indicates significant difference between full sample and potential sample

5.3.2 Exploratory results: Activity space size

5.3.2.1 Activity space size by sociodemographic characteristics

Table 5.4 shows the mean activity space size according to sample characteristics at both phases, as well as the mean within-person changes. As highlighted in Chapter 4, weekday activity spaces appeared to be larger than weekend activity spaces. A mean increase of weekday activity space size was shown for the whole sample, and a decrease for weekend activity spaces which appeared to drive the overall week data, although there was a large degree of variation around the means.

No significant differences in activity space size were observed for age, sex, or education. Although non-significant, a mean decrease in activity space size was recorded for males for all temporal scales while females showed an increase in the size of week and weekday activity spaces, but not weekend. The youngest age group (<40 years) appeared to have the greatest decrease in mean activity space size, except for at the weekend where the oldest age group (>50 years) showed a greater decrease.

Significant differences in the size of activity spaces between urban and rural dwellers were shown at phase 4 of the study for week and weekday temporal scales with urban activity spaces being smaller. A significant difference was also shown for weekend activity spaces for the number of cars owned in each household. Participants owning no car showed larger decreases in activity space size, compared to those with two or more cars. The changes appeared largest for weekend activity spaces. However, there was only a small number of individuals in the strata for no car ownership.

Table 5.4: Mean activity space size (km^2) by sociodemographic characteristics

	Week			Weekday			Weekend		
	Phase 2	Phase 4	**Individual change between phases 2 and 4**	Phase 2	Phase 4	**Individual change between phases 2 and 4**	Phase 2	Phase 4	**Individual change between phases 2 and 4**
	Mean (SD)	Mean (SD)	**Mean (SD)**	Mean (SD)	Mean (SD)	**Mean (SD)**	Mean (SD)	Mean (SD)	**Mean (SD)**
	14.8 (12.5)	13.2 (13.2)	**-1.5 (15.8)**	8.8 (8.1)	9.1 (11.4)	**0.3 (12.6)**	7.5 (7.8)	6.1 (7.5)	**-1.5 (10)**
Sex									
Male	16.1 (16.1)	12.3 (11.3)	**-3.8 (14.9)**	9.2 (10.3)	8.9 (9.8)	**-0.3 (11.1)**	8.4 (9)	5.7 (6)	**-2.7 (9.2)**
Female	13.8 (9.2)	13.8 (14.6)	**0.1 (16.4)**	8.5 (5.8)	9.4 (12.8)	**0.9 (14)**	6.9 (6.8)	6.4 (8.6)	**-0.5 (10.6)**
Age [Years]									
<40	16.5 (15.3)	10.9 (7.8)	**-5.6 (15.3)**	9.5 (9.2)	5.0 (5.2)	**-4.4 (10.0)**	8.2 (9.3)	6.9 (7.2)	**-1.3 (9.3)**
40-50	11.8 (8.6)	15.8 (17.4)	**3.9 (18.2)**	6.3 (5.1)	11.2 (14.6)	**4.9 (15.8)**	6.7 (6.9)	6.6 (9.4)	**-0.1 (12.2)**
>50	16.1 (13.1)	12.71 (12.2)	**-3.4 (12.5)**	10.8 (9.3)	10.1 (10.9)	**-0.7 (9.3)**	7.8 (7.5)	4.9 (5.8)	**-2.9 (8.4)**
Education									
Less than degree	13.8 (9.1)	10.8 (9.6)	**-2.9 (10.5)**	7.7 (4.2)	8.7 (6.8)	**0.9 (8.4)**	7.8 (7.1)	5.4 (8.2)	**-2.4 (10.3)**
Degree or higher	15 (13.4)	13.8 (14)	**-1.2 (17)**	9.0 (8.7)	9.2 (12.2)	**0.2 (13.4)**	7.5 (8.1)	6.3 (7.4)	**-1.2 (10)**
Urbanicity		*			*				
Urban	13.7 (12.6)	9.9 (7.7)	**-3.8 (12.9)**	8.1 (7.1)	6.2 (5.6)	**-1.8 (8.7)**	7.2 (8.7)	5.3 (6.6)	**-1.9 (9.9)**
Rural	16.2 (12.6)	17.8 (17.5)	**1.7 (18.9)**	9.7 (9.3)	12.8 (15.4)	**3.1 (16)**	7.9 (6.7)	7.1 (8.6)	**-0.8 (10.3)**
Car ownership									**†**
None	22.6 (15.2)	4.4 (2.5)	**-18.2 (14.7)**	7.1 (4.5)	4.0 (2.3)	**-3.2 (6.1)**	16.7 (13.8)	1.9 (1.2)	**-14.9 (13.4)**
One	13.8 (11.0)	13.5 (15.4)	**-0.3 (17.4)**	7.0 (6.2)	9.8 (14.3)	**2.8 (13.7)**	7.9 (7)	6.1 (6.6)	**-1.8 (9.5)**
Two or more	15.1 (14.0)	13.8 (10.9)	**-1.3 (13.4)**	10.8 (9.7)	9.0 (8.4)	**-1.7 (11.8)**	6.3 (7.9)	6.5 (8.8)	**0.2 (9.6)**

**$p<0.001$ *$p<0.01$ †$p<0.05$ indicates significant difference between groups at each phase. Differences between phases for each group also tested, symbols shown in bold.

5.3.2.2 *Change in activity space size by travel options and behaviours*

Figures 5.5 to 5.8 show the distribution of participants across categories of distance to work, usual commute mode, proximity to the busway, and use of the busway (self-reported and GPS-derived), respectively. The bars represent the absolute percentage of the total number of participants with valid week data and are stratified by categories of change in size of weekly activity space. The mean changes in activity space size (km^2) are shown as points for the full sample and for non-movers only. Results for weekday and weekend activity spaces are included in Appendix D1.

Twenty-five per cent of participants lived within 5 km of their workplace (Figure 5.5). The majority of whom showed an increase in activity space size after the opening of the busway which appeared to be driven by weekend results (Appendix D1, Figure D.4). The greatest proportion of participants lived over 20 km from their workplace (36%) and for those living furthest away, categories of change in activity space size were evenly distributed for all temporal scales.

Most participants actively commuted at both study phases (36%) but 34% of participants did not actively commute at all (Figure 5.6). The relative differences across categories of change were similar for both groups at each temporal scale. Very few participants switched from an active mode of travel to passive and 18% showed an uptake of active travel at phase 4. For those that did start actively commuting, most did not change their activity space size, although this was not shown for weekday days where an increase in activity space size was recorded for the majority (Appendix D1, Figure D.2).

A small majority of participants living closest to the busway decreased their activity space size (Figure 5.7), however, for weekday data, an increase was shown (Appendix D1, Figure D.3). For most participants living furthest from the busway, their activity space size did not change.

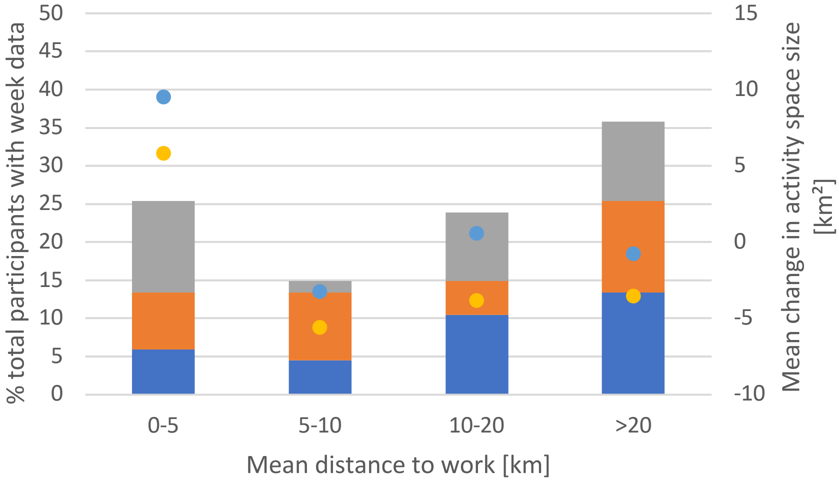


Figure 5.5: Change in activity space size by mean distance to work between study phases

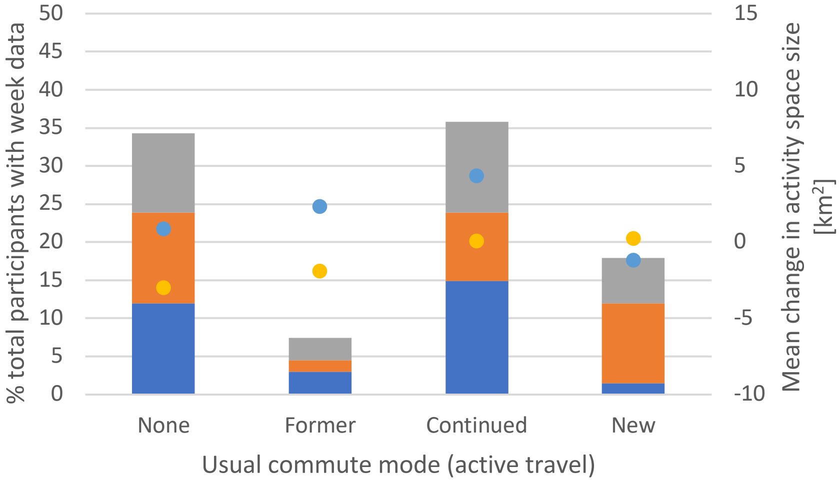


Figure 5.6: Change in activity space size by whether participants actively commuted

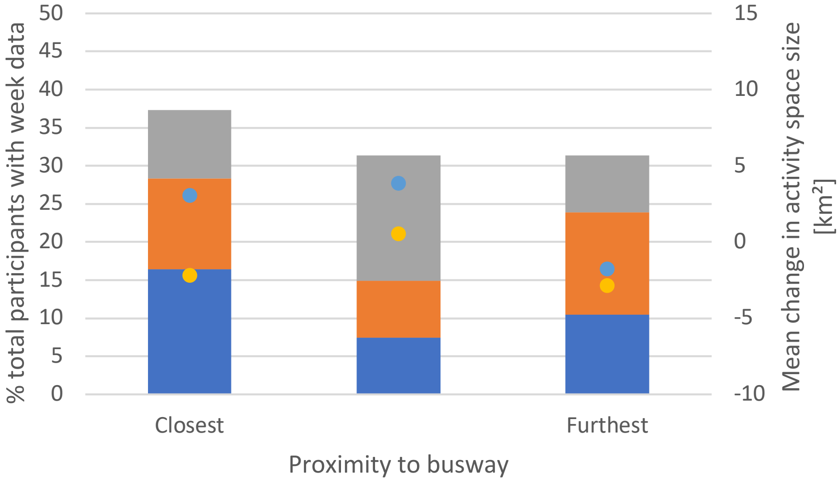


Figure 5.7: Change in activity space size by mean proximity from home address to busway between phases 2 and 4

Change in activity space size: Decrease No change Increase
Mean change in activity space size: All sample Non-movers only

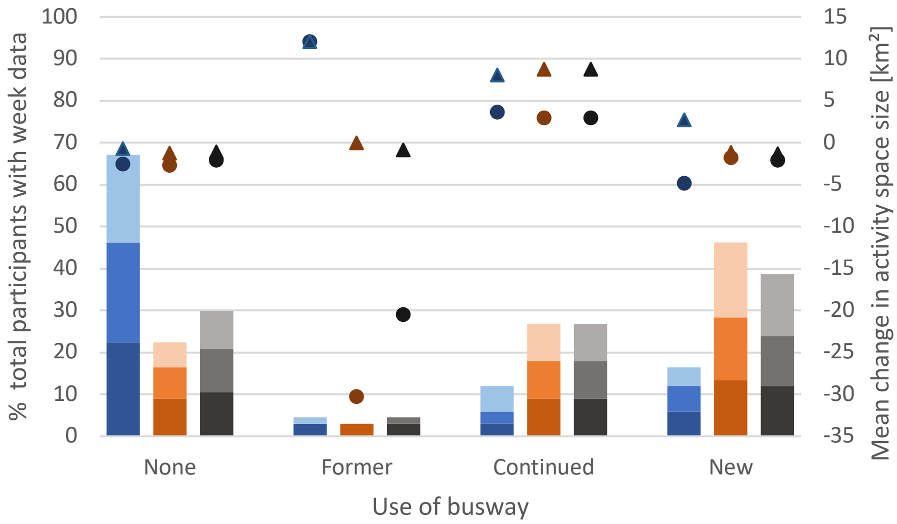


Figure 5.8: Change in activity space size by use of busway

GPS measure of use: Decrease No change Increase
Self-reported use: Decrease No change Increase
Self-reported walk or cycle: Decrease No change increase
Mean change in activity space size: ○ All sample △ Non-movers only

Figure 5.8 shows busway use as measured from different data sources. The GPS data showed that most participants' activity spaces did not intersect the busway with 94% of the sample with weekend data being categorised as non-users in the last 7 days (Appendix D1, Figure D.8). The categories of change for each temporal scale were equal for GPS-measured non-users. In contrast, the self-reported measures of busway use (indicating any use) showed a much smaller proportion of non-users (25-36%) with the majority reporting some new use of the busway at phase 4 (43%), mostly for walking or cycling (37%). For weekday activity spaces, a small majority of continued and new users of the busway showed an increase in activity space size which was reflected in the mean values of change. This observation was less clear for weekend activity spaces (Appendix D1, Figure D.8).

Across Figures 5.5 to 5.8, the mean change in activity space size was similar for movers and non-movers. The largest differences were shown for self-reported former busway users where activity space size appeared to decrease for the whole sample and not for non-movers. However, the sample size was very small for these groups (3-4% of all participants with week data).

5.3.2.3 Associations between sociodemographic and geographical characteristics and changes in activity space size

Adjusted associations of sociodemographic characteristics and exposure to the busway with changes in activity space size are presented in Table 5.5. In univariate models, urban-rural status was associated with a change in the size of weekend activity spaces. After adjustment, the association persisted with rural dwellers less likely to increase their weekend activity space size compared with urban dwellers (relative risk ratio [RRR]: 0.22, 95% CI: 0.06, 0.81). Those living further from the busway were less likely to have increased their weekday activity space size than those living further away (RRR: 0.49, 95% CI: 0.27, 0.86). This suggests spatial patterning of weekday behaviour was different for those more exposed to the busway.

Sensitivity analyses showed that moving home or workplace was not associated with change in activity space size in the univariate regression. Performing multivariate regression on the sample of participants that did not move between phases 2 and 4, the effects for urban-rural status and busway proximity remained broadly similar both in terms of direction and statistical significance (Appendix D1, Table D.1).

Table 5.5: Adjusted associations between sociodemographic and geographic characteristics and exposure to the busway with change in activity space size

	Week		**Weekday**		**Weekend**	
	Decrease RRR (95% CI)	Increase RRR (95% CI)	Decrease RRR (95% CI)	Increase RRR (95% CI)	Decrease RRR (95% CI)	Increase RRR (95% CI)
Urban rural status						
(ref: urban)	*n.i*	*n.i*	*n.i*	*n.i*		*
Rural					0.41 (0.12, 1.44)	**0.22 (0.06, 0.81)**
Proximity to busway						
(ref: closest)				*		
[square root of mean distance]	0.94 (0.60, 1.46)	0.96 (0.61, 1.52)	0.72 (0.44, 1.18)	**0.49 (0.27, 0.86)**	0.75 (0.47, 1.20)	0.77 (0.48, 1.24)

Model adjusted for age, sex, and significant variables from univariate analyses.
Bold text indicates statistical significance (***p*<0.001 **p*<0.01 †*p*<0.05)
RRR – relative risk ratio; CI – confidence interval; n.i – not included in model

5.3.3 Exploratory results: activity space shape

5.3.3.1 Activity space shape by sociodemographic characteristics

Table 5.6 shows the mean measure of activity space compactness according to the sample characteristics, as well as the mean within-person change in compactness between both phases. There appears to be little difference in the mean compactness of activity spaces for the whole sample for each temporal scale. However, weekend activity spaces became less compact on average unlike the week activity spaces which became more compact over time.

Differences in mean changes in compactness for sex, age, weight status, and urban rural status were all minimal and non-significant. Despite this, different trends were shown for sex with week and weekday activity spaces becoming less compact for males and more compact for females. On average, activity spaces became more compact for younger participants as well as those living in urban areas.

There is some evidence of a trend for week activity spaces by car ownership at phase 2 with less compact shapes shown for those with more than one car in the household. The trends for mean change show activity spaces for participants with one car becoming less compact for all temporal scales, however this is not reflected for those with more than one car and changes are all non-significant.

Table 5.6: Mean activity space shape (compactness) by sociodemographic characteristics

	Week			Weekday			Weekend		
	Phase 2	Phase 4	**Individual change between phases 2 and 4**	Phase 2	Phase 4	**Individual change between phases 2 and 4**	Phase 2	Phase 4	**Individual change between phases 2 and 4**
	Mean (SD)	Mean (SD)	**Mean (SD)**	Mean (SD)	Mean (SD)	**Mean (SD)**	Mean (SD)	Mean (SD)	**Mean (SD)**
	0.34 (0.16)	0.35 (0.15)	**0.01 (0.22)**	0.34 (0.18)	0.34 (0.16)	**0 (0.2)**	0.35 (0.19)	0.35 (0.17)	**-0.01 (0.23)**
Sex									
Male	0.36 (0.14)	0.33 (0.13)	**-0.03 (0.17)**	0.32 (0.15)	0.31 (0.16)	**-0.02 (0.18)**	0.37 (0.21)	0.36 (0.14)	**0 (0.23)**
Female	0.32 (0.17)	0.36 (0.17)	**0.04 (0.25)**	0.35 (0.2)	0.36 (0.16)	**0.01 (0.21)**	0.35 (0.18)	0.33 (0.19)	**-0.01 (0.22)**
Age [Years]									
<40	0.34 (0.17)	0.35 (0.14)	**0.01 (0.22)**	0.37 (0.2)	0.38 (0.18)	**0.01 (0.23)**	0.31 (0.16)	0.32 (0.14)	**0.01 (0.19)**
40-50	0.33 (0.15)	0.35 (0.14)	**0.02 (0.21)**	0.3 (0.16)	0.31 (0.11)	**0 (0.18)**	0.37 (0.17)	0.35 (0.19)	**-0.01 (0.24)**
>50	0.35 (0.16)	0.35 (0.17)	**0 (0.24)**	0.35 (0.18)	0.33 (0.18)	**-0.02 (0.21)**	0.38 (0.23)	0.36 (0.17)	**-0.02 (0.25)**
Weight status [BMI kg/m²]									
Underweight/normal	0.32 (0.14)	0.34 (0.13)	**0.01 (0.18)**	0.31 (0.15)	0.34 (0.12)	**0.04 (0.15)**	0.35 (0.18)	0.35 (0.16)	**-0.01 (0.23)**
Overweight/obese	0.34 (0.16)	0.41 (0.15)	**0.08 (0.23)**	0.33 (0.16)	0.37 (0.19)	**0.04 (0.21)**	0.37 (0.2)	0.39 (0.19)	**0.01 (0.22)**
Education									
Less than degree	0.36 (0.17)	0.36 (0.17)	**0 (0.27)**	0.34 (0.15)	0.36 (0.17)	**0.02 (0.2)**	0.39 (0.2)	0.37 (0.16)	**-0.02 (0.23)**
Degree or higher	0.33 (0.16)	0.35 (0.15)	**0.01 (0.21)**	0.34 (0.18)	0.33 (0.16)	**-0.01 (0.2)**	0.35 (0.19)	0.34 (0.17)	**-0.01 (0.23)**
Urbanicity									
Urban	0.35 (0.17)	0.36 (0.17)	**0.01 (0.24)**	0.37 (0.2)	0.37 (0.18)	**0 (0.21)**	0.35 (0.18)	0.36 (0.17)	**0.02 (0.22)**
Rural	0.32 (0.15)	0.33 (0.12)	**0 (0.19)**	0.3 (0.14)	0.3 (0.12)	**0 (0.19)**	0.37 (0.2)	0.33 (0.17)	**-0.04 (0.23)**
Car ownership	†								
None	0.31 (0.15)	0.36 (0.13)	**0.05 (0.22)**	0.28 (0.18)	0.33 (0.15)	**0.05 (0.27)**	0.24 (0.09)	0.31 (0.06)	**0.07 (0.05)**
One	0.39 (0.14)	0.34 (0.15)	**-0.05 (0.21)**	0.38 (0.19)	0.33 (0.17)	**-0.04 (0.2)**	0.38 (0.18)	0.37 (0.19)	**-0.01 (0.21)**
More than one	0.29 (0.17)	0.36 (0.15)	**0.07 (0.22)**	0.31 (0.17)	0.34 (0.15)	**0.03 (0.19)**	0.34 (0.21)	0.32 (0.14)	**-0.01 (0.26)**

**p<0.001 *p<0.01 †p<0.05 indicates significant difference between groups at each phase. Differences between phases for each group also tested, symbols shown in bold.

5.3.3.2 *Change in activity space shape by travel options and behaviours*

Figures 5.9 to 5.12 show the distribution of participants across categories of travel options and behaviours, stratified by categories of change in activity space shape based on week activity spaces. Results for weekday and weekend activity spaces are included in Appendix D1.

For participants living closest to their workplace, week activity spaces became less compact for a small majority (Figure 5.9). In contrast, weekday activity spaces became more compact for most participants living within 5 km of their workplace but less change was shown for weekend activity spaces (Appendix D1, Figure D.12). In general, the mean values of change showed that activity spaces of those living closest to their workplace became less compact while those of participants furthest away became more compact.

The categories of change in activity space compactness for participants who continued to actively commute or did not actively commute at all were distributed evenly with similarities in the relative differences across categories for both groups (Figure 5.10). For those who started to actively commute, most recorded no change in compactness in weekday activity spaces but a mean increase in compactness for weekend activity spaces (Appendix D1, Figure D.13).

Most participants living closest to the busway did not change the compactness of their week activity space (Figure 5.11), however more variability is shown for weekday activity spaces with an increase in compactness shown for a small majority (Appendix D1, Figure D.11). For those living furthest from the busway, a decrease in week activity space compactness is shown for the majority which appears to be driven by weekend data (Appendix D1, Figure D.14).

Figure 5.12 shows change in activity space shape by busway use. In general, new users (classified using both self-reported and GPS-measures) recorded an increase in activity space compactness whereas less change was shown for continued and non-users.

As with mean changes in activity space shape, there was little difference in mean change in compactness for movers and non-movers.

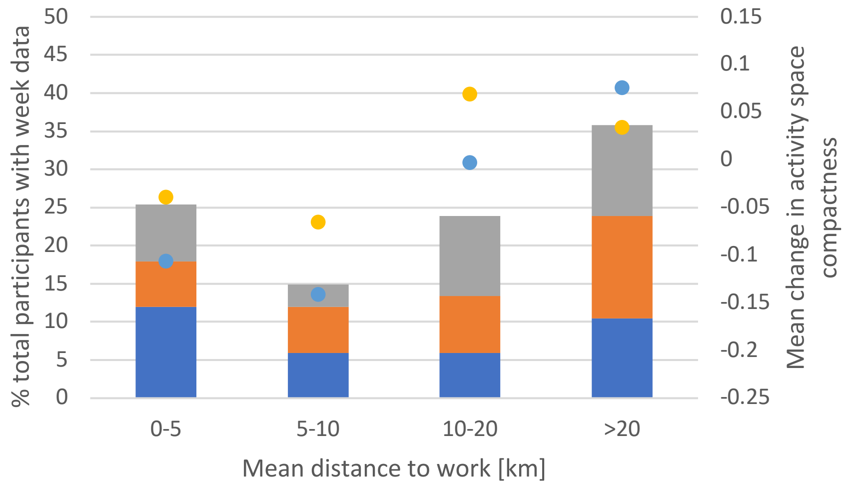


Figure 5.9: Change in activity space shape by mean distance to work between study phases

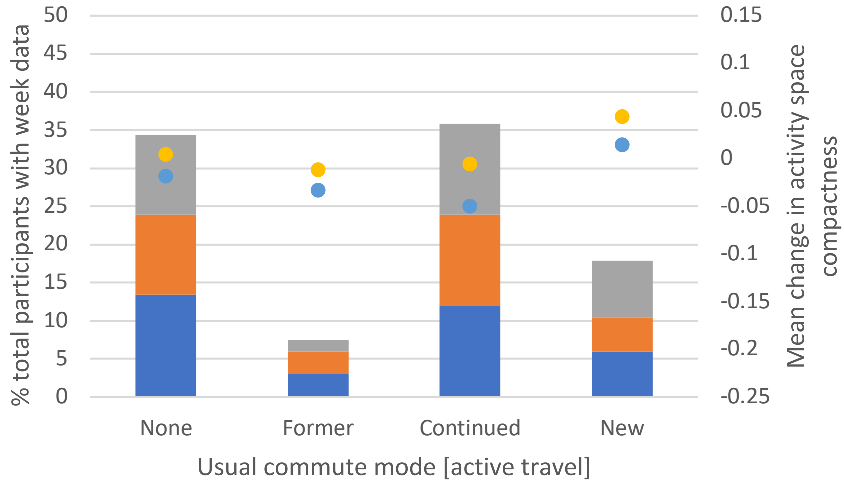


Figure 5.10: Change in activity space shape by whether participants actively commuted

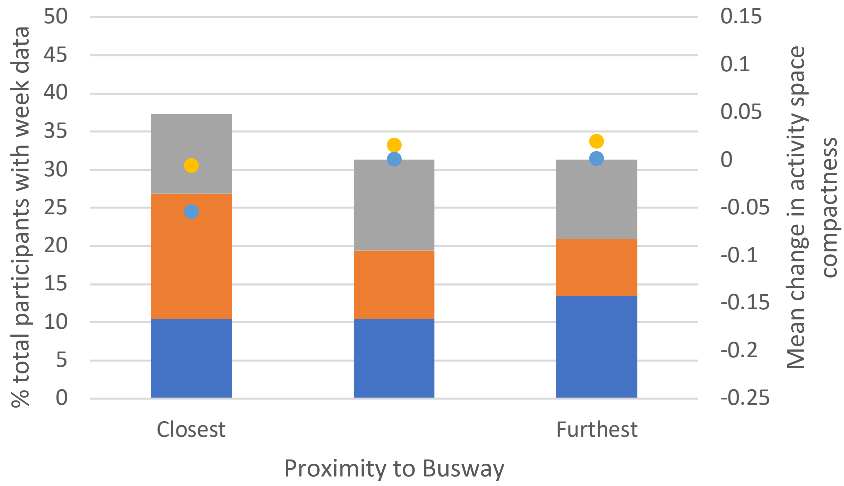


Figure 5.11: Change in activity space shape by proximity to busway

Change in activity space shape: Less compact No change More compact
Mean change in activity space size: All sample Non-movers only

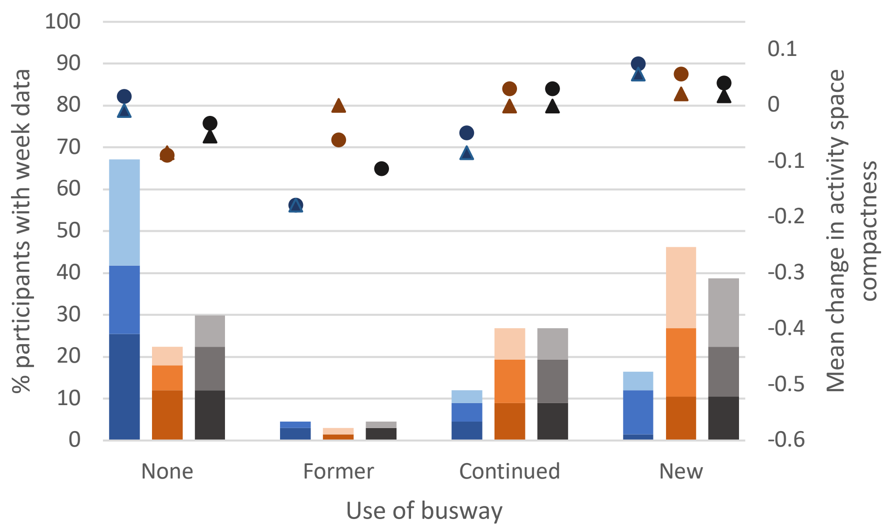


Figure 5.12: Change in activity space shape by use of busway

GPS measure of use:	Less compact	No change	More compact
Self-reported use:	Less compact	No change	More compact
Self-reported walk or cycle:	Less compact	No change	More compact
Mean change in activity space shape :	○ All sample	△ Non-movers only	

5.3.3.3 Associations between sociodemographic and geographical characteristics and changes in activity space shape

Unlike activity space size, urban-rural status was not associated with change in activity space shape in univariate analyses, but car ownership was. However, after adjustment (Table 5.7), the relationship between car ownership and activity space compactness became non-significant. Proximity to the busway was not significantly associated with change in activity space shape at any temporal scale.

In sensitivity analyses, univariate regression showed that moving home or work was not associated with changes in activity spaces shape. After repeating the multivariate regression on the sample that did not move, all relationships remained non-significant (Appendix D1, Table D.2).

Table 5.7: Adjusted associations between sociodemographic and travel characteristics change in activity space shape

	Week		Weekday		Weekend	
	Less compact RRR (95% CI)	More compact RRR (95% CI)	Less compact RRR (95% CI)	More compact RRR (95% CI)	Less compact RRR (95% CI)	More compact RRR (95% CI)
Car ownership (ref: none)	*n.i*	*n.i*			*n.i*	*n.*
One or more cars			0.32 (0.10, 1.08)	0.82 (0.24, 2.76)		
Proximity to busway (ref: closest)						
[square root of mean distance]	1.25 (0.78, 2.01)	1.28 (0.79, 2.05)	0.81 (0.49, 1.33)	0.78 (0.47, 1.29)	1.05 (0.68, 1.61)	1.03 (0.66, 1.60)

Model adjusted for age, sex, and significant variables from univariate analyses.
Bold text indicates statistical significance (**p<0.001 *p<0.01 †p<0.05)
RRR – relative risk ratio; CI – confidence interval; n.i – not included in model

5.3.4 Individual profiles

Based on the exploratory analysis from the quantitative questionnaire data (n=67 week; n=63 weekday; and n=71 weekend analysis), it appeared that individual characteristics were only associated with activity space size, not shape. The four individual profiles bring together data from the quantitative and qualitative datasets to provide more detailed contextual information. The characteristics for each individual, their self-reported and GPS-measured busway use, and changes in shape and size of activity space are shown in Table 5.8.

Distance to work, usual commute mode, and proximity to the busway varied across participants and participant 3 was the only person without access to a car. Use of the busway was self-reported by both urban and rural dwellers and by participants with long and short commutes (participants 2, 3, and 4). In contrast, GPS-measured use of the busway was only recorded for urban dwellers (participants 3 and 4), both of whom lived in towns outside of Cambridge.

Maps of activity spaces at different temporal scales for GPS-measured users of the busway (participants 3 and 4) are presented in Figure 5.13 and Figure 5.14. The maps show the extent and locations of each individual's movement, and how their activity spaces changed between the two study phases. Spatial patterns of movement are also presented alongside qualitative interview data to provide insight into how and why the busway was used and the effect of its use on activity spaces and physical activity. Drawing on the information shown in the maps and qualitative data, these three topics are discussed in greater detail in the following text.

Table 5.8: Characteristics and travel behaviours of participants included in qualitative analysis

Participant ID	1	2	3	4
Characteristics				
Age (phase 2)	63	31	44	42
Sex	Female	Female	Male	Female
Urban rural status (phase 2)	Rural	Rural	Urban	Urban
Moved home	No	No	Yes	No
Moved work	No	No	No	No
Travel options and behaviours				
No. cars (phase 2)	2	1	0	1
No. cars (phase 4)	2	1	0	1
Distance to work (phase 2)	>20 km	0-5 km	10-20 km	>20 km
Distance to work (phase 4)	>20 km	0-5 km	>20 km	>20 km
Usual active commute	None	None	Former	New
Proximity to busway (phase 2)	Close	Mid	Close	Far
Proximity to busway (phase 4)	Close	Mid	Close	Far
Self-reported measures of busway use				
Use of busway	None	New	Continued	Continued
Use of busway for walking or cycling	None	New	Continued	Continued
Use of guided bus	No	Yes	Yes	No
GPS measures of busway use and change in activity space features				
Week:				
Use of busway	None	None	New	Continued
Change in activity space size	Decrease	Increase	Decrease	No change
Change in activity space shape	No change	Less compact	No change	Less compact
Weekday:				
Use of busway	None	None	New	New
Change in activity space size	Decrease	No change	Increase	Increase
Change in activity space shape	No change	More compact	More compact	No change
Weekend:				
Use of busway	None	None	None	Continued
Change in activity space size	Decrease	Increase	Decrease	No change
Change in activity space shape	More compact	Less compact	No change	More compact

Please note that this figure has been redacted for online submission as it contains potentially sensitive participant data. Please contact the author to request access to the figure

Figure 5.13: Maps of weekday and weekend activity spaces for participant 3

Please note that this figure has been redacted for online submission as it contains potentially sensitive participant data. Please contact the author to request access to the figure

Figure 5.14: Maps of weekday and weekend activity space for participant 4
Contains OS data © Crown copyright and database right 2019

Topic 1: Patterns of busway use

After the opening of the busway, both the northern and southern sections of the route were used as alternative routes for commuting by participants 3 and 4. Rather than reporting a regular mode of travel, both the interview and GPS data for participant 3 show how a range of travel options to make the journey between St Ives and Cambridge were available to them. Participant 3 described how they switched between modes; travelling by guided bus or bicycle along the busway and the regular bus along the A14. Figure 5.13 shows how participant 3 started to use the busway at phase 4. In Panel A, the weekday activity space for phase 4 shown in green captures the participant's new commute route into Cambridge after relocating from Histon to St Ives. The route includes the entire northern stretch of the busway and use of an alternative route along the A14 to make the same journey is also shown. Their interview data corroborates this:

"I can go on the regular bus, the guided bus or I can cycle, and I do a bit of all of them" [Participant 3, 44 years]

In Figure 5.14, use of the northern stretch of the busway is captured for participant 4 for both weekdays and weekends. In contrast to participant 3, participant 4 described how they typically drive into Cambridge during the week and then walk from a free parking space to their workplace:

"Generally I drive from the village where I live, approximately 25 miles away… to where my partner works in Cambridge, and he has free parking there… then I walk to the Addenbrookes site" [Participant 4, 42 years]

This regular pattern of travel is shown in their GPS data (Figure 5.14, Panel A) through use of the A14 at both study phases. The participant also described how they occasionally use the busway to cycle or walk along for commuting purposes, which may be reflected in their weekday activity space at phase 4, and regularly use the busway to walk or cycle for leisure at weekends which is captured clearly in Panel B.

"Occasionally if my partner's not working then I will park at the Trumpington Park and Ride and then I'll walk along the Guided Busway route through the fields"

"we'll often use the busway at weekends to cycle on" [Participant 4, 42 years]

Although participant 2's activity space did not intersect the busway, they reported new use of the busway and guided bus since it opened. Participant 2 regularly commuted via scooter and similar to participant 4, they described how the have used the busway to cycle along for leisure.

"the Vespa is my main mode of commuting, particularly if I need to go into town… I have a wholly unsuitable bike for commuting"

"I have [used] the north route, and maybe cycled five miles… it's good to do time trials on it" [Participant 2, 31 years]

The only participant who did not report or record any use of the busway lived in a rural area, had two cars in the household and did not actively commute due to health reasons (participant 1). After falling unwell, they switched from using the local bus service to driving.

"when I first started the study I was taking the bus… [then] I had a car parking pass for the duration of my illness" [Participant 1, 63 years]

<u>Topic 2: Reasons for use and non-use</u>

The GPS and interview data shed light on the reasons for use at particular times or for particular purposes. New weekday use of the busway captured in Figure 5.13, Panel A by participant 3 was driven by their new route to work after relocation. Travel from St Ives into Cambridge was not previously required, however, the presence of the busway and convenience of travel routes influenced their choice to relocate.

"Because I don't drive, moving house and transport… had to be combined… We could have a bigger house that we could afford and… we looked into the alternatives for commuting and they were OK from St Ives" [Participant 3, 44 years]

The weekday use of the busway was dependent on the convenience of the guided bus. The quality and cleanliness of the guided buses were recognised, however, barriers to use were also revealed. Due to irregular times or incoherence with the local buses, both local buses and the busway were used by Participant 3, as shown by their activity space. The participant described concerns over operating times and busyness of the guided bus service which was corroborated by Participant 1 who, despite living close to the busway, described how they would use the local bus over the guided bus to support local services. Whether the busway was used was often therefore influenced by alternative options available, or lack of.

"If you're at the Park and Ride and you try and get a bus at the Park and Ride at 7.30 in the morning, some of the buses are very full. And by the time they get to Longstanton, you know, it has been known that there are people standing" [Participant 3, 44 years]

"[the local bus is] a good service and I know the people on it and you get to know the drivers and it's just a friendly atmosphere, so I would consider that before taking the Guided Bus"

"I want to get in [work] early and the one [guided] bus that starts that time to get me here is always full… and also I do like to support local firms" [Participant 1, 63 years]

The mode of travel used on the busway (Figure 5.13) was also dependent on weather conditions, with participant 3 choosing to cycle for enjoyment and efficiency when the weather suited. The participant also noted that they were only able to cycle the long distance shown by their activity space due to shower facilities available at their workplace.

"I quite like cycling, so the distance isn't so bad. I mean, I try and do it several times a week… it really just depends what the weather's like."

"the cycle path by the side of the busway is the quickest route. And it's quite nice for cycling, you know. It is black tarmac all the way and is flat."

"The only reason I can [cycle] and do that distance is that we do have facilities at work… if there wasn't a shower there, I wouldn't even contemplate it." [Participant 3, 44 years]

The intention to cycle via the busway in good weather was echoed by participant 4 despite the long commute, which may explain the use of the busway captured in Figure 5.14, Panel A. The quality of the cycle path and directness of route was appraised for pleasant cycling and reducing travel times. In Panel B, participant 4's GPS trace at phase 4 of the study showed a deviation from the busway to local nature reserves. The participant welcomed the access to local nature reserves that the busway provided which they walked through at weekends after the busway opened.

"…it's a 50 mile round cycle ride so the weather conditions have to be perfect and I have to be full of energy… but it's really nice cycling along the busway"

"it's brought a lot of positives in terms of the cycle way along it, access to various wildlife areas that I wasn't aware of before the busway, certainly around St Ives, there are a lot of nice little nature reserves that you can access… I think the busway has made that more accessible to people" [participant 4, 42 years]

<u>Topic 3: Potential displacement of activity and take up of new activity</u>

Using the GPS data, the weekday activity spaces increased for both participants who used the busway during the period of monitoring. Figure 5.13 and Figure 5.14 show how both participants used the busway at phase 4 in addition to alternative routes used at phase 2, to travel to and from the north of Cambridge to the city centre. Participant 3's weekend activity space decreased over time which was driven by a long trip taken at phase 2 (Figure 5.13, Panel

B). In contrast, there was no change for the size of participant 4's activity space as they used the busway path before the busway opened in 2011 and continued to use the busway for leisure at weekends during phase 4 (Figure 5.14, Panel B).

The interviews revealed that the use of the busway captured in Figure 5.13 and Figure 5.14 allowed for active commuting or walking or cycling for leisure to be undertaken in a new space. Most participants discussed the personal importance of physical activity with some preferring to be active during leisure time whilst others described how they could incorporate it into their habitual routines.

"the bicycle is a… really good exercise thing, and it's a very good enjoyment thing, but going into the city centre, the enjoyment is cut out" [Participant 2, 31 years]

"I use my commute as part of my exercise strategy, really. It saves me having to go to the gym. The gym's OK, but when the sun's shining I prefer to be out on my bike and exercising that way" [Participant 3, 44 years]

The distance of commute shown in Figure 5.13 and Figure 5.14 for participants 3 and 4 are too long to walk although the possibility to use the busway and incorporate mixed modes of travel into the commute, including walking and cycling, was acknowledged. However, concerns about carrying items as well as security and locking up bikes near the busway were raised which limits the desire to use the guided bus and cycle.

"There's a health and safety issue with transporting carbon dry ice which on certain public transport we wouldn't, the Trust… doesn't have the necessary insurance" [Participant 2, 31 years]

"I'd… consider driving part of the way. Just for convenience… [for] carrying thing."

"I don't think the security at any of the stops along the way is good enough to really leave my bike there for any extended period" [Participant 3, 44 years]

5.4 Discussion

5.4.1 Principal findings

Divergent findings for different temporal measures of activity spaces were observed in both the quantitative and qualitative data. In general, weekday activity spaces appeared larger than weekend activity spaces and showed a mean increase in size following the opening of the busway.

Exploratory quantitative analysis indicated that the size and shape of activity spaces, and how these changed in response to the intervention, varied according to urban-rural status, car ownership, and available travel options. Weekday activity spaces for rural dwellers were larger and less likely to change in size compared with urban dwellers, and weekday activity spaces for non-car owners were more compact than car owners' but more likely to change in shape and size over time. Living further from the busway was also associated with a lower likelihood of increasing the size of weekday activity spaces. The combination of mapped GPS traces and interview data revealed that the busway was used as a new space for walking and cycling for both commuting purposes and leisure.

5.4.2 Comparisons with existing evidence

Differences were observed for urban and rural dwellers, with participants living in rural areas typically having larger weekday activity spaces and being less likely to change the size of their activity space in response to the intervention. The size of activity spaces may be reflected in the need for rural dwellers to regularly travel greater distances to access services and workplaces. This aligns with existing evidence in the literature that suggests access to more urban environments is associated with smaller activity spaces [168], [178], [180].

Non-car owners tended to have more compact weekday activity spaces but were more likely to change the shape and size of their weekend activity spaces over time. This may be reflected in a propensity to change travel behaviours. For example, a longitudinal study of a UK sample which showed that 91.4 % of participants who commuted by car continued with the same mode of travel over the course of a year [258] and a study of students in Northern Ireland showed that that majority of car owners chose to commute by car [171]. Although a previous study of the Commuting and Health in Cambridge dataset showed a reduction in the proportion of trips made entirely by car [251], when considering car ownership and the responses of interviewees in my study, those who travelled by car or scooter continued with the same commuting patterns. Conversely, the participant without access to a car changed their commute to frequently incorporate the busway. This corresponds with a previous study that found activity spaces to be sensitive to the accessibility of public transport and car ownership [164], although there are few other studies on the topic.

Participants who lived furthest from the busway were less likely to increase the size of their weekday activity spaces. Intuitively, those participants may be less likely to make a change to their travel patterns in response to the opening of the busway due to its inaccessibility and reduced desire to incorporate its use. Distance from home locations to interventions has shown to be an important factor in determining use and having an effect on behaviour in previous studies [171], [176], including a study of the same dataset [249]. However, the individual profiles showed that the participants with GPS-measured use of the busway incorporated it into their mobility patterns, irrespective of how far they lived from an access point. Those participants not only used the busway but also alternative routes when perceived to be more favourable or convenient, which contributed to a larger activity space post-intervention. The two GPS-measured users of the busway lived in towns and villages and a previous study found effects of the busway to be stronger for individuals living in such areas [249]. Similarly, Kamruzzaman and colleagues found that participants living furthest from a transport intervention were less likely to use its service in the evening, not because the service itself was inaccessible, but because they were less inclined to make long trips at this time [171]. This highlights distinct patterns of behaviour within groups of participants, the complexity in promoting use of an intervention for all individuals and the importance of context.

The qualitative evidence confirmed that the busway provided a new space for travel and physical activity for some participants. Its use may therefore contribute to achieving recommended levels of activity with participants using it for both passive and active commuting, and walking and cycling for leisure [175], [176], [254]. Previous studies, including those of comparable walking and cycling infrastructure projects around the UK, have shown that increases in active travel are commensurate with increases in overall physical activity without compensating levels of recreational physical activity [253], [259]. However, regular and sustained use of the busway for physical activity may be challenging with issues regarding security for bikes and the provision of showers at workplaces identified as barriers to use, alongside poor provision of lighting highlighted in an existing study [248]. At present, it is unknown whether changes in spatial patterning of movement to incorporate the busway contributes to increases in overall activity or compensates previous activity undertaken elsewhere. Changes in the spatial patterning of physical activity are therefore explored further in Chapter 6.

5.4.3 Strengths and limitations

This is one of the first studies to use a longitudinal study design to evaluate changes in use of space after an intervention, and to use the concept of the activity space to achieve this understanding. Using an experimental design, it was possible to approximate a true change in transport infrastructure and provide interpretations of possible effects [54]. Comparisons were also drawn with other studies of the same and similar interventions to position findings in the context of related outcomes.

Although the sample size was small, the dataset provided an example of how a strong study design with a range of concurrent measures can mitigate issues of inferring causality from conventional regression analyses alone [260], [261]. From the outset of the study it was apparent that relying on quantitative data and statistical models was not appropriate. For example, focusing on effect sizes would have done little to strengthen causal inference and would not have provided insight into the role of the busway in facilitating potential changes in behaviour. Instead, due to the strength of the research design, it was possible to draw on quantitative and qualitative data and employ distinct and complementary methods to investigate how the use of an intervention may drive changes in spatial patterning of mobility.

A detailed descriptive analysis of change in activity spaces was performed. This provided a basis for analyses in Chapter 6 which focuses specifically on suitability of methods to measure changes in the location of physical activity. As the aim of the study was exploratory, the quantitative results were sufficient in providing insight into potential changes in travel behaviour. A key strength of the study was the ability to supplement quantitative results with interview data and visualisations of spatial information. In doing so, detailed case-based information was provided and triangulated with descriptive results to afford a stronger, more contextual package of evidence [262]. Qualitative data was used in this study to provide possible causal explanation and to identify potential mechanisms for change. For example, it was possible to ascertain that the busway provided an alternative route for commuting, an additional space for leisure activity, and a new route for accessing greenspaces which may lead to potential changes in physical activity and wellbeing. Exploring individual case studies therefore allowed for potential pathways and assumptions made from the quantitative results to be challenged in a way that cannot be achieved with larger effect sizes.

The use of GPS data allowed all locations visited over the course of a week by each participant to be objectively measured. It was therefore possible to derive activity space metrics which assessed the geographical extent of mobility and to objectively measure whether an individual’s GPS trace intersected the busway. Visually inspecting GPS traces against the mapped route of the busway accounted for potential issues with signal stray and accuracy of GPS receivers, meaning no users were identified as false positives. However, investigation into changes in specific routes taken before and after the busway opened was limited to the four participants with interview data. Additionally, using data from 4 to 7 days of GPS wear captured a narrow temporal window of participants’ mobility. In doing so, behaviours undertaken and locations visited less regularly or in different seasons may not have been included in the activity space metrics. Irregular trips captured in the week of GPS data collection may also have obscured more general behaviours. However, a sensitivity test performed in Chapter 5 comparing the outcomes of mean and total space sizes showed no significant differences, suggesting this effect may be small.

A difficulty with experimental studies is often the reliance on discrete pre and post measures of an intervention to categorise exposure. The use of GPS data within this study showed that participants accessed the busway path at phase 2 before the busway opened. This highlights the challenges with evaluative studies in real world settings where there is a lack of clean pre and post measures and potential unintended consequences in studies of similar interventions. I overcame this to some degree by categorising busway use as new, continued, former, or none, and comparing findings with self-reported results which may capture more occasional use. However, this meant that the number of participants in each group was small. The results are therefore exploratory rather than definitive and conclusions and implications take this into account.

The sample was not generalisable to the UK or the local population, with an over-representation of women, high levels of education, and a high prevalence of cycling at baseline compared with the rest of the UK [245]. However, there are other places in the UK, such as Oxford, with similar populations where results may be applicable. Qualitative data were analysed for only four participants, none of whom lived in Cambridge. However, sociodemographic characteristics and travel behaviours of the sample were heterogeneous. Individuals were also included who intuitively would be more likely to change travel routes to incorporate the busway, given its location in relation to their home addresses and Cambridge city centre.

5.4.4 Future research

The findings show that changes in the shape and size of activity spaces were divergent across different groups of participants and were sensitive to exposure to and use of the busway. Future research could aim to locate and quantify changes in physical activity to understand whether a change in mobility patterns is due to use of the busway, or other interventions, and whether physical activity increases, decreases or is substituted in a new location. This will be the subject of the next chapter.

I have shown that the methods developed and adopted in this study, including the cleaning of data, derivation of activity spaces, and combination of spatial and qualitative data could be used and further developed in the context of observational cross-sectional, longitudinal and evaluative studies. The methods for processing of GPS could also be rolled out in larger datasets.

Chapter 6

Changes in physical activity in response to a built environment intervention: evaluating the applicability of geospatial analysis methods

6.1 Introduction

Changes to the physical environment have the potential to affect physical activity behaviours and health at the population level. Chapter 5 explored how use of space changed after the opening of the Cambridgeshire Guided Busway using a combination of questionnaire, GPS and interview data. Findings suggest that some individuals' spatial patterning of movement changed to incorporate the busway. Specifically, the busway provided a new space for active and passive commuting, and was used in addition to alternative routes for the same journey over the course of a week. The busway was also used as a new location for walking and cycling for leisure and provided access to greenspaces which were previously inaccessible.

However, the previous chapter focused on spatial mobility patterns and did not provide any information on changes in levels of physical activity. It is therefore unknown whether new walking or cycling taking place on the busway results in increases in total physical activity, whether it substitutes another type of activity, or whether it is spatially displacing walking or cycling from elsewhere. Investigating the potential spatial displacement of activity through the use of the busway helps to strengthen the basis for causal inference by providing more weight to the increases in physical activity that have been observed in previous evaluation papers [251]–[253].

6.1.1 Chapter overview

This chapter explores the potential of different geospatial analysis methods for visualising changes in spaces used for physical activity, and how physical activity accrued on the busway contributes to overall changes in levels of physical activity.

6.1.2 Background: Geospatial analysis of physical activity

Geospatial analysis methods help to improve understanding of the geographic patterns of specified outcomes or exposures and have been widely applied in environmental epidemiology. A suite of methods is available to visualise spatial and spatiotemporal data. For example, hot spot and spatial cluster analysis has been used to identify areas with a high concentration of outcomes, including the prevalence of chronic diseases, such as breast cancer [263], [264], and communicable disease outbreaks, such as malaria and tuberculosis [265]–[267]. When presented alongside information about the local environment, it is possible to explore outcomes in the context of risk factors and characteristics of the environment pertaining to specific locations [268] and, as demonstrated by John Snow's antecedent study of the Broad Street pump, provide insight into where to direct interventions geographically.

In the field of physical activity, the integration of spatial and temporal data has long been recognised as important for representing and understanding travel and activity patterns [269], [270]. Technological developments have led to the increasing integration of GPS, accelerometry, and GIS to investigate relationships between the environment, space, and activity-related behaviours [69], [155], [271]. However, studies that combine these data typically employ rudimentary spatial techniques. A recent review showed that the primary application has been to quantify physical activity by the domain or environment in which it occurs [69] without mapping the range of spaces visited and where time is spent active. This approach is useful for describing the types of environments used for physical activity but provides no information on the spatial distribution of activity, the accessibility of spaces, or the compensation of time spent active in some locations relative to others. Similarly, in the review in Chapter 3, studies which derived walking or MVPA-specific activity spaces for individuals typically quantified and investigated environmental characteristics experienced during the specified activity, without identifying key locations in which physical activity occurs. Intuitively, when modifying the environment it is useful to understand not only which types of environments are conducive to activity, but where changes should be made and how the spatial patterning of activity is affected following such changes.

There is some evidence of physical activity studies using geospatial analysis methods to identify clusters of physical activity at the population level. Some use hot spot analysis to locate popular walking routes or neighbourhoods where the prevalence of physical activity is high [272]–[274] or kernel density estimation (KDE) to identify potential opportunities for increasing physical activity in particular populations [275], [276]. These provide an important pre-cursor for generating evidence of causal relationships between the built environment and physical activity whilst providing examples of how to apply spatial analysis methods. Intervention studies which map data at multiple time points build on this by analysing whether a particular new space is used and how this affects the location and levels of overall physical activity [277], [278]. However, these types of studies are few and the application of techniques to geovisualise these data and perform analysis is in its relative infancy.

To date there has been heterogeneity in the geospatial analysis methods that have been employed to locate physical activity and the rationale for the methods chosen is often unclear. Despite the importance of temporal data in understanding activity patterns being well known [269], [270], [279], there is a dearth of longitudinal study designs and integration of temporal information [69], [142], [155]. Some studies have used 3D imagery to illustrate spatial concentrations of travel modes by times of day [280], [281] but the presentation of this data must be carefully considered to allow for meaningful comparisons. Although few studies measure and assess longitudinal changes in locations of physical activity, lessons can be learnt from fields such as ecology where gridded maps have been overlaid and subtracted from one another to measure absolute changes in type and coverage of land use over time [282], [283]. These methods are useful for performing calculations on location-specific data and measuring rates of change in locations of interest.

There is therefore a need to test and refine geospatial analysis methods used to locate physical activity and understand their technical and conceptual strengths in relation to different research questions and causality [67], [68], [152].

6.1.3 Aims and scope

Building on methods used to assess changes in spatial patterns of movements and use of space after the opening of the Cambridgeshire Guided Busway presented in Chapter 5, this chapter aims to investigate how locations of physical activity changed. Due to the small study sample size, this study does not aim to be definitive. Rather, the feasibility of different spatial analysis methods are reviewed by piloting them on the available data with the view of being able to roll out the most applicable in larger datasets or cohorts where GPS data is being collected alongside behavioural and health outcomes. The chapter also seeks to identify and quantify physical activity which occurred on the busway and explore whether it could have contributed to an increase in overall levels of physical activity or displaced physical activity from elsewhere.

6.2 Methods

6.2.1 Data

Data were used from phases 2 and 4 of the Commuting and Health in Cambridge Study, as detailed in Chapter 4, Section 4.2.1. In brief, participants in the cohort were invited to take part in objective activity monitoring by wearing physical activity and GPS devices. Physical activity devices and GPS receivers were issued at a visit made by participants to the research institute. Participants were instructed to wear the devices simultaneously for seven consecutive days [89]. The physical activity devices were worn constantly and for all activities as they were waterproof and attached directly to the skin. In contrast, the GPS devices were worn on an elastic waist belt.

GPS devices (Qstarz [BT-1000X]) were set to collect data at 5 second epochs at phase 2 and 10 second epochs at phase 4 of the study. This change was made in order to preserve battery life and reduce the burden for participants through device charging. Phase 2 was the first phase of data collection where GPS data were available and phase 4 was the first complete phase of data collection when the busway was formally opened. The methods used to clean the GPS data were described in detail in Chapter 4. These were used as the basis for locating physical activity.

Physical activity was assessed using combined heart rate and movement sensors (ActiHeart, CamNtech, Papworth, UK). The ActiHeart records a measure of physical activity energy expenditure and has been shown to be a valid and reliable tool, particularly for measuring activities such as cycling, which are often not detected by wrist-worn accelerometers [284]. ActiHeart devices were set to record activity at 60 second epochs at both phases 2 and 4 of the study. Calibration based on age, sex, and sleeping heart rate has previously been performed on a sample of 51 adults in the Cambridge area [285] and therefore the group calibration equation was applied and no individual calibration was used.

For inclusion in this analysis, participants were required to have repeat ActiHeart and GPS measures at both phases of the study (pre and post-intervention) which created a potential sample of 72. As the purpose of the study was about assessing feasibility of methods and exploratory analysis, participants who moved home or work were included in the final sample to maximise the sample size.

6.2.2 Data pre-processing

6.2.2.1 Matching ActiHeart and GPS data

ActiHeart data points were matched to the closest recorded GPS location based on date and timestamp through a combination of Activity Data Analyser (ADA) software developed at the University of East Anglia and manual checks in STATA.

ActiHeart timestamp	GPS timestamp	Time difference between devices [seconds]	GPS Index
01/06/2010 11:12:00	01/06/2010 11:14:43	163	1
01/06/2010 11:13:00	01/06/2010 11:14:43	103	1
01/06/2010 11:14:40	01/06/2010 11:14:43	3	1
01/06/2010 11:14:45	01/06/2010 11:14:48	3	2
01/06/2010 11:14:50	01/06/2010 11:14:53	3	3
01/06/2010 11:14:55	01/06/2010 11:14:58	3	4
01/06/2010 11:15:00	01/06/2010 11:14:58	2	4

Dummy times generated between original ActiHeart timestamps (shown in **bold**) to match GPS time intervals

Data deleted due to time differences greater than 60 seconds

Duplicate data point deleted

Figure 6.1: Example of time-matched ActiHeart and GPS data

Periods of non-wear recorded by the ActiHeart were dropped prior to data-matching. As the epochs used for data collection differed between ActiHeart and GPS and between phases for the GPS data, data collection epochs were input as parameters into the ADA interface. Dummy times between each ActiHeart timestamp were generated by ADA to match the shorter time intervals in the GPS data so no GPS data were lost (Figure 6.1). The physical activity estimate created for each newly imputed time was the same as that recorded at the previous ActiHeart timestamp.

Data were cleaned in order to ensure good quality matches were achieved. Data points were removed if time differences between the matched ActiHeart and GPS data were greater than 60 seconds or if duplicate GPS points existed. In the case of the latter, the ActiHeart value from the closest matched time was retained (Figure 6.1). Eastings and northings associated with each GPS data were retained in the dataset and used to plot points in ArcGIS.

Using adapted versions of the Python functions in Appendix C, total wear time of both devices was calculated for each day, and each participant. In line with inclusion criteria used in Chapters 4 and 5, days with fewer than 8 hours of wear time were excluded from the analysis.

Participants were retained based on the same total week criteria used in Chapter 4 of at least 3 weekdays and 1 weekend day (Table 4.4). This approach was chosen as I was interested in the broader patterns of physical activity over weekdays and weekends and the capabilities of different geospatial methods, not in describing detailed within-individual changes in temporal patterns.

6.2.2.2 *Physical activity measures*

A metabolic equivalent (MET) is the ratio of energy expended during activity to the rate of energy expended at rest [286], and was recorded by ActiHeart devices. One MET is a measure of energy expenditure at rest. Activity at 2 METs therefore requires twice the energy used when at rest [287].

MVPA has been shown to confer health benefits in line with recommended guidelines [4] and has been estimated using a threshold of 3 METs in previous studies of adults [251], [288], including those of the same sample. Based on the value of METs recorded by the ActiHeart and assigned to each GPS data point, a binary variable was created to indicate whether the participant was in MVPA (above 3 METs) at each data point or not. The amount of time spent in each episode of MVPA (time spent continuously above 3 METs) was also calculated using adapted versions of the CreateIndex and SegmentTotal functions I wrote in Python to clean GPS data in Appendix C. These two variables were used to create population maps of MVPA.

To investigate changes in levels of physical activity between study phases, I first created a relative measure of time spent in MVPA by calculating the percentage of total device wear time spent in MVPA for each participant at phase 2 and phase 4 separately. The relative measure of MVPA at phase 2 was then subtracted from the relative measure at phase 4 with a positive outcome indicating an increase time spent active and a negative value indicating a reduction in time spent active. Based on the distribution of the data and the logic used in Chapter 5 to assess changes in activity spaces (whereby some change is almost inevitable because it is unlikely that a participant would record exactly the same level of physical activity over a 7 day period, 2 years apart), tertiles were used to categorise time spent in MVPA into categories of increase, decrease, and no substantial change.

6.2.2.3 *Quantifying physical activity on busway*

Using a 20 m buffer of the busway in ArcGIS and identifying all GPS points located within it, the total time spent in MVPA on the busway was estimated. This was translated into a relative measure based on device wear time for each participant. As with total MVPA, measures at phase 2 were subtracted from those at phase 4. Due to the large number of participants who did not change their time spent in MVPA on the busway, tertiles were not appropriate. Instead, change was categorised based on relative measures where; (i) those with negative change values decreased MVPA on the busway, (ii) those with change values of zero stayed the same, and (iii) those with positive change values increased time in MVPA.

6.2.3 Geovisualisation: population-level maps

Due to the small sample size, it was not appropriate to investigate whether exposure to or use of the busway was associated with within-individual changes in physical activity. Instead, valid data points from each participant were merged into a single dataset for the sample. Drawing on previous experience of using the ArcGIS suite to visualise spatial data, I generated four types of population maps to test the suitability of different methods and to visualise shifts in the spatial patterns of physical activity. The methods chosen reflected methods previously used in studies which have aimed to locate physical activity and health outcomes [272]–[276]. Rather than reviewing all potential geovisualiation methods, I chose methods that ranged in approach and processing power in order to understand the types of processes that may be feasible for identifying locations used for physical activity by a sample.

All data were clipped to the study area (30 km of Cambridge city centre) for data manageability and to capture likely commuting behaviours and use of the busway.

6.2.3.1 Point-to-raster

Raster data is a type of gridded data where each cell is assigned a value [289] and can be created in a number of ways. In this case, each cell represented a measure of physical activity, using either a binary or continuous value. All GPS point data were aggregated to a raster format using the 'Point to Raster' tool in ArcGIS. Two raster maps were created, as visualised in Figure 6.2, using this relatively simplistic approach to create a direct representation of point values using limited processing power. The first represented locations where any MVPA had or had not occurred based on the binary MVPA variable of all points located within a cell. The second provided graded maps of mean minutes spent in MVPA episodes.

Initially, a large cell size of 500 m was tested, based on the maximum distance that could be travelled between GPS points, as described in Chapter 4, Section 4.2.4.3. The output was useful for testing the methods, however, it appeared too coarse to detect walking and cycling behaviours in relation to the busway. A cell size of 10 m, based on distances travelled whilst walking or cycling, was also tested but deemed to be too detailed and fine grained and took a large amount of processing power. After consideration, a cell size of 50 m was used which achieved a balance between detail, processing power, and suitability to answer the research question.

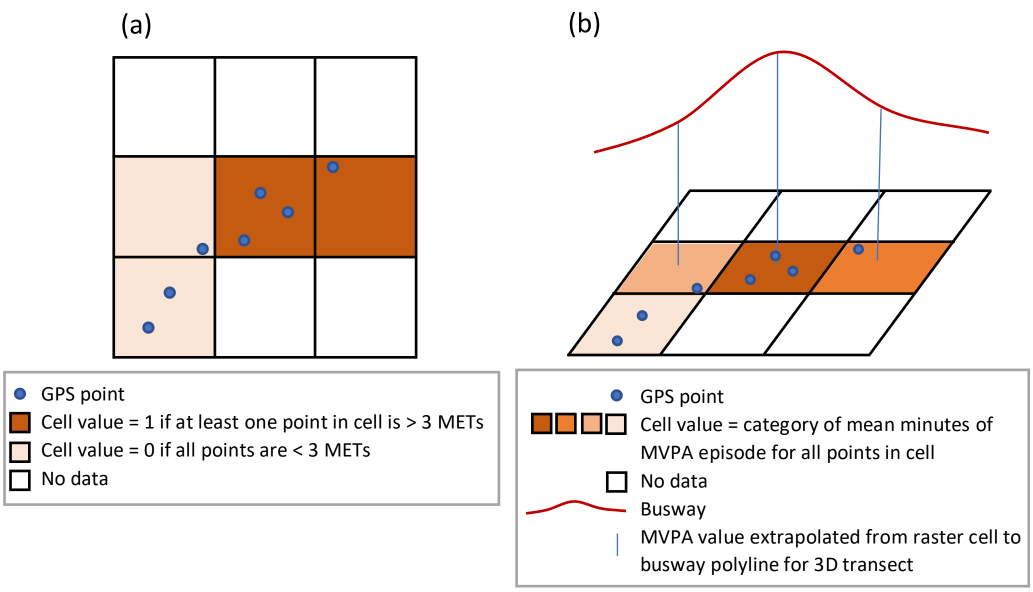


Figure 6.2: Binary (a) and gradient (b) point-to-raster map method

As mean minutes spent in MVPA episodes comprised both a spatial and temporal element, I also generated a 3D transect by interpolating surface values of time spent in MVPA data to polylines of the route network.

6.2.3.2 Spatial autocorrelation

Spatial autocorrelation techniques can be used to reveal groupings of points based on their value and location in relation to neighbouring points. Two approaches to identify concentrations of values were tested: hot spot analysis and cluster analysis. Both are complementary in their capabilities, however, cluster analysis identifies outlying points where values of surrounding points differ. For the purposes of this analysis, I therefore used hot spot analysis to measure concentrations of MVPA points, and spatial clustering to measure concentrations of long episodes of MVPA, and locations of outlying episodes.

<u>Hot spot analysis</u>

Maps showing the statistically significant locations of physical activity were derived based on the binary MVPA variable. To identify hot spots of MVPA, the Getis-Ord Gi* statistic score was calculated for each data point using the Hot Spot Analysis tool in the ArcGIS suite. The tool works by evaluating each point in the context of its neighbouring points and identifies where concentrations of points with high values exist. As erroneous spatial outliers have been removed in the GPS data cleaning process (Chapter 4), all points were considered of interest and so the default fixed distance used to assess neighbouring points was used for both study phases. Points within the fixed distance were weighted equally and those outside had no influence on the calculations. The Gi* statistic returned for each feature in the dataset is a z-score. All points with the highest significant positive z-scores, indicating the most concentrated groupings of MVPA (hot spots), were aggregated into polygons.

Spatial cluster and outlier analysis

Using the cluster and outlier tool in ArcGIS, I calculated the local Moran's I value for each GPS point, based on the continuous value of minutes spent in an episode of MVPA. This statistic identifies locations where high and low values are clustered and anomalous areas where outlier high values are surrounded by primarily low values. As with the hot spot analysis, the fixed distance band was used to assess neighbouring points. Significant points with high values were aggregated into polygons to represent locations where the longest episodes of MVPA occur. I also created polygons of outlying points to represent locations where anomalous long episodes have been recorded amongst areas typical of little or no MVPA.

6.2.3.3 Kernel density estimation (KDE)

Kernel estimation was used to create density surfaces of physical activity locations for each study phase. Kernel estimation calculates the density of points by first fitting a smoothed surface (kernel surface) over each GPS point (Figure 6.3). The number of kernel surfaces that overlap the centre of each cell are then summed to create a value of density (kernel density estimation) in each cell of the map.

Density maps were created for the binary measure of any MVPA and the continuous measure of time spent in MVPA episodes. For the binary measure, the kernel surface for each GPS point where MVPA was recorded equated to 1, as illustrated in Figure 6.3. For example, if the kernel surface of 3 GPS points overlapped the centre of a cell, the kernel estimation for that cell would be 3. The kernel density map of MVPA episodes is weighted by the minutes spent in each episode. For example, if 2 GPS points were located in a cell, one with a value of 3 minutes and another with a value of 8 minutes, the kernel surface of each point would represent the number of minutes and the kernel density estimation value would sum these, equating to 11. This surface therefore represents the locations where the longest episodes of MVPA most frequently occur.

A cell size of 50 m was selected based on the processing of raster maps.

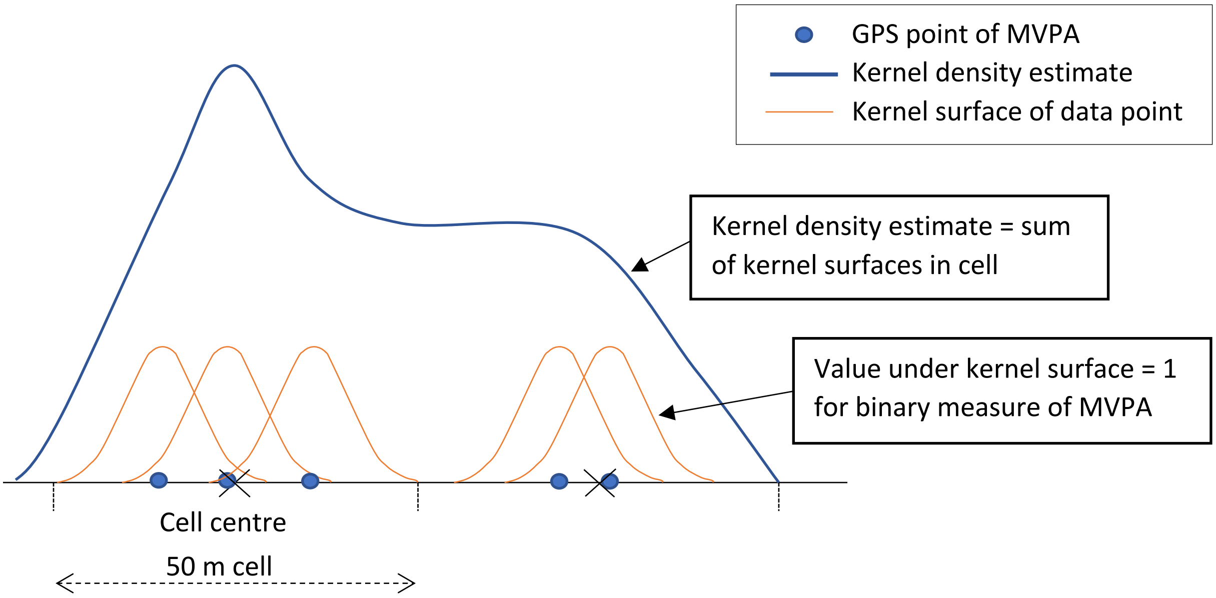


Figure 6.3: Kernel density estimation method

6.2.4 Applicability assessment

Data are presented and interpreted to highlight potential changes in locations of physical activity shown by each output map. For each method, the technical challenges, practicality, and ability to answer my research questions were reviewed narratively.

6.2.5 Descriptive analysis

To investigate how physical activity on the busway contributed to overall levels of physical activity, descriptive analysis was performed using the relative measures of physical activity outlined in Section 6.2.2. The analysis was designed to complement the geovisual outputs by quantifying within-individual changes in the location and level of physical activity.

Due to the small sample size, regression analyses were not considered appropriate for this study.

6.3 Results

6.3.1 Mapping results

Valid matched data were available for 53 participants and used to derive population-level maps of physical activity in the study area. Each type of map is represented in Figures 6.4 to 6.8. Maps are annotated and key sections highlighted to demonstrate possible key findings that may be interpreted, as well as the capabilities of each geospatial analysis method.

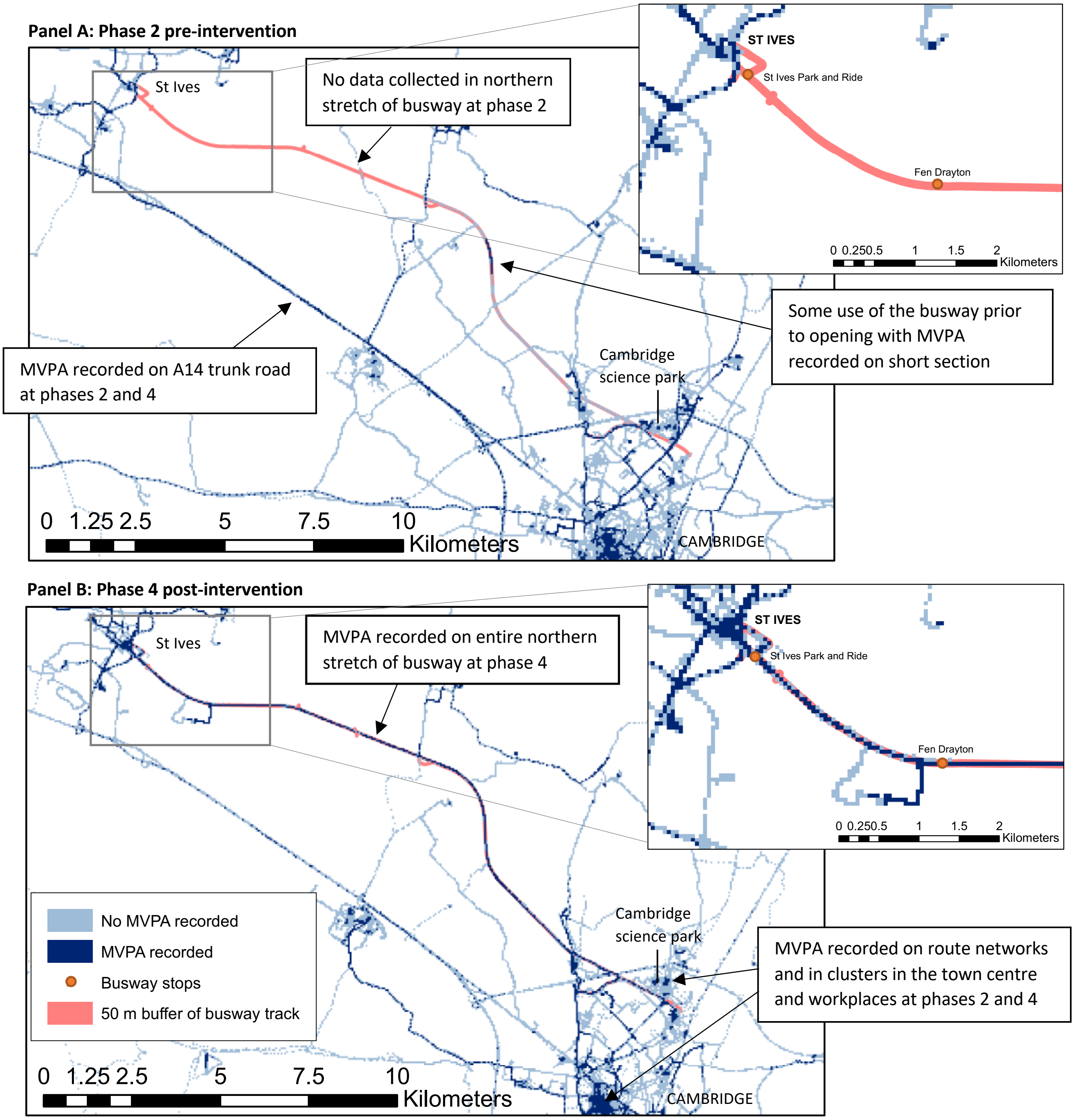


Figure 6.4: Binary point-to-raster maps of MVPA at phases 2 and 4 of the Commuting and Health in Cambridge Study

The binary point-to-raster maps in Figure 6.4 provide a simple measure of locations in which MVPA occurred. The values represented are not weighted by neighbouring points and do not account for the density of points measured in each cell, which may indicate the frequency of use or time spent active. Comparing the outputs for phases 2 and 4 of the study, a clear uptake of MVPA is measured on the busway and on attributing roads to the northern section of the busway. MVPA is typically recorded on route networks, however, some routes do not record any MVPA at all suggesting they favour more passive modes of travel. Some MVPA is recorded in small clusters which may be representative of residential or workplace locations, as shown near the science park. A large cluster of MVPA is shown in Cambridge at both phases which is likely indicative of the large number of people walking or cycling through the city centre. Interestingly, some MVPA is recorded on the A14 trunk road at both phases where the most suitable modes of travel are passive. This may be due to the use of a motorbike or scooter which requires some exertion and involves movements and vibrations that might be captured by the ActiHeart.

Graded raster maps build on the binary maps by representing where MVPA occurred and the mean time people in that location were continuously active for. The average time spent in episodes of MVPA was less than 5 minutes in most locations where MVPA was recorded (not shown). In Figure 6.5, I therefore focus on new use of the busway at phase 4 (Panel A). Episodes of more than 10 minutes were recorded in the most northerly section and in nature reserves accessible from the busway, as identified from the analysis of individual profiles in Chapter 5. As with the binary maps, due to the point-to-raster approach used to create these visual outputs, it is unknown whether these observations of MVPA are representative of a number of participants or driven by data from a single participant.

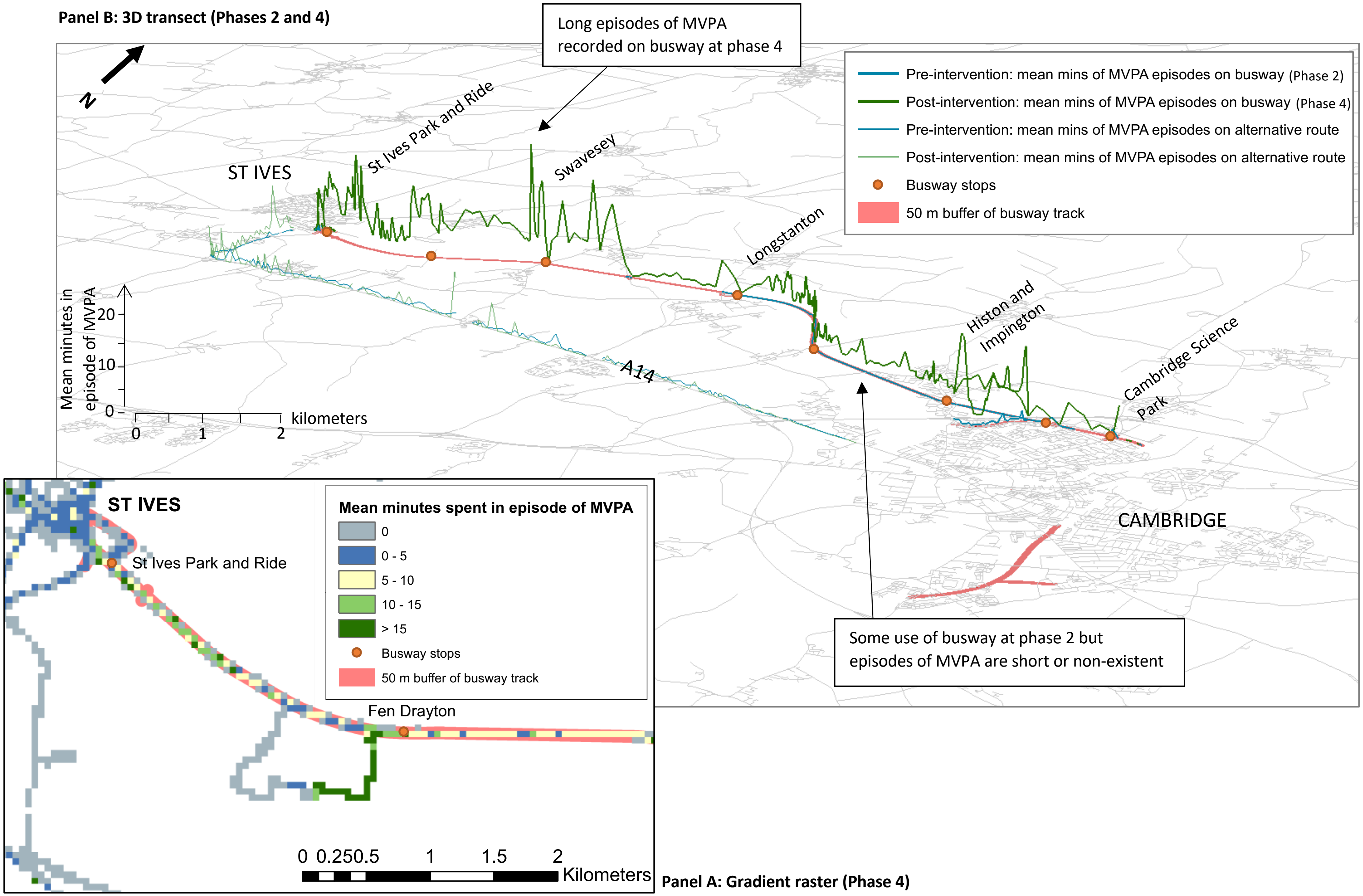


Figure 6.5: Point-to-raster gradient map and 3D transect of mean minutes spent in episodes of MVPA along the busway

Scales added in Panel B approximate for visual reference. Contains OS data © Crown copyright and database right 2019

Values from the graded point-to-raster maps have been extruded along the busway in the 3D transect (Figure 6.5, Panel B) to compare episodes of MVPA at phases 2 and 4 (shown in blue and green respectively). The transect shows some data were collected along the busway at phase 2, but little or no MVPA is shown. In contrast, data at phase 4 clearly show episodes of MVPA recorded along the whole length of the busway which are consistently longer than those at phase 2, as would be expected post-intervention. The longest episodes are recorded in the most northerly section which may be indicative of longer journeys made for most travellers using this section of the busway. For example, a participant cycling continuously from St Ives to Cambridge will record a longer episode of MVPA than a participant cycling from Longstanton, whose recording will lower the average time spent in an episode of MVPA. Some of the bus stops appear to coincide with shorter episodes of MVPA which may be due to participants exiting or entering the busway at these points, capturing the start or end of an episode. An alternative route along the A14 from St Ives to Cambridge is also illustrated for comparison, showing short or no episodes of MVPA. This suggests that the measures of binary MVPA shown along the A14 in Figure 6.4 were anomalous. Some peaks indicating longer episodes do appear close to St Ives and at cross-roads, possibly capturing MVPA on routes that dissect the A14.

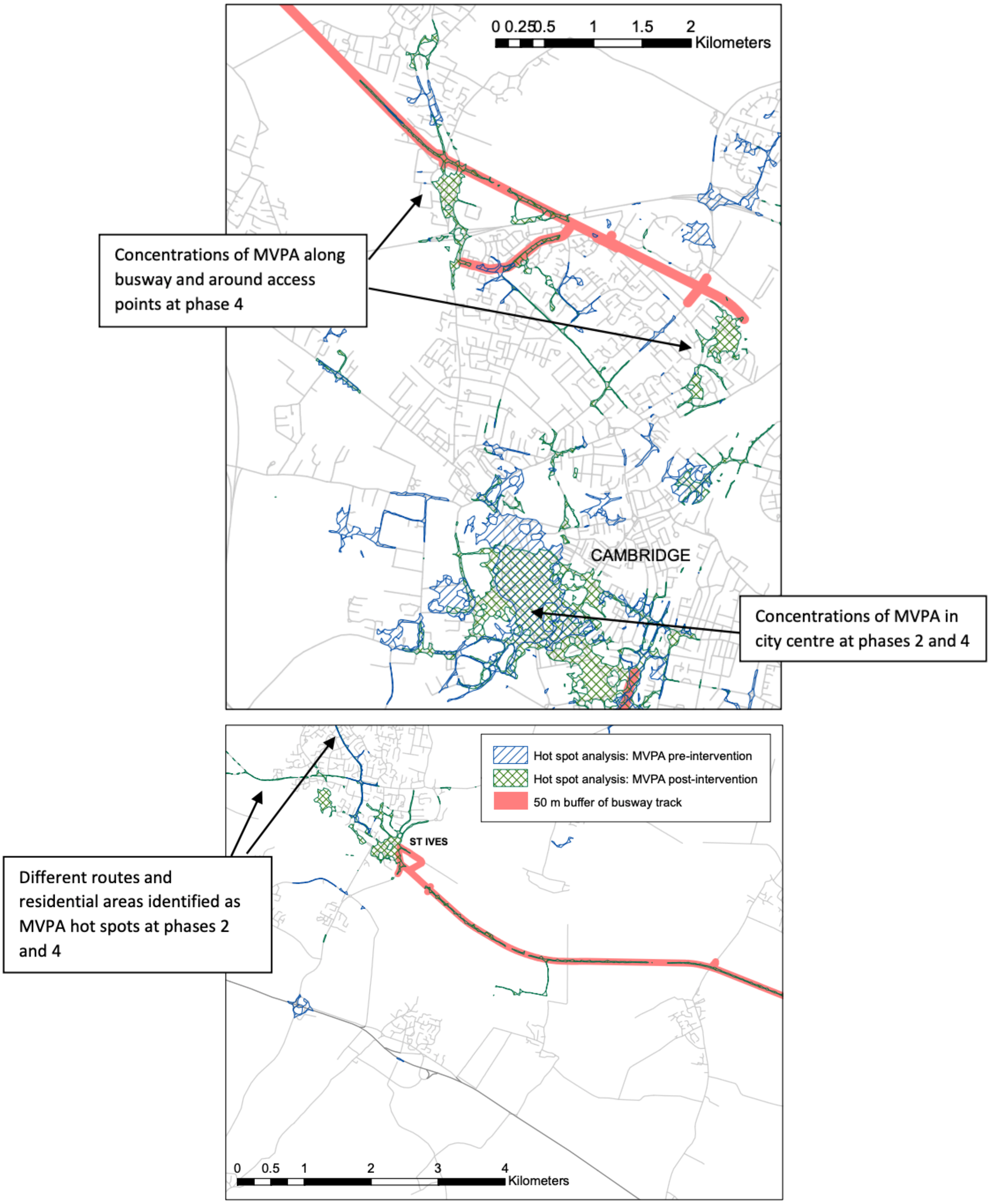


Figure 6.6: Hot spot analysis maps of MVPA at phases 2 and 4 of the Commuting and Health in Cambridge Study
Contains OS data © Crown copyright and database right 2019

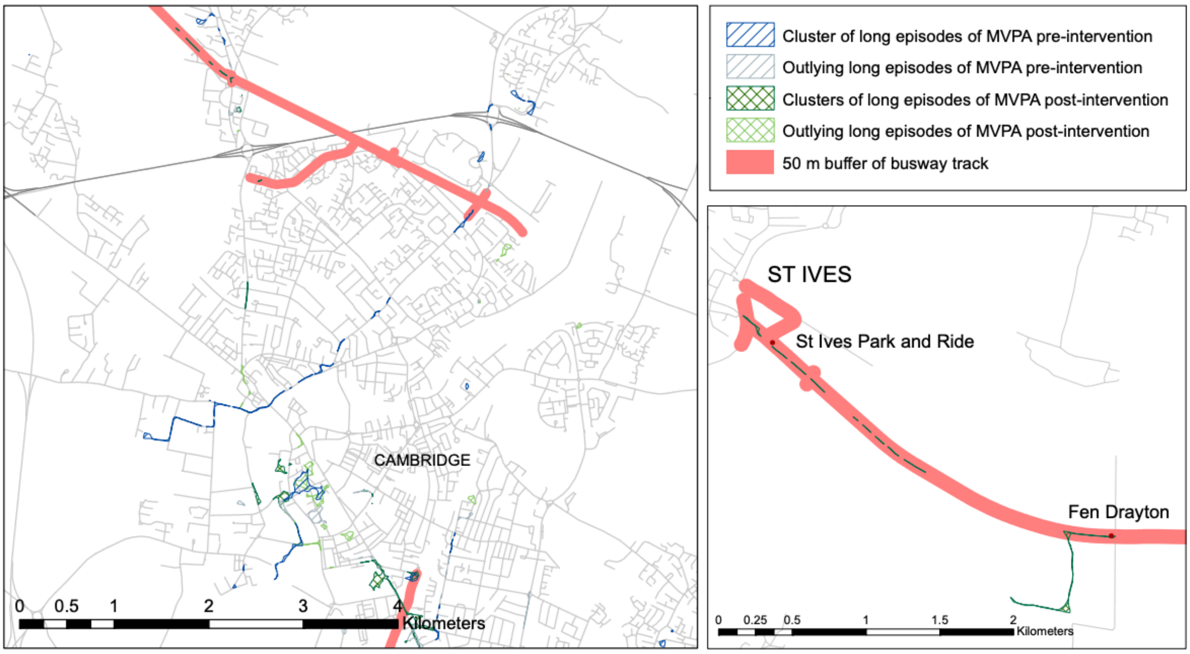


Figure 6.7: Spatial cluster maps of mean minutes spent in episodes of MVPA
Contains OS data © Crown copyright and database right 2019

In contrast to the point-to-raster maps which do not account for neighbouring points, the spatial autocorrelation maps in Figure 6.6. and Figure 6.7 show the statistically significant locations where MVPA occurred at phase 2 and phase 4 of the study.

In Figure 6.6, Cambridge city centre is identified as a significant hot spot for MVPA at both phase 2 (shown in blue) and phase 4 (shown in green). Some network routes and residential areas are highlighted as hot spots, although these locations are rarely consistent between phases. The busway is shown clearly as a new significant location for MVPA at phase 4, as are areas around access points in St Ives and Cambridge. Unlike the data presented on the point-to-raster maps, concentrations of MVPA are only shown on some sections of the busway.

The spatial cluster maps in Figure 6.7 complement the hot spot maps by illustrating significant locations where long episodes of MVPA occur. In contrast to the hot spot maps, the area in the city centre of Cambridge identified for long episodes of MVPA is much smaller with outlying clusters shown at phase 4 (shown in light green), suggesting the long episodes occur in a location where no MVPA or short episodes are most common. As with the hot spot analysis, some clusters are shown along network routes although these are much fewer and there doesn't tend to be any long episodes in residential areas. Similarly, while some clusters are shown on the busway at phase 4, highlighting the new use of space for long episodes of MVPA, these are presented for much smaller sections. The clusters of long episodes shown to the north of the busway and in the nature reserves near Fen Drayton correspond with those shown in the graded raster (Figure 6.5, Panel A), despite accounting for neighbouring points.

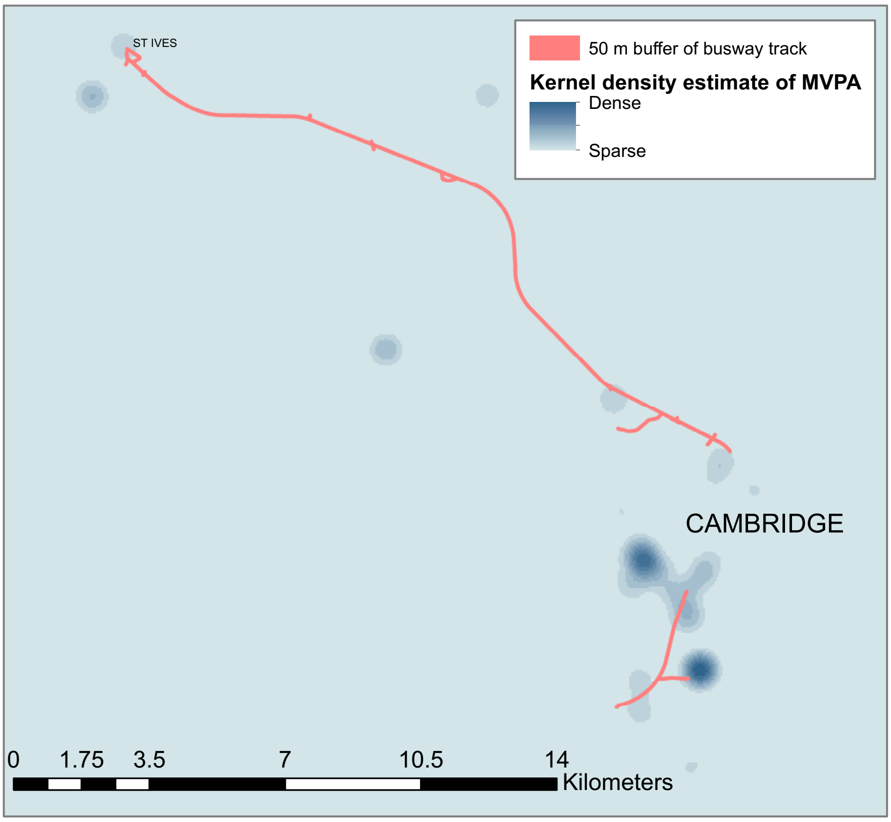


Figure 6.8: Example output of kernel density map of MVPA at phase 4 of the Commuting and Health in Cambridge study (post-intervention)

Kernel density maps were created to identify locations where the density of MVPA was greatest. However, the process was not suitable for identifying use of key infrastructure, such as the busway, from population level data. Figure 6.8 provides an example output from the kernel density estimation, favouring locations such as workplaces where participants spend the most amount of time (in the city centre and at the Cambridge biomedical campus). MVPA accrued when actively travelling is unlikely to be densely clustered due to the distance travelled and the relatively short time spent en route compared to that in a workplace. The outputs produced for both phases and outcomes were therefore similar and provided little information on how the location of physical activity changed over time.

6.3.2 Strengths and limitations of geospatial methods

The key strengths and limitations of each geospatial analysis methods used are outlined in Table 6.1.

Table 6.1: Strengths and limitations of geospatial methods

Geospatial method	Interpretation	Strengths	Limitations
Point-to-raster maps			
Binary	Locations of absolute measures of any MVPA	Easy to determine whether specific location has been used for physical activity or not Simple to calculate differences between data at different study phases Low processing power required	No weighting by number of people, frequency of visits, or time spent in locations Statistically non-significant locations - does not account for neighbouring points and cells and anomalous values are given equal weight Difficult to identify changes in spatial patterns
Gradient	Average time spent in episode of MVPA across study area	Easy to identify locations where episodes of MVPA are longest Temporal trends easily visualised in 3D transect Low processing power required	Statistically non-significant locations - does not account for neighbouring points and cells and anomalous values are given equal weight Difficult to identify changes in spatial patterns
Spatial autocorrelation			
Hot spot analysis	Significant locations where MVPA occurs	Values of neighbouring points assessed Significant points can be grouped into polygon or raster maps Easy to identify and compare locations where MVPA is concentrated	Additional steps required to create raster or polygon maps following calculation of clusters High processing power required
Cluster and outlier analysis	Significant locations where MVPA episodes are longest and where outlying long episodes exist	Values of neighbouring points assessed Significant points can be grouped into polygon or raster maps Easy to identify locations where episodes of MVPA are longest	Additional steps required to create raster or polygon maps following calculation of clusters High processing power required
Kernel density maps	Density of points where MVPA is measured, weighted by length of episode	Weights locations by time spent in them and length of MVPA episodes	Concentrated locations where most time is spent (such as workplaces) always favoured Locations of MVPA where points are less dense (such as route networks) not identified Less easy to interpret use of specific locations High processing power required

A strength of using point-to-raster maps is the low processing power and time required to produce outputs. Accordingly, it is possible to test different cell sizes and outcome variables to map, such as most frequent binary value in a cell or mean METs. However, as the point-to-raster maps represent absolute measures, the locations of MVPA presented are not necessarily significantly different from surrounding points. For example, the spatial representation of population-level MVPA may be driven by a single participant which makes it difficult to observe generalised trends in the data. A key limitation of the point-to-raster maps which were created therefore lies in their interpretation. As neighbouring points are not considered when creating a cell value, anomalous values, such as those records of MVPA values shown on the A14 in Figure 6.4, receive equal weight to genuine clusters of MVPA. However, raster maps in general provide a useful format for performing calculations as a range of mathematical functions can be quickly and easily performed on overlaying cells from different layers of raster data. By creating graded raster maps from continuous data, it is also possible to extrude values and generate 3D transects. Transects provide a useful visualisation for comparing data from different time points or populations, particularly along route networks, and in the case of this analysis allowed for temporal data to be incorporated.

In contrast to point-to-raster maps, the spatial autocorrelation maps highlight significant locations where MVPA and the longest episodes of MVPA take place, accounting for data from neighbouring GPS points. Anomalous values are therefore removed from the outputs. A strength of the cluster and outlier analysis is the ability to detect outlying long episodes of MVPA in locations where sedentary behaviour or short periods of MVPA is most common. With regards to the busway, this helps to shed light on whether active or passive travel is most prevalent. Both hot spot and cluster analysis maps showed concentrations of MVPA in very specific locations, such as routes or residential areas, highlighting the sensitivity of the calculation process and difficulty in making comparisons between two groups of data. However, the outputs were helpful for identifying new locations of MVPA and although I chose to aggregate significant points into polygons to represent clusters, smoothed polygons or raster maps may be produced to present that data. The benefits of using raster maps to perform calculations and extrude 3D profiles may then also be afforded. The primary issue of creating cluster maps is the processing power required to firstly calculate scores for each point and secondly produce a meaningful visual representation of points. This poses issues for working with large datasets but may be applicable when working with data at the individual level.

Kernel density maps allowed for locations to be weighted by time spent active in them. However, contained locations where most time is spent by high volumes of people skew the data and limit the ability to detect the use of infrastructure for physical activity. It might be applicable for use in datasets where there is more variation in the locations where people spend time.

6.3.3 Descriptive results: relative changes in MVPA on the busway

Table 6.2 shows the proportion of participants by change in the amount of MVPA undertaken on the busway between phases 2 and 4. Participants who spent a smaller proportion of their device wear time active on the busway at phase 4 than phase 2 were typically urban dwellers and former users of the busway. The largest number of participants measured no change the proportion of time spent active on the busway between phases. This is because 75% of this group recorded no use of the busway at either phase.

Those that saw a relative increase in MVPA on the busway owned at least one car and were largely new users. The majority of this group also recorded a relative increase in overall levels of physical activity. This suggests that the busway may provide a space for new and additional MVPA to occur and the spatial displacement of MVPA from a previous location to the busway may be minimal. There was also some indication that those who decreased their relative amount of MVPA on the busway (largely former users) tended to decrease their overall levels of physical activity.

Table 6.2: Changes in MVPA on busway

	Change in MVPA on busway		
	Decrease (n = 7)	**No change (n = 34)**	**Increase (n = 12)**
	n (%)	n (%)	n (%)
Sex			
Male	2 (29)	16 (47)	7 (58)
Female	5 (71)	18 (53)	5 (42)
Age [Years]			
<40	3 (43)	8 (24)	3 (25)
40-50	1 (14)	11 (32)	8 (67)
>50	3 (43)	15 (44)	1 (8)
Urbanicity			
Urban	6 (86)	17 (50)	7 (58)
Rural	1 (14)	17 (50)	5 (42)
Car ownership			
None	1 (14)	2 (6)	0
One	5 (71)	15 (44)	6 (50)
More than one	1 (14)	17 (50)	6 (50)
GPS-measured use of busway			
None	0	26 (76)	0
Former	5 (71)	1 (3)	0
Continued	2 (29)	3 (9)	3 (25)
New	0	4 (12)	9 (75)
Change in overall MVPA			
Decrease	3 (43)	13 (38)	2 (17)
No change	2 (29)	15 (44)	1 (8)
Increase	2 (29)	6 (18)	9 (75)

6.4 Discussion

6.4.1 Principal findings

This study explored three different geospatial analysis methods and assessed their capacity for identifying locations of physical activity, and how they change in relation to a built-environment intervention.

The geospatial maps indicated the possible uptake of MVPA along the busway. The 3D transect and cluster and outlier analyses appeared to illustrate longer episodes of MVPA accrued in the most northerly section where potentially longer active trips may have been made. The city centre of Cambridge was shown to be a consistently concentrated area of physical activity, but in the same location, sedentary and short episodes of MVPA appeared most common as suggested by the cluster and outlier analyses. Clusters of MVPA were recorded along routes (particularly those connecting to the busway), reflecting possible travel behaviours, and in nature reserves accessible from the busway, as shown in Chapter 5. The maps suggest that use of the busway for physical activity may be important for increasing overall levels of physical activity, providing a new space for additional activity to occur.

Technical and conceptual limitations varied depending on the analysis method piloted. Point-to-raster maps provided a quick way of visualising physical activity outcomes which may include temporal elements but do not account for the frequency of visits or anomalous points. Spatial autocorrelation methods indicated significant locations of physical activity and were useful for visualising spatial changes, although significant processing power is required to compute outcomes. Kernel density estimation has the potential to weight locations by time spent in them but appeared less appropriate for identifying physical activity within a defined area where relatively little time of the day is spent, such as the busway.

6.4.2 Applicability of geospatial methods and recommendations

Although maps representing behaviour at the population level cannot be used to identify statistically significant relationships, the geovisualisation of data can bring meaning and local relevance to quantitative analysis. They complement methods used in studies that quantify physical activity by environments and domains by illustrating potential trends in physical activity over time and space.

Point-to-raster maps provide a means for initially exploring data to identify whether locations of interest were being used. The low processing requirements means that the use of raster data is flexible and can form the basis for subsequent calculations or 3D visualisations that enable comparisons. However, the outputs generated were based on absolute values which must be interpreted with caution. Spatial autocorrelation maps provide a more discriminating assessment of locations used for physical activity, although their application in the literature has previously been simplistic with, for example, ecological studies identifying the prevalent locations of physical activity [272]–[274]. There is opportunity to apply these approaches to more complex study designs to assess how use of spaces change over time. Given the low processing power of point-to-raster maps, it is also feasible to combine multiple geospatial outputs to add greater depth and incorporate temporal information into findings. For example, Miller and colleagues overlaid spatial clusters of high activity over raster maps of transport related physical activity and assessed how the distribution of each changed following an intervention [277]. This provided context for findings and a clear illustration of change over time.

A key limitation to the use of spatial autocorrelation maps using tools within ArcGIS is the significant processing time required which makes its application less suitable for larger datasets. However, creating spatial clusters of physical activity at the individual level may be feasible as the density and number of neighbouring points in an individual's GPS trace is more manageable than those in a population dataset. This would allow for within-individual spatial changes to be investigated. Some studies have attempted to assess the contribution of transport related physical activity to overall physical activity at the individual level [253], [277], [278], as tentatively explored in this study. However, none have incorporated geospatial methods to measure potential spatial displacement of physical activity due to a specific use of space, highlighting a key area for future research.

Kernel density maps appear less appropriate for identifying physical activity in relation to specific locations and interventions. However, one strength of this method is the ability to identify areas where most time is spent, rather than where specific outcomes occur. The method can therefore be applied to temporally weight measures of environmental exposure in line with the application of the activity space concept, as has been employed in studies of food environments [290]. Building on the method used by Miller and colleagues to combine complementary geospatial methods, future studies could incorporate time-weighted data through the use of kernel density information.

6.4.3 Strengths and limitations

This study was novel in its attempt to assess a range of geospatial methods in terms of their application for measuring locations of physical activity, and how these change in response to a built environment intervention. Although further testing on different samples and study areas is required, as high volume locational data becomes increasingly available in physical activity research, this work provides timely insight into potential ways to better understand the use of space for physical activity alongside more traditional epidemiological analyses to ultimately strengthen the basis for causal inference.

A key limitation of the study was the sample size of the available data which meant that population-level maps formed the focus of the study. However, the sample of working adults and availability of data pre- and post-intervention meant that participants with diverse spatial patterns and agency to change their travel behaviour were included. An additional limitation was the use of ArcGIS software for analysis. A range of alternative GIS software are available, including opensource platforms such as QGIS and RSpatial [291]. It also possible to write programs to manipulate and visualise spatial data using Python and geographical coding libraries, such as GeoPandas [292], without the need for a software interface. It is likely that the processing times experienced when generating spatial autocorreletaion and kernel density maps related to the use of ArcGIS and may be mitigated through the use of alternative software. To test alternative approaches would have required a significant amount of learning and time resources which, given the time already directed to develop skill writing Python, were considered to be beyond the scope of this PhD. However, I have identified this as an area of key learning for future development as a researcher in health geography which I intend to employ in future roles and projects.

Although only a descriptive analysis of change in MVPA in relation to the busway was performed, this allowed for potential spatial displacement of physical activity to be explored; a concept that has received little attention before. While the strength of conclusions that could be drawn in this specific case are limited, this study plays an important role in the wider evaluation studies of the busway. Future research may apply similar methods and draw together a wide range of data to provide answers about how interventions are used and why they may (or may not) be effective.

6.5 Conclusions

This study is the first to examine methods for measuring spatial changes in physical activity after the opening of a major new walking and cycling infrastructure project. Findings suggest that the busway may have provided a location for new physical activity and that active use of the busway may contribute to an increase in overall physical activity.

Maps can provide a valuable input alongside traditional epidemiological analyses to help understand use of space and target public health interventions. Future studies could combine geospatial methods with in-depth individual analyses in order to more fully understand environmental influences on physical activity, and its potential compensation or displacement.

Chapter 7

Discussion

7.1 Introduction

This thesis aimed to improve understanding of the relationships between the environment and physical activity by reviewing, developing, and applying different concepts and methods. Associations between a range of environmental characteristics and physical activity outcomes were examined and different conceptualisations of the activity space and their potential for furthering causality and answering scientific questions were presented. Key methods used to delineate the activity space were taken forwards using objective measures of location, and qualitative data were incorporated to examine how use of space changes in the context of a built environment intervention. Furthermore, ways to visualise the spatial distribution of changes in locations of physical activity were tested, and their scalability and replicability in future research evaluated.

7.1.1 Chapter Outline

This chapter initially summarises findings from the previous chapters. The implications of the methodological and scientific contributions made by the thesis are subsequently discussed, considering their impact on the field of physical activity and public health research and policy. The strengths of the work presented are reflected upon before outlining potential directions for future research.

7.2 Summary of principal findings

In Chapter 2, neighbourhood environmental characteristics of varying scales were considered in combination. Characteristics were grouped within five facets (spaces for physical activity, walkability, disturbance, natural environment, and the sociodemographic environment) and their associations with objective ('recorded') and self-reported ('reported') physical activity were investigated. The findings indicated that participants living in areas with higher concentrations of air pollution recorded and reported lower levels of physical activity, while those in rural and more walkable areas had higher levels of both recorded and reported activity. Some associations varied according to the specificity of the outcome, for example, those living in most deprived areas were less likely to record higher levels of MVPA but were more likely to report higher levels of walking. These findings suggest that environmental characteristics have the potential to contribute to different types of physical activity. However, interventions that focus on a single environmental attribute or physical activity outcome may not have the greatest overall benefits for physical activity or health, given the adverse effects of greater exposure to air pollution and social inequalities.

Chapter 3 addressed the limitations of focusing on static neighbourhood measures of the environment used in Chapter 2. Literature that investigated the relationships between the environment and physical activity using the concept of the activity space was systematically reviewed, exploring methodological, analytical, and conceptual issues relevant to causal inference. Included studies answered research questions about features of (shape or size) or environmental characteristics contained within activity spaces using a range of spatial and temporal summary techniques. A key issue related to the conflation of access to and use of the environment and the issue of selective daily mobility bias whereby spaces used as a result of activity were used as a measure of exposure. Distinguishing between potential and actual spaces used for physical activity, and using appropriate measures for research questions, will help to overcome this in future research. Most studies were cross-sectional and the conceptual challenge of using activity spaces to strengthen causal inference was rarely considered, although some studies discussed important markers of causality including circularity, temporality, and plausibility. Findings from the review suggest that the use of longitudinal and experimental designs, as well as qualitative data, in future studies may be useful to strengthen the basis for causal inference.

The subsequent section of the thesis applied the concept of the activity space in the context of a built environment intervention. Drawing on methods identified to delineate activity spaces in Chapter 3, a replicable data processing method to clean and prepare GPS data was developed in Chapter 4. The process was planned, developed, and refined using a random test sample (10% of potential sample) and involved a two-step approach. The first step identified erroneous points based on attribute values available in all GPS datasets. The second step assessed spatial and temporal differences between consecutive points and accounted for technical limitations of GPS data including signal stray and signal loss. The process allowed for erroneous points to be removed from the dataset and for activity space polygons to be derived, and subsequently described by their shape and size. The latter measures formed the basis of analyses investigating changes in activity spaces following a change to the built environment in the following chapter.

Chapter 5 investigated if, how and why spatial patterns of movement changed for individuals, following a change to the local built environment; in this case, the opening of the Cambridgeshire guided busway. I used both qualitative and quantitative data to explore changes in the shape and size of activity spaces and associated sociodemographic characteristics and travel behaviours. This exploratory work was completed using regression analysis and participant profiles which drew together interview data and visualised activity spaces. The quantitative findings showed that non-car owners had more compact activity spaces and participants living in rural areas had larger activity spaces and were less likely to change the size of their activity space in response to the intervention. The temporal patterning of behaviour was important with some associations only shown for weekday or weekend data. For example, participants who lived furthest from the busway were less likely to increase the size of their weekday activity space. The qualitative data suggested that increases in the size of individuals' activity spaces to incorporate the busway was due to the busway being used as an additional commute route, rather than an alternative, and as a new space for recreational activity. Understanding the mechanisms of changes in spatial habits is useful for evaluating how and why interventions are used, their wider effects on existing behaviours, and for strengthening the basis for causal inference.

Building upon methods used to describe and measure changes in spatial patterns of movement in Chapter 5, Chapter 6 explored geospatial analysis methods that could be used to assess changes in the spatial locations of physical activity. GPS data prepared in Chapter 4 were matched with objective physical activity data and methods for analysing and displaying data were piloted and their feasibility reviewed. Physical activity that occurred on the busway was also quantified and its contribution to changes in overall levels of physical activity explored. Population-level point-to-raster maps were effective in identifying absolute changes in locations of physical activity and provided a useful foundation for more complex spatial analysis. Spatial autocorrelation maps identified significant locations used for physical activity and were useful for dealing with anomalous data. In contrast, kernel density estimation prioritised locations where most time is spent such as workplaces. While less appropriate for identifying locations of activity due to the spatially transient nature of activities such as walking and cycling, kernel density methods may be applicable for temporally weighting exposure to environmental characteristics. Descriptive results showed that the majority of new busway users increased the physical activity that was undertaken on the busway and their overall levels of physical activity. The geospatial methods presented can provide context and meaning to quantitative analysis. The combination of geospatial analysis at the population level with individual level analysis should be considered to more fully understand how changes in the environment affect locations of physical activity, and whether use of a new space increases or displaces overall activity levels.

7.3 Implications of research

The research in this thesis explored the relationship between the environment and physical activity and different methods to measure this in order to strengthen the basis for causal inference. Consequently, the findings have implications for the field and for public health policy.

7.3.1 Implications for the field of environment and physical activity

7.3.1.1 Theoretical implications

Key theoretical implications of the research relate to the consideration of a range of environmental factors in combination, the application of the activity space to further causality, and the contribution of particular behaviours or use of spaces to overall levels of physical activity.

Much of existing evidence in the field has focused on a specific microscale characteristic of the environment and its association with a single measure of physical activity [14], [45], [59]. This thesis showed that environmental conditions are not experienced in isolation. Without accounting for wider determinants of physical activity and health, it is difficult to understand the role of the environment in influencing physical activity behaviours, and therefore which environments might be conducive to activity. For example, walkable neighbourhoods are associated with higher levels of walking but so too are more deprived neighbourhoods where walking may occur out of necessity and where social and health inequalities are greater. This study was also one of the first to show that higher concentrations of air pollution were associated with more walking, suggesting that specific behaviours may still occur, irrespective of the quality of the environment, which may have adverse implications for health. Greater theoretical consideration is therefore needed of how characteristics might interact and how interventions could be designed given the potential harms and benefits.

The limitations of using static neighbourhood measures of the environment alongside the increased availability of location data has seen recent use of the concept of the activity space in studies of the environment and physical activity [70]. The activity space provides a new way of thinking in understanding interactions with the environment at increasingly refined spatial and temporal scales [70]. However, its use has given rise to a new set of challenges with regards to defining exposure and causality, including selective daily mobility bias [152]. The thesis synthesised these challenges and opportunities and explicitly documented the different spatial methods employed to delineate activity spaces and different research questions which may be answered. It showed that there has been inconsistency in the methods applied, a lack of consideration for the difference between potentially accessible and used spaces, and limited focus on causality in general. In doing so, the findings provide a timely resource for researchers making use of location data to apply the most appropriate method for the specified research question and to translate their use into better inference in future studies.

Lastly, the research considered the contribution of specific behaviours and activity in new spaces to overall physical activity. In the UK, increases in active travel have been shown to contribute to increases in physical activity [253], [259], [293]. Although exploratory, analysis within the thesis showed how use of new infrastructure may contribute to increases in physical activity. This suggests that changes to the built environment have the potential to provide locations for new active behaviours. However, this might be dependent on individual characteristics, as supported by the different associations for different types of physical activity based on sociodemographic environments, such as area level deprivation shown in Chapter 2, and the diverse responses to the busway shown in Chapter 5. Whilst small and divergent changes are shown in the population, descriptive analysis and qualitative data highlight the potential for big changes for certain individuals [248] which may also explain small or conflicting aggregate effects previously observed [252], [253]. The way in which exposure is theorised and who is exposed and whose behaviour may change requires careful thought to assist further investigation into the impact of changes to the built environment on population health [294].

7.3.1.2 Methodological implications

The key methodological implications of the thesis are centred around the GPS data cleaning process, the application of combined data sources, and the testing of different geospatial methods to visualise changing locations of physical activity.

A key limitation of using GPS data to locate physical activity in the current literature is the lack of openly available software to remove signal stray and account for signal loss, and lack of clear description of methods used in the cleaning process by others, which would permit replication and improvement [67]. The cleaning process developed in Chapter 4 removes erroneous and outlying points. I intend to make the code available as a tool in the ArcGIS Toolbox for researchers following publication of work from Chapter 5 which, due to its limited dependence on attribute values, may be applied to alternative datasets. Although it is acknowledged that not all points where signal stray has occurred may be removed, it provides an important first step in allowing consistency in GPS data processing in future studies. The method can be adapted to suit data collected at different epochs and as temporal differences between consecutive points are calculated, signal loss is easily identified. By estimating information on signal quality, the method also provides a useful resource for researchers who choose to impute to improve completeness of datasets.

The thesis uses data in an exploratory way with the application of qualitative data, alongside quantitative analysis and maps of activity spaces to understand how and why changes in spatial patterning of movement and behaviour occur. Five of the 47 studies identified in the systematic review used both qualitative and quantitative methods. However, only one of those studies combining quantitative and qualitative data was longitudinal and none applied methods in the context of an intervention, highlighting the novelty of the approach [244]. The inclusion of data from multiple datasets allowed for a more granular interpretation of an individual's use of space and more in depth understanding of how and why spaces may or may not be used. The application and further development of such methods in future studies provides an opportunity to identify mechanisms and develop stronger evidence on the pathways which act to influence use of spaces and changes in behaviour.

While the field has made progress in its inclusion of spatial and temporal data to better understand mobility and use of spaces, there are few widely recognised and accepted methods for analysing changes in the spatial patterning of physical activity over time. The thesis makes an important contribution in addressing this by presenting different geospatial methods to visualise population changes in physical activity and reviewing their feasibility to do so in alternative datasets. Although these techniques need further refinement, the preliminary findings from this thesis underscore potential directions for future research efforts and methodologies, which may help to strength evidence for future policy and planning.

7.3.2 Public health context, and implications for policy and planning

Taken together, the findings from the analyses highlight the potential of the environment to contribute to levels of physical activity. Changing the environmental determinants of physical activity through the development of new infrastructure, such as a traffic-free walking and cycling routes, may enable a population shift in physical activity to be achieved. However, the uptake of additional active behaviours such as walking and cycling through the use of new infrastructure is likely to depend on the accessibility of the facility and the presence of other environmental and sociodemographic factors. For example, the findings from Chapter 5 suggest that access to the busway and convenience of the guided busway service, over and above other alternatives, encouraged its use. In order for longer and more active journeys to be made, however, the provision of lighting and security for bicycles may be required, complementing findings from previous studies [247]–[249].

Socioecological models suggest that there are many drivers of behaviour [13], [24]. Interventions that target a single aspect of the environment may be insufficient in achieving large scale shifts in behaviour, but provide an important lever for change. For example, the findings presented in this thesis support the availability of walkable environments and accessibility of a new walk and cycle path for promoting overall physical activity. However, adults who lived closest to the pathway were more responsive to using it for active travel (Chapter 5), and while people who lived in more deprived or polluted areas may have walked more, their overall levels of physical activity were typically lower and exposure to alternative health risks higher (Chapter 2). The location of new infrastructure developments and their connection to other environmental factors are therefore important and may be experienced differently by different groups of people. For example, connecting rural communities to services such as employment centres and local shops may facilitate trip mode transition for groups previously dependent on car travel and allow for longer active journeys to be made, as illustrated in the 3D transect for those travelling further distances along the busway path (Chapter 6). Conversely, Chapter 5 showed that urban dwellers typically had smaller activity spaces and achieved lower levels of physical activity overall but that new infrastructure allowed for new spaces to be used. Enabling urban dwellers to access greenspaces and nature reserves that were previously inaccessible for recreational activity may have positive implications, particularly if targeted at residents of more polluted or deprived areas. Although exploratory, if similar findings are replicated in larger studies, these highlight potential types of interventions and implications for overall physical activity and health that may be communicated to urban and transport planners.

This thesis focused on the relationship between the environment and physical activity outcomes and showed that environmental interventions have the potential to create new spaces for physical activity. Considered in a broader context, the development of active environments, their use, and resultant increases in physical activity have the potential to contribute to a range of physical and mental health co-benefits, including reduced risk of cardiovascular disease, type 2 diabetes, premature mortality, and improved mental wellbeing at the population level [1], [2], [35], [36]. Changes to transport infrastructure which encourage active over passive modes of travel have the potential for increased social interaction [5], [6], and improved environmental sustainability through the reduced car use, congestion and carbon emissions [7], [8]. Several of these effects align with sustainable development goals [47]. However, findings from Chapter 2 also showed that some interventions that focus on specific outcomes may have trade-offs for health, such as the effect on cardiorespiratory health from of walking in polluted areas. Methods presented in this thesis, complemented by high quality study designs and data, demonstrate transferrable ways in which people's interactions with place may be investigated. For example, maps created in Chapters 5 and 6 suggest the busway provided new access to greenspaces for leisure physical. Its use may therefore be a mechanism for improved wellbeing, as supported by literature on greenspaces and physical and mental health [37], [38], [41], [42].

It is important to conceptualise health in a complex system model, considering the interplay of different elements and actors within a connected whole [24]. This thesis reiterates the importance of investigating multiple factors and physical activity outcomes in combination (Chapter 2). However, its focus is on environmental factors, while Rutter and colleagues identify societal, socio-political, individual and biological factors as related determinants in their systems map of physical activity [48]. Drawing on methods used in this thesis, and considering multiple related outputs of physical activity, may improve understanding of potential feedback mechanisms within the system and ways to best coordinate interventions across multiple domains.

The latter chapters of the thesis investigate a specific change in transport infrastructure. Investment in similar infrastructural changes such as walkable city centres, public transport with attention to pedestrians and cyclists, parks, and public safety can have major implications for levels of physical activity. However, investment may be driven by alternative concerns relating to economic development or climate change [295]. An isolated public health strategy is therefore unlikely to be successful. Many actions necessary for investment and development require broader political support and coordination from multiple actors and sectors additional to public health [50], [55]. For effective change, wider effects on behaviour and health, including socioeconomic and cultural factors, should also be considered in research and decision-making [295], [296]. In Chapter 2, the role of culture in place for specific activities such as walking for pleasure was highlighted. In Chapters 5 and 6, the potential uptake of additional active travel on new transport infrastructure was shown in Cambridge, a city where cycling is prevalent and a widely accepted mode of transport. Interventions which target the built environment must therefore adapt to the social norms and consider the realities of the context in which the intervention is being implemented.

Changes to the built environment align with a shift in focus and funding from individual-level to population-level approaches in public health [50]. There is evidence that such approaches have contributed to the large scale promotion of physical activity in some settings [295]. However, considering the notions behind complex system and socioecological theories whereby multiple factors interact, physical activity can be targeted at multiple levels of influence. To improve population levels of physical activity requires the recognition of individual behaviours as key elements that affect population health. In Chapter 5, qualitative data revealed that some participants were unable to shift their behaviour to incorporate active use of the busway due to intrapersonal factors such as health issues. Consequently, some population groups are missed where more specialist support may have an impact and help to reduce inequalities [296]. The qualitative data also showed that new spaces should be designed to provide a compelling alternative to present infrastructure for additional and sustained physical activity and promoted in a way that prevents the displacement of activity to alternative locations. Potential organisational-level approaches to promote walking and cycling were highlighted by participants and corroborated with those described in systematic reviews [297]–[299]. These include access to showers in the workplace, security provisions such as adequate lighting and bicycle parking, or financial incentives to actively commute or use public transport. It is likely that in isolation, environmental changes might be necessary but not sufficient [300]. Instead, multi-level approaches, such as the combination of behavioural and environmental interventions, may be more effective at facilitating behaviour change than one that targets only one level.

7.4 Strengths and limitations

Strengths and limitations of studies and methods have been discussed in detail in each chapter. In this section, I therefore focus on the broad benefits and weakness of the thesis in its totality.

A key strength of the thesis was the use of two complementary datasets that allowed for a broader exploration of environmental determinants of different physical activity outcomes, and the triangulation of qualitative and quantitative spatial data. Self-reported and objective measures of physical activity and use of a built environment intervention (the Cambridgeshire Guided busway) allowed for different behaviours and use from different timeframes to be investigated. This meant that over-reporting from questions relating to any use of the busway could be accounted for through GPS-measures which may capture more regular use and the contribution of behaviours such as walking to overall levels of activity could be considered.

Jankowska and colleagues describe the sensitivity and accuracy of various methodological combinations as placed on a continuum with self-reported measures of the environment and physical activity offering an indication of relevant factors which can be feasibly applied in large datasets and the combination of GPS, GIS, and accelerometer data providing the most specific and accurate measure of physical activity and contexts, but at high cost [68]. The use of two different datasets in the thesis allowed for the benefits of both ends of the continuum to be realised and for different research questions to be answered. By incorporating qualitative data, I was further able to build on the use of GPS data and gain insight into physical activity behaviours and mechanisms for how spaces are used and why that might change.

The review in Chapter 3 provided in depth reflection on the concept of the activity space which is increasingly being used in the literature to address limitations of static measures of the environment but with little justification for the methods chosen or consideration for their implications for causality. Based on the findings of the review I was able to apply the concept of the activity space within a longitudinal study of an intervention, building on primarily cross-sectional study designs that have gone before. The availability of data both pre- and post-intervention allowed for methods to detect changes in the spatial patterning of movement and physical activity to be tested. However, due to the small sample size of the Commuting and Health dataset, the latter chapters of the thesis were exploratory in nature and it was not possible to investigate associations due to limited statistical power. Although studies in Chapters 5 and 6 were not definitive and clear scientific conclusions cannot be drawn, they provide valuable lessons in key methodological considerations and hypotheses for exploration in future research.

7.5 Recommendations for future work

In order to provide stronger evidence for the development of effective policies to promote physical activity, more longitudinal and intervention studies are required. The use of natural experiments overcome practical difficulties and ethical concerns associated with designing a randomised control trial to explore responses in behaviour to changes in the environment [301]. Environmental interventions may include new greenspaces or changes to transport infrastructure such as traffic calming or the development of walk and cycle paths.

Future studies should consider a broad range of environmental factors and the ways in which they may interrelate. Although specific environments may be closely related to specific types of activity, changing just one element of the environment may have trade-offs for health such as an increase of walking in polluted areas. Promoting activity in a new space may also spatially displace activity from one place to another, without increasing overall levels of physical activity. Further studies which investigate changes in spatial patterning of physical activity and substitution effects are therefore required. These may incorporate novel concepts which are emerging in the literature such as time use analysis and compensation effects of activities [253], [277], [302].

In recent decades, the use of GPS trajectories and advanced GIS methods in physical activity research have emerged as viable options for enhancing understanding of associations between physical activity and physical and social environments [151]. Their use fills a gap in the literature. Where previous research has focused on physical activity intensities, largely ignoring the role of different places, GPS data enable individual health behaviours and specific types of activity over time and space to be investigated [68]. As rich locational data and automated processes for data cleaning and modelling become increasingly available, I anticipate and encourage this line of research to continue. The use of smartphones and their sensing capabilities allow for data to be collected and communicated in real time with minimal outlay of time and effort [303]. Smartphone technology is a fast-developing resource for capturing big data and estimating different types of physical activity in free-living populations [304]. Its widespread use creates opportunity to research previously underrepresented populations in lower and middle income countries and in lower-SES groups in higher income countries where smartphone proliferation and use is highest and health risks greatest [304]–[306].

To use GPS data to its full potential and conduct meaningful and comparable research requires robust and replicable data processing and analysis methods. Findings from the systematic review in Chapter 3 showed that authors did not always report data cleaning processes, the level at which data had been accumulated, or provide justification for the metrics used to define exposure. Moving forwards, work should be reported with greater transparency, particularly in relation to the filtering, cleaning and aggregation of spatial data [307], [308]. To my knowledge, there is no standard method for cleaning GPS data. Following publication of a manuscript summarising findings from work described in Chapters 4 and 5, I plan to share my code for cleaning GPS data via github.com, an opensource repository for sharing, storing and managing code. In doing so, I welcome researchers to use and test the process on different GPS datasets and to discuss and iterate changes with the view of developing a standard automated procedure for high volume data.

Findings from Chapter 2 suggest that it is important to understand specific activities, as well as overall levels of physical activity at different intensities. Algorithms and machine learning approaches to classify different behaviours from smartphone and combined GPS and accelerometer data have been tested and published [309], [310]. Whilst these show promise for detecting and locating behaviours from large objective datasets, further development is required to improve their accuracy in free-living conditions [310]. The field should build on existing examples from other disciplines such as transportation where the application of machine learning has been used to automatically detect travel modes [311], [312]. The collection of supplementary data using daily activity diaries should also be considered in future research to provide more detailed information on different types of activity, and for ground truth purposes.

As well as improved and standardised methods to process spatial and physical activity data, the metrics used to measure environmental context require careful consideration. The activity space is an important concept for examining relationships between the environment and health more accurately. However, there is a need to move beyond the simplistic spatial designs largely employed in studies in Chapter 3, to incorporate temporal dimensions, drawing on methods used in fields such as ecology, and qualitative data. The use of methods to delineate the activity space must be applicable for the research question and the issue of selective daily mobility bias must be addressed. A consistent challenge with environment and physical activity research has been the inability to make comparisons across studies due to heterogeneity in methods employed. The studies included in the review in Chapter 3 further highlight that this issue remains prevalent in studies applying the activity space. To mitigate bias and strengthen the basis for causal inference, studies applying the concept of the activity space should distinguish between potentially accessible environments and those used. A standard measure of the activity space may also be used in sensitivity analyses to test for selective daily mobility bias and enable comparisons between studies.

The work in this thesis quantified environmental characteristics within the residential neighbourhood and assessed features of activity spaces using metrics relating to their shape and size. Moving forwards, it may be useful for studies to quantify environmental characteristics within activity spaces, to understand whether environments accessible to individuals are used for physical activity. Methods to achieve this may include kernel density estimation to account for temporally weighted exposures based on where individuals spend most of their time, alongside measures of spaces used for physical activity. Although focused at developing population-level maps, the findings from the feasibility study of geospatial methods in Chapter 6 provide potential ways to identify spaces used for physical activity to understand within-individual associations and change.

Lastly, a greater focus on causality is required. The findings from Chapter 3 showed that the focus on causality in much of the activity space literature has been limited. To develop causal explanation and to better understand how and why environments are used, well designed studies, sufficient data and deep-thinking about causal relationships are required. The use of both qualitative and quantitative data in combination is recommended. Where this is not applicable in larger studies, more detail may be collected from select individuals to provide further exploration of social and psychological factors that relate to physical activity, and insight into the mechanisms behind the use of space and changes in behaviour.

7.6 Overall summary

This thesis explored different methods and types of data to better understand environmental influences and use of space for physical activity.

Results from the thesis showed that the environment has the potential to contribute to different physical activities. Spatial data and the application of the activity space enable the environmental contexts of physical activity to be more accurately identified. However, methods used to measure spaces used for physical activity must be carefully aligned with research questions to limit issues relating to selective daily mobility bias.

Changes to the built environment may shape people's use of space, behaviours, and overall levels of physical activity. In order for public health initiatives to effectively promote the use of new spaces for additional physical activity, consideration should be given to the location and connectedness of interventions, potential trade-offs and exposures to alternative environments, as well as individual characteristics.

References

[1] I. M. Lee *et al.*, “Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy,” *Lancet*, vol. 380, no. 9838, pp. 219–229, 2012.

[2] US Department of Health and Human Services, “Physical Activity Guidelines Advisory Committee report, 2008.,” *Nutr. Rev.*, vol. 67, no. 2, pp. 114–120, 2009.

[3] World Health Organization, “Physical activity,” 2016. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs385/en/. [Accessed: 04-Jun-2016].

[4] World Health Organization, “Global Recommendations on Pysical Activity For Health,” *WHO Libr. Cat. Data*, pp. 1–60, 2010.

[5] World Health Organization (WHO), “Physical activity promotion in socially disadvantaged groups: principles for action. PHAN Work Package 4 Final Report.,” Copenhagen, 2013.

[6] R. F. Hunter, M. Boeri, M. A. Tully, P. Donnelly, and F. Kee, “Addressing inequalities in physical activity participation: Implications for public health policy and practice,” *Prev. Med. (Baltim).*, vol. 72, pp. 64–69, 2015.

[7] J. Woodcock, D. Banister, P. Edwards, A. M. Prentice, and I. Roberts, “Energy and transport,” *Lancet*, vol. 370, no. 9592, pp. 1078–1088, Sep. 2007.

[8] B. Giles-Corti, S. Foster, T. Shilton, and R. Falconer, “The co-benefits for health of investing in active transportation,” *N. S. W. Public Health Bull.*, vol. 21, no. 5–6, pp. 122–127, 2010.

[9] G. Stevens, “Global Health Risks: Mortality and burden of disease attributable to selected major risks,” *Bull. World Health Organ.*, vol. 87, pp. 646–646, 2009.

[10] World Health Organization, “Global Health Observatory Data Repository: Prevalence of insufficient physical activity among adults,” 2016. [Online]. Available: http://apps.who.int/gho/data/node.main.A893?lang=en. [Accessed: 04-Jun-2016].

[11] Department of Health Physical Activity Health Improvement and Protection, “Start Active , Stay Active,” *Report*, p. 62, 2011.

[12] UKactive, “Turning the tide of inactivity,” 2014. [Online]. Available: http://www.ukactive.com/turningthetide/pdf/Turning the tide of inactivity.pdf. [Accessed: 18-Jun-2016].

[13] J. F. Sallis, N. Owen, and E. B. Fisher, *Ecological models of health behavior*. 2008.

[14] A. E. Bauman, R. S. Reis, J. F. Sallis, J. C. Wells, R. J. F. Loos, and B. W. Martin, “Correlates of physical activity: Why are some people physically active and others not?,” *Lancet*, vol. 380, no. 9838, pp. 258–271, 2012.

[15] J. N. Morris, S. P. W. Chave, C. Adam, C. Sirey, L. Epstein, and D. J. Sheehan, “VIGOROUS EXERCISE IN LEISURE-TIME AND THE INCIDENCE OF CORONARY HEART-DISEASE,” *Lancet*, 1973.

[16] N. T. Artinian *et al.*, “Interventions to promote physical activity and dietary lifestyle changes for cardiovascular risk factor reduction in adults: A scientific statement from the american heart association,” *Circulation*. 2010.

[17] M. Hillsdon, C. Foster, and M. Thorogood, “Interventions for promoting physical activity.,” *Cochrane database Syst. Rev.*, no. 1, p. CD003180, 2005.

[18] L. McLaren, L. McIntyre, and S. Kirkpatrick, "Rose's population strategy of prevention need not increase social inequalities in health," *Int. J. Epidemiol.*, 2010.

[19] S. Capewell and H. Graham, "Will cardiovascular disease prevention widen health inequalities?," *PLoS Med.*, 2010.

[20] J. Adams, O. Mytton, M. White, and P. Monsivais, "Why Are Some Population Interventions for Diet and Obesity More Equitable and Effective Than Others? The Role of Individual Agency," *PLoS Medicine*. 2016.

[21] T. R. Frieden, "A framework for public health action: The health impact pyramid," *American Journal of Public Health*. 2010.

[22] G. Rose, K. T. Khaw, and M. Marmot, *Rose's Strategy of Preventive Medicine*. 2009.

[23] G. Rose, "Sick individuals and sick populations," *Int. J. Epidemiol.*, 1985.

[24] G. Dahlgren and M. Whitehead, "Policies and strategies to promote social equity in health," *Stock. Inst. Futur. Stud.*, 1991.

[25] C. Dora, "A different route to health: Implications of transport policies," *British Medical Journal*. 1999.

[26] P. Das and R. Horton, "Rethinking our approach to physical activity," *The Lancet*. 2012.

[27] Public Health England, "Everybody active, every day. An evidence-based approach to physical activity," 2014.

[28] National Institute for Health and Clinical Excellence and Nice, "Promoting and creating built or natural environments that encourage and support physical activity," 2008.

[29] NICE, "Walking and cycling overview - NICE Pathways," *National Institute for Health and Care Excellence.*, 2012. .

[30] D. Mozaffarian *et al.*, "Population Approaches to Improve Diet, Physical Activity, and Smoking Habits: A Scientific Statement From the American Heart Association.," *Circulation*, vol. 126, no. 12, pp. 1514–1563, 2012.

[31] J. F. Sallis *et al.*, "Co-benefits of designing communities for active living: an exploration of literature," *Int. J. Behav. Nutr. Phys. Act.*, 2015.

[32] J. F. Sallis, R. B. Cervero, W. Ascher, K. a. Henderson, M. K. Kraft, and J. Kerr, "An Ecological Approach To Creating Active Living Communities," *Annu. Rev. Public Health*, vol. 27, no. 1, pp. 297–322, 2006.

[33] K. McPherson, T. Marsh, and M. Brown, "Foresight: Tackling Obesities: Future Choices – Modelling Future Trends in Obesity and the Impact on Health," *Gov. Off. Sci.*, 2007.

[34] S. A. Lear *et al.*, "The effect of physical activity on mortality and cardiovascular disease in 130 000 people from 17 high-income, middle-income, and low-income countries: the PURE study," *Lancet*, 2017.

[35] R. D. Goodwin, "Association between physical activity and mental disorders among adults in the United States," *Prev. Med. (Baltim).*, 2003.

[36] A. Ströhle, "Physical activity, exercise, depression and anxiety disorders," *Journal of Neural Transmission*. 2009.

[37] T. Hartig, G. W. Evans, L. D. Jamner, D. S. Davis, and T. Gärling, "Tracking restoration in

natural and urban field settings," *J. Environ. Psychol.*, 2003.

[38] J. Pretty, J. Peacock, R. Hine, M. Sellens, N. South, and M. Griffin, "Green exercise in the UK countryside: Effects on health and psychological well-being, and implications for policy and planning," *J. Environ. Plan. Manag.*, 2007.

[39] S. Steinmo, G. Hagger-Johnson, and L. Shahab, "Bidirectional association between mental health and physical activity in older adults: Whitehall II prospective cohort study," *Prev. Med. (Baltim).*, 2014.

[40] M. A. Da Silva *et al.*, "Bidirectional association between physical activity and symptoms of anxiety and depression: The whitehall II study," *Eur. J. Epidemiol.*, 2012.

[41] T. Takano, K. Nakamura, and M. Watanabe, "Urban residential environments and senior citizens' longevity in megacity areas: the importance of walkable green spaces.," *J. Epidemiol. Community Health*, vol. 56, no. 12, pp. 913–918, Dec. 2002.

[42] T. Astell-Burt, X. Feng, and G. S. Kolt, "Is neighborhood green space associated with a lower risk of type 2 diabetes evidence from 267,072 australians," *Diabetes Care*, vol. 37, no. 1, pp. 197–201, Jan. 2014.

[43] I. C. Mills, R. W. Atkinson, S. Kang, H. Walton, and H. R. Anderson, "Quantitative systematic review of the associations between short-term exposure to nitrogen dioxide and mortality and hospital admissions," *BMJ Open*. 2015.

[44] R. D. Brook *et al.*, "Particulate matter air pollution and cardiovascular disease: An update to the scientific statement from the American Heart Association," *Circulation*, 2010.

[45] B. Giles-Corti *et al.*, "City planning and population health: a global challenge," *The Lancet*. 2016.

[46] J. F. Sallis *et al.*, "Use of science to guide city planning policy and practice: how to achieve healthy and sustainable future cities," *The Lancet*. 2016.

[47] World Health Organization, "Global Action Plan on Physical Activity 2018 - 2030 - More active people for a healthier world," 2018.

[48] H. Rutter, N. Cavill, A. Bauman, and F. Bull, "Systems approaches to global and national physical activity plans," *Bulletin of the World Health Organization*. 2019.

[49] J. D. Wright, *International Encyclopedia of the Social & Behavioral Sciences: Second Edition*. 2015.

[50] H. Rutter *et al.*, "The need for a complex systems model of evidence for public health," *Lancet*, 2017.

[51] A. M. Bagnall *et al.*, "Whole systems approaches to obesity and other complex public health challenges: A systematic review," *BMC Public Health*. 2019.

[52] S. Galea, M. Riddle, and G. A. Kaplan, "Causal thinking and complex system approaches in epidemiology," *Int. J. Epidemiol.*, 2010.

[53] T. A. Glass, S. N. Goodman, M. A. Hernán, and J. M. Samet, "Causal Inference in Public Health," *Annu. Rev. Public Health*, 2013.

[54] T. Dunning, *Natural experiments in the social sciences: A design-based approach*. 2009.

[55] D. T. Finegood, "The Complex Systems Science of Obesity," in *The Oxford Handbook of the Social Science of Obesity*, 2012.

[56] K. K. Davison and C. T. Lawson, “Do attributes in the physical environment influence children’s physical activity? A review of the literature,” *Int. J. Behav. Nutr. Phys. Act.*, 2006.

[57] D. Ding, J. F. Sallis, J. Kerr, S. Lee, and D. E. Rosenberg, “Neighborhood environment and physical activity among youth: A review,” *American Journal of Preventive Medicine*, vol. 41, no. 4. pp. 442–455, 2011.

[58] G. R. McCormack and A. Shiell, “In search of causality: a systematic review of the relationship between the built environment and physical activity among adults,” *Int. J. Behav. Nutr. Phys. Act.*, vol. 8, no. 1, p. 125, 2011.

[59] V. Van Holle *et al.*, “Relationship between the physical environment and different domains of physical activity in European adults: a systematic review,” *BMC Public Health*, vol. 12, no. 1, p. 807, 2012.

[60] W. Wendel-Vos, M. Droomers, S. Kremers, J. Brug, and F. van Lenthe, “Potential environmental determinants of physical activity in adults: a systematic review,” *Obes Rev*, vol. 8, no. 5, pp. 425–440, 2007.

[61] B. E. Saelens, J. F. Sallis, and L. D. Frank, “Environmental correlates of walking and cycling: findings from the transportation, urban design, and planning literatures,” *Ann. Behav. Med. a Publ. Soc. Behav. Med.*, vol. 25, no. 2, pp. 80–91, 2003.

[62] N. Owen, N. Humpel, E. Leslie, A. Bauman, J. F. Sallis, and R. J. Donovan, “Understanding environmental influences on walking,” *Am. J. Prev. Med.*, 2004.

[63] J. Van Cauwenberg, I. De Bourdeaudhuij, F. De Meester, and A. Et, “Relationship between the physical environment and physical activity in older adults: a systematic review.,” *Health Place*, vol. 17, pp. 458–469, 2011.

[64] E. Cerin, A. Nathan, J. van Cauwenberg, D. W. Barnett, and A. Barnett, “The neighbourhood physical environment and active travel in older adults: a systematic review and meta-analysis,” *Int. J. Behav. Nutr. Phys. Act.*, 2017.

[65] S. Cummins, S. Curtis, A. V. Diez-Roux, and S. Macintyre, “Understanding and representing ‘place’ in health research: A relational approach,” *Soc. Sci. Med.*, 2007.

[66] P. M. Hurvitz, A. V. Moudon, B. Kang, B. E. Saelens, and G. E. Duncan, “Emerging technologies for assessing physical activity behaviors in space and time.,” *Front. public Heal.*, vol. 2, p. 2, 2014.

[67] P. James *et al.*, “‘Spatial Energetics’: Integrating Data From GPS, Accelerometry, and GIS to Address Obesity and Inactivity.,” *Am. J. Prev. Med.*, vol. 51, no. 5, pp. 792–800, Nov. 2016.

[68] M. M. Jankowska, J. Schipperijn, and J. Kerr, “A framework for using GPS data in physical activity and sedentary behavior studies,” *Exerc. Sport Sci. Rev.*, vol. 43, no. 1, pp. 48–56, Jan. 2015.

[69] L. Yi, J. P. Wilson, T. B. Mason, R. Habre, S. Wang, and G. Dunton, “Methodologies for assessing contextual exposure to the built environment in physical activity studies: A systematic review,” *Health Place*, vol. 60, 2019.

[70] C. Perchoux, B. Chaix, S. Cummins, and Y. Kestens, “Conceptualization and measurement of environmental exposure in epidemiology: Accounting for activity space related to daily mobility,” *Heal. Place*, vol. 21, pp. 86–93, 2013.

[71] P. Craig *et al.*, “Using natural experiments to evaluate population health interventions: New medical research council guidance,” *J. Epidemiol. Community Health*, 2012.

[72] Public Health England, “Health profile for England. Chapter 6: social determinants of health,” 2017. [Online]. Available: https://www.gov.uk/government/publications/health-profile-for-england/chapter-6-social-determinants-of-health. [Accessed: 20-Apr-2017].

[73] M.-P. Kwan, “The Uncertain Geographic Context Problem,” *Ann. Assoc. Am. Geogr.*, vol. 102, no. 5, SI, pp. 958–968, 2012.

[74] S. E. Spielman and E. Yoo, “The spatial dimensions of neighborhood effects,” *Soc. Sci. Med.*, 2009.

[75] S. Cummins, “Commentary: Investigating neighbourhood effects on health - Avoiding the ‘local trap,’” *International Journal of Epidemiology*. 2007.

[76] M. P. Kwan, “How GIS can help address the uncertain geographic context problem in social science research,” *Ann. GIS*, 2012.

[77] S. M. Santos, D. Chor, and G. L. Werneck, “Demarcation of local neighborhoods to study relations between contextual factors and health,” *Int. J. Health Geogr.*, 2010.

[78] UK Biobank, “UK Biobank Protocol for a large-scale prospective epidemiological resource.” 2010.

[79] A. Doherty *et al.*, “Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK Biobank Study,” *PLoS One*, vol. 12, no. 2, 2017.

[80] C. Sarkar, C. Webster, and J. Gallacher, “UK Biobank Urban Morphometric Platform (UKBUMP) – a nationwide resource for evidence-based healthy city planning and public health interventions,” *Ann. GIS*, vol. 21, no. 2, pp. 135–148, 2015.

[81] C. Sarkar, J. Gallacher, and C. Webster, “Morphometric Analysis of the Built Environment in UK Biobank: Data Analyses and Specification Manual,” 2014. [Online]. Available: https://biobank.ctsu.ox.ac.uk/crystal/docs/ard-1195_UKBBuiltEnvWales.pdf. [Accessed: 17-May-2017].

[82] G. D. Batty, C. R. Gale, M. Kivimäki, I. J. Deary, and S. Bell, “Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis,” *BMJ*, 2020.

[83] I. M. Lee and E. J. Shiroma, “Using accelerometers to measure physical activity in large-scale epidemiological studies: Issues and challenges,” *British Journal of Sports Medicine*. 2014.

[84] E. J. Shiroma, P. S. Freedson, S. G. Trost, and I. M. Lee, “Patterns of accelerometer-assessed sedentary behavior in older women,” *JAMA - J. Am. Med. Assoc.*, vol. 310, no. 23, pp. 2562–2563, 2013.

[85] A. J. Swerdlow *et al.*, “The Breakthrough Generations Study: Design of a long-term UK cohort study to investigate breast cancer aetiology,” *Br. J. Cancer*, 2011.

[86] V. J. Howard *et al.*, “Obtaining accelerometer data in a national cohort of black and white adults,” *Med. Sci. Sports Exerc.*, 2015.

[87] M. Menai, V. T. Van Hees, A. Elbaz, M. Kivimaki, A. Singh-Manoux, and S. Sabia, “Accelerometer assessed moderate-To-vigorous physical activity and successful ageing: Results from the Whitehall II study,” *Sci. Rep.*, 2017.

[88] I. C. M. Da silva *et al.*, “Physical activity levels in three Brazilian birth cohorts as assessed with raw triaxial wrist accelerometry,” *Int. J. Epidemiol.*, 2014.

[89] D. Ogilvie *et al.*, “Commuting and health in Cambridge : a study of a ‘ natural experiment ’ in the provision of new transport infrastructure,” *BMC Public Health*, vol. 10, no. 1, p. 703, 2010.

[90] D. Ogilvie, J. Panter, C. Guell, A. Jones, R. Mackett, and S. Griffin, “Health impacts of the Cambridgeshire Guided Busway: a natural experimental study,” *Public Heal. Res*, vol. 4, no. 1, 2016.

[91] L. Smith, J. Panter, and D. Ogilvie, “Characteristics of the environment and physical activity in midlife: Findings from UK Biobank,” *Prev. Med. (Baltim).*, vol. 118, 2019.

[92] G. Dahlgren and M. Whitehead, “European strategies for tackling social inequities in health: Levelling up part 2,” *Stud. Soc. Econ. Determ. Popul. Heal.*, 2006.

[93] S. Hankey, J. D. Marshall, and M. Brauer, “Health impacts of the built environment: Within-urban variability in physical inactivity, air pollution, and ischemic heart disease mortality,” *Environ. Health Perspect.*, vol. 120, no. 2, pp. 247–253, 2012.

[94] A. Hajat, C. Hsia, and M. S. O’Neill, “Socioeconomic Disparities and Air Pollution Exposure: A Global Review,” *Curr. Environ. Heal. reports*, vol. 2, no. 4, pp. 440–450, Dec. 2015.

[95] R. An, S. Zhang, M. Ji, and C. Guan, “Impact of ambient air pollution on physical activity among adults: a systematic review and meta-analysis,” *Perspect. Public Health*, vol. 138, no. 2, pp. 111–121, Aug. 2017.

[96] D. W. Esliger and M. S. Tremblay, “Physical activity and inactivity profiling: the next generation.,” *Can. J. public Heal. Rev. Can. santé publique*, vol. 98 (Suppl., 2007.

[97] J. F. Sallis *et al.*, “Physical activity in relation to urban environments in 14 cities worldwide: A cross-sectional study,” *Lancet*, vol. 387, no. 10034, pp. 2207–2217, 2016.

[98] C. Sarkar, “Residential greenness and adiposity: Findings from the UK Biobank,” *Environ. Int.*, vol. 106, pp. 1–10, 2017.

[99] UK Biobank, “UK Biobank Touchscreen Questionnaire: final version,” 2006. [Online]. Available: https://www.ukbiobank.ac.uk/wp-content/uploads/2011/06/Touch_screen_questionnaire.pdf?phpMyAdmin=trmKQlYdjjnQIgJ%2CfAzikMhEnx6. [Accessed: 17-May-2017].

[100] UK Biobank, “UK Biobank Ethics and Governance Framework: version 3.0,” 2007. [Online]. Available: http://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf?phpMyAdmin=trmKQlYdjjnQIgJ%2CfAzikMhEnx6. [Accessed: 17-May-2017].

[101] T. White, K. Westgate, N. J. Wareham, and S. Brage, “Estimation of physical activity energy expenditure during free-living from wrist accelerometry in UK adults,” *PLoS One*, vol. 11, no. 12, 2016.

[102] S. Sabia *et al.*, “Association between questionnaire-and accelerometer-assessed physical activity: The role of sociodemographic factors,” *Am. J. Epidemiol.*, 2014.

[103] G. Peeters, Y. van Gellecum, G. Ryde, N. A. Farías, and W. J. Brown, “Is the pain of activity log-books worth the gain in precision when distinguishing wear and non-wear time for tri-axial accelerometers?,” *J. Sci. Med. Sport*, 2013.

[104] K. R. Evenson, F. Wen, A. Hillier, and D. A. Cohen, “Assessing the contribution of parks to physical activity using global positioning system and accelerometry.,” *Med. Sci. Sports Exerc.*, vol. 45, no. 10, pp. 1981–1987, Oct. 2013.

[105] The IPAQ Group, “International Physical Activity Questionnaire,” 2015. [Online]. Available: https://sites.google.com/site/theipaq/home.

[106] N. J. Wareham, R. W. Jakes, K. L. Rennie, J. Mitchell, S. Hennings, and N. E. Day, “Validity and repeatability of the EPIC-Norfolk Physical Activity Questionnaire.,” *Int. J. Epidemiol.*, vol. 31, no. 1, pp. 168–174, 2002.

[107] C. Bancroft *et al.*, “Association of proximity and density of parks and objectively measured physical activity in the United States: A systematic review,” *Soc. Sci. Med.*, vol. 138, pp. 22–30, 2015.

[108] K. E. Mason, N. Pearce, and S. Cummins, “Associations between fast food and physical activity environments and adiposity in mid-life: cross-sectional, observational evidence from UK Biobank,” *Lancet Public Heal.*, 2018.

[109] M. I. Creatore *et al.*, “Association of Neighborhood Walkability With Change in Overweight, Obesity, and Diabetes,” *Jama*, vol. 315, no. 20, p. 2211, 2016.

[110] S. Hajna *et al.*, “Neighbourhood walkability and home neighbourhood-based physical activity: an observational study of adults with type 2 diabetes,” *BMC Public Health*, vol. 16, no. 1, p. 957, 2016.

[111] J. Van Cauwenberg, V. Van Holle, I. De Bourdeaudhuij, D. Van Dyck, and B. Deforche, “Neighborhood walkability and health outcomes among older adults: The mediating role of physical activity,” *Health Place*, vol. 37, pp. 16–25, 2016.

[112] A. G. Rundle *et al.*, “Using GPS Data to Study Neighborhood Walkability and Physical Activity.,” *Am. J. Prev. Med.*, vol. 50, no. 3, pp. e65–e72, Mar. 2016.

[113] L. D. Frank, T. L. Schmid, J. F. Sallis, J. Chapman, and B. E. Saelens, “Linking objectively measured physical activity with objectively measured urban form: Findings from SMARTRAQ,” *Am. J. Prev. Med.*, vol. 28, no. 2, pp. 117–125, 2005.

[114] L. D. Frank, J. F. Sallis, T. L. Conway, J. E. Chapman, B. E. Saelens, and W. Bachman, “Many Pathways from Land Use to Health: Associations between Neighborhood Walkability and Active Transportation, Body Mass Index, and Air Quality,” *J. Am. Plan. Assoc.*, vol. 72, no. 1, pp. 75–87, Mar. 2006.

[115] H. E. Christian *et al.*, “How important is the land use mix measure in understanding walking behaviour? Results from the RESIDE study,” *Int. J. Behav. Nutr. Phys. Act.*, vol. 8, no. 1, p. 55, 2011.

[116] J. C. Stockton, O. Duke-Williams, E. Stamatakis, J. S. Mindell, E. J. Brunner, and N. J. Shelton, “Development of a novel walkability index for London, United Kingdom: cross-sectional application to the Whitehall II Study,” *BMC Public Health*, vol. 16, no. 1, p. 416, May 2016.

[117] A. Carver, J. R. Panter, A. P. Jones, and E. M. F. van Sluijs, “Independent mobility on the journey to school: A joint cross-sectional and prospective exploration of social and physical environmental influences,” *J. Transp. Heal.*, vol. 1, no. 1, pp. 25–32, 2014.

[118] D. Arvidsson, N. Kawakami, H. Ohlsson, and K. Sundquist, “Physical activity and concordance between objective and perceived walkability,” *Med. Sci. Sports Exerc.*, vol. 44, no. 2, pp. 280–287, 2012.

[119] J. R. Panter, A. P. Jones, E. M. F. Van Sluijs, and S. J. Griffin, “Neighborhood, Route, and School Environments and Children’s Active Commuting,” *Am. J. Prev. Med.*, vol. 38, no. 3, pp. 268–278, 2010.

[120] K. Sundquist, U. Eriksson, N. Kawakami, L. Skog, H. Ohlsson, and D. Arvidsson, “Neighborhood walkability, physical activity, and walking behavior: The Swedish Neighborhood and Physical Activity (SNAP) study,” *Soc. Sci. Med.*, vol. 72, no. 8, pp. 1266–1273, 2011.

[121] R. Beelen *et al.*, “Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe - The ESCAPE project,” *Atmos. Environ.*, vol. 72, pp. 10–23, 2013.

[122] ESCAPE, “ESCAPE Exposure Assessment Manual: version July 2010,” 2010. [Online]. Available: http://www.escapeproject.eu/manuals/ESCAPE_Exposure-manualv9.pdf.

[123] GBD 2015 Risk Factors Collaborators, “Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015,” *Lancet*, vol. 388, no. 10053, pp. 1659–1724, 2016.

[124] E. Dons *et al.*, “NO2 but Not PM2.5 at the Home Address is Associated with Concern over Health Effects of Air Pollution,” *J. Transp. Heal.*, vol. 5, pp. S84–S114, Jun. 2017.

[125] P. Nafstad *et al.*, “Urban air pollution and mortality in a cohort of norwegian men,” *Env. Heal. Perspect*, vol. 112, no. 5, pp. 610–615, 2004.

[126] M. Krzyzanowski and A. Cohen, “Update of WHO air quality guidelines,” *Air Qual. Atmos. Heal.*, vol. 1, no. 1, pp. 7–13, 2008.

[127] S. Kephalopoulos, M. Paviotti, and F. Anfosso-Lédée, “Common noise assessment methods in EU: CNOSSOS-EU,” 2012.

[128] P. James *et al.*, “GPS-based exposure to greenness and walkability and accelerometry-based physical activity,” *Cancer Epidemiol. Biomarkers Prev.*, vol. 26, no. 4, pp. 525–532, 2017.

[129] J. D. Marshall, M. Brauer, and L. D. Frank, “Healthy neighborhoods: Walkability and air pollution,” *Environ. Health Perspect.*, vol. 117, no. 11, pp. 1752–1759, 2009.

[130] R. An and X. Xiang, “Ambient fine particulate matter air pollution and leisure-time physical inactivity among US adults,” *Public Health*, vol. 129, no. 12, pp. 1637–1644, 2015.

[131] G. Sun, R. Haining, H. Lin, N. M. Oreskovic, and J. He, “Comparing the perception with the reality of walking in a hilly environment: An accessibility method applied to a University campus in Hong Kong,” *Geospat. Health*, 2015.

[132] P. Gordon-Larsen, M. C. Nelson, P. Page, and B. M. Popkin, “Inequality in the built environment underlies key health disparities in physical activity and obesity,” *Pediatrics*, 2006.

[133] N. Eliasoph and P. Lichterman, “Culture in interaction,” *Am. J. Sociol.*, 2003.

[134] A. J. Perrin, N. Caren, A. C. Skinner, A. Odulana, and E. M. Perrin, “The unbuilt environment: Culture moderates the built environment for physical activity,” *BMC Public Health*, 2016.

[135] E. A. Richardson, R. Mitchell, N. K. Shortt, J. Pearce, and T. P. Dawson, “Developing summary measures of health-related multiple physical environmental deprivation for epidemiological research,” *Environ. Plan. A*, 2010.

[136] E. A. Richardson, R. J. Mitchell, N. K. Shortt, J. Pearce, and T. P. Dawson, “Evidence-based selection of environmental factors and datasets for measuring multiple environmental deprivation in epidemiological research,” *Environ. Heal. A Glob. Access Sci. Source*, 2009.

[137] J. R. Pearce, E. A. Richardson, R. J. Mitchell, and N. K. Shortt, "Environmental justice and health: the implications of the socio-spatial distribution of multiple environmental deprivation for health inequalities in the United Kingdom," *Trans. Inst. Br. Geogr.*, vol. 35, no. 4, pp. 522–539, Oct. 2010.

[138] E. Rind, N. Shortt, R. Mitchell, E. A. Richardson, and J. Pearce, "Are income-related differences in active travel associated with physical environmental characteristics? A multi-level ecological approach.," *Int. J. Behav. Nutr. Phys. Act.*, vol. 12, p. 73, Jun. 2015.

[139] UK Biobank, "Access matter: representativeness of the UK Biobank resource," 2017. [Online]. Available: http://www.ukbiobank.ac.uk/wp-content/uploads/2017/03/access-matters-representativeness-1.pdf. [Accessed: 17-May-2017].

[140] A. Fry *et al.*, "Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population," *Am. J. Epidemiol.*, 2017.

[141] T. Burgoine and P. Monsivais, "Characterising food environment exposure at home, at work, and along commuting journeys using data on adults in the UK.," *Int. J. Behav. Nutr. Phys. Act.*, vol. 10, no. 1, p. 85, 2013.

[142] L. Smith, L. Foley, and J. Panter, "Activity spaces in studies of the environment and physical activity: A review and synthesis of implications for causality," *Health Place*, 2019.

[143] W. Haskell *et al.*, "Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association," *Circulation*, vol. 116, pp. 1081–1093, 2007.

[144] P. C. Hallal *et al.*, "Global physical activity levels: Surveillance progress, pitfalls, and prospects," *Lancet*, vol. 380, no. 9838, pp. 247–257, 2012.

[145] B. Chaix, "Geographic Life Environments and Coronary Heart Disease: A Literature Review, Theoretical Contributions, Methodological Updates, and a Research Agenda," *Annu. Rev. Public Health*, vol. 30, no. 1, pp. 81–105, 2009.

[146] R. G. Golledge and R. J. Stimson, *Spatial Behavior: A Geographic Perspective*. Guilford Publications, 1997.

[147] K. W. Axhausen, A. Zimmermann, S. Schönfelder, G. Rindsfüser, and T. Haupt, "Observing the rhythms of daily life: A six-week travel diary," *Transportation (Amst).*, vol. 29, no. 95, 2002.

[148] T. Hägerstraand, "What about people in regional science?," *Pap. Reg. Sci.*, vol. 24, no. 1, pp. 7–24, 1970.

[149] S. Schönfelder and K. W. Axhausen, "Activity spaces: Measures of social exclusion?," *Transp. Policy*, vol. 10, no. 4, pp. 273–286, 2003.

[150] C. Harding, Z. Patterson, L. Miranda-Moreno, and S. Zahabi, "Modeling the Effect of Land Use on Activity Spaces," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2323, no. 1, pp. 67–74, 2012.

[151] P. J. Krenn, S. Titze, P. Oja, A. Jones, and D. Ogilvie, "Use of global positioning systems to study physical activity and the environment: A systematic review," *Am. J. Prev. Med.*, vol. 41, no. 5, pp. 508–515, 2011.

[152] B. Chaix *et al.*, "GPS tracking in neighborhood and health studies: A step forward for environmental exposure assessment, A step backward for causal inference?," *Heal. Place*, vol. 21, pp. 46–51, 2013.

[153] Z. Patterson and S. Farber, "Potential Path Areas and Activity Spaces in Application: A

Review," *Transp. Rev.*, vol. 35, no. 6, pp. 679–700, 2015.

[154] R. Maddison and C. Ni Mhurchu, "Global positioning system: a new opportunity in physical activity measurement.," *Int. J. Behav. Nutr. Phys. Act.*, vol. 6, no. 73, 2009.

[155] P. R. Mccrorie, C. Fenton, and A. Ellaway, "Combining GPS, GIS, and accelerometry to explore the physical activity and environment relationship in children and young people – a review," *Int. J. Behav. Nutr. Phys. Act.*, vol. 11, p. 93, 2014.

[156] A. Loveday, L. B. Sherar, J. P. Sanders, P. W. Sanderson, and D. W. Esliger, "Technologies that assess the location of physical activity and sedentary behavior: A systematic review," *J. Med. Internet Res.*, vol. 17, no. 8, p. e192, 2015.

[157] A. Cetateanu and A. Jones, "How can GPS technology help us better understand exposure to the food environment? A systematic review.," *SSM - Popul. Heal.*, vol. 2, pp. 196–205, Dec. 2016.

[158] L. Smith, J. Panter, and L. Foley, "Systematic review of activity spaces in studies of environment, physical activity, and health," *PROSPERO*, 2018. [Online]. Available: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42018087095.

[159] S. N. Zenk *et al.*, "Activity space environment and dietary and physical activity behaviors: a pilot study.," *Health Place*, vol. 17, no. 5, pp. 1150–1161, Sep. 2011.

[160] A. B. Hill, "The Environment and Disease: Association or Causation?," *J. R. Soc. Med.*, vol. 58, no. 5, pp. 295–300, 1965.

[161] J. Pearl, "Causal inference in statistics: An overview," *Stat. Surv.*, vol. 3, pp. 96–146, 2009.

[162] G. W. Imbens and D. B. Rubin, "Rubin Causal Model," in *Microeconometrics*, London: Palgrave Macmillan, 2016, pp. 229–241.

[163] M. Sanchez *et al.*, "Predictors of daily mobility of adults in peri-urban south India," *Int. J. Environ. Res. Public Health*, vol. 14, no. 7, p. 783, Jul. 2017.

[164] M. Kamruzzaman and J. Hine, "Analysis of rural activity spaces and transport disadvantage using a multi-method approach," *Transp. POLICY*, vol. 19, no. 1, pp. 105–120, Jan. 2012.

[165] D. A. Rodriguez *et al.*, "Influence of the Built Environment on Pedestrian Route Choices of Adolescent Girls," *Environ. Behav.*, vol. 47, no. 4, pp. 359–394, May 2015.

[166] D. A. Rodríguez *et al.*, "Out and about: Association of the built environment with physical activity behaviors of adolescent females," *Health Place*, vol. 18, no. 1, pp. 55–62, 2012.

[167] J. A. Hirsch, M. Winters, M. C. Ashe, P. J. Clarke, and H. A. McKay, "Destinations That Older Adults Experience Within Their GPS Activity Spaces: Relation to Objectively Measured Physical Activity," *Environ. Behav.*, vol. 48, no. 1, SI, pp. 55–77, Jan. 2016.

[168] J. A. Hirsch, M. Winters, P. Clarke, and H. McKay, "Generating GPS activity spaces that shed light upon the mobility habits of older adults: a descriptive analysis.," *Int. J. Health Geogr.*, vol. 13, p. 51, Dec. 2014.

[169] T. Franke, M. Winters, H. McKay, and H. Chaudhury, "A grounded visualization approach to explore sociospatial and temporal complexities of older adults' mobility," *Soc. Sci. Med.*, vol. 193, pp. 59–69, 2017.

[170] G. Vich, O. Marquet, and C. Miralles-Guasch, "Suburban commuting and activity spaces: using smartphone tracking data to understand the spatial extent of travel behaviour," *Geogr. J.*,

vol. 183, no. 4, pp. 426–439, Dec. 2017.

[171] M. Kamruzzaman, J. Hine, B. Gunay, and N. Blair, “Using GIS to visualise and evaluate student travel behaviour,” *J. Transp. Geogr.*, vol. 19, no. 1, pp. 13–32, Jan. 2011.

[172] P. A. Plazier, G. Weitkamp, and A. E. van den Berg, “‘Cycling was never so easy!’ An analysis of e-bike commuters’ motives, travel behaviour and experiences using GPS-tracking and interviews,” *J. Transp. Geogr.*, vol. 65, pp. 25–34, Dec. 2017.

[173] N. Colabianchi, C. J. Coulton, J. D. Hibbert, S. M. McClure, C. E. Ievers-Landis, and E. M. Davis, “Adolescent self-defined neighborhoods and activity spaces: spatial overlap and relations to physical activity and obesity.,” *Health Place*, vol. 27, pp. 22–29, May 2014.

[174] B. Chaix, D. Duncan, J. Vallee, A. Vernez-Moudon, T. Benmarhnia, and Y. Kestens, “The ‘Residential’ Effect Fallacy in Neighborhood and Health Studies Formal Definition, Empirical Identification, and Correction,” *EPIDEMIOLOGY*, vol. 28, no. 6, pp. 789–797, Nov. 2017.

[175] S. Kosaka, M. Umezaki, M. Ishikawa, and C. Watanabe, “Physical activity and the neighborhood environment in a heavy snowfall area in Japan: The role of ‘Gangi-dori,’” *Landsc. Urban Plan.*, vol. 123, pp. 124–133, Mar. 2014.

[176] C. P. Tribby, H. J. Miller, B. B. Brown, K. R. Smith, and C. M. Werner, “Geographic regions for assessing built environmental correlates with walking trips: A comparison using different metrics and model designs,” *Heal. Place*, vol. 45, pp. 1–9, 2017.

[177] T. Burgoine, A. P. Jones, R. J. Namenek Brouwer, and S. E. Benjamin Neelon, “Associations between BMI and home, school and route environmental exposures estimated using GPS and GIS: do we see evidence of selective daily mobility bias in children?,” *Int. J. Health Geogr.*, vol. 14, p. 8, Feb. 2015.

[178] N. Chen, C.-H. Wang, and G. Akar, “Geographically Weighted Regression Approach to Investigate Spatial Variations in Activity Space,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2671, pp. 40–50, Jan. 2017.

[179] N. C. Lee, C. Voss, A. D. Frazer, J. A. Hirsch, H. A. McKay, and M. Winters, “Does Activity Space Size Influence Physical Activity Levels of Adolescents? - A GPS study of an urban environment.,” *Prev. Med. reports*, vol. 3, pp. 75–78, 2016.

[180] Y. Fan, “The built environment, activity space, and time allocation: An activity-based framework for modeling the land use and travel connection,” The University of North Carolina at Chapel Hill, Ann Arbor, 2007.

[181] J. E. Loebach and J. A. Gilliland, “Free Range Kids? Using GPS-Derived Activity Spaces to Examine Children’s Neighborhood Activity and Mobility,” *Environ. Behav.*, vol. 48, no. 3, pp. 421–453, Apr. 2016.

[182] D. A. Quistberg *et al.*, “The relationship between objectively measured walking and risk of pedestrian-motor vehicle collision,” *Am. J. Epidemiol.*, vol. 185, no. 9, pp. 810–821, 2017.

[183] D. Houston, “Implications of the modifiable areal unit problem for assessing built environment correlates of moderate and vigorous physical activity,” *Appl. Geogr.*, vol. 50, pp. 40–47, Jun. 2014.

[184] K. Larsen, J. Gilliland, P. Hess, P. Tucker, J. Irwin, and M. He, “The Influence of the Physical Environment and Sociodemographic Characteristics on Children’s Mode of Travel to and From School.,” *Am J Public Heal.*, vol. 99, no. 3, pp. 520–526, 2009.

[185] C. Hand, D. Rudman, S. Huot, J. Gilliland, Ma. MA, and R. Pack, “Toward Understanding

Person-Place Transactions in Neighborhoods: A Qualitative-Participatory Geospatial Approach.," *Gerontologist*, vol. 58, no. 1, pp. 89–100, 2018.

[186] K. Larsen, R. N. Buliung, and G. Faulkner, "School travel route measurement and built environment effects in models of children's school travel behavior," *J. Transp. Land Use*, vol. 9, no. 2, pp. 5–23, 2016.

[187] J. Loebach and J. Gilliland, "Neighbourhood play on the endangered list: examining patterns in children's local activity and mobility using GPS monitoring and qualitative GIS," *Child. Geogr.*, vol. 14, no. 5, pp. 573–589, Oct. 2016.

[188] T. van Heeswijck, C. Paquet, Y. Kestens, B. Thierry, C. Morency, and M. Daniel, "Differences in associations between active transportation and built environmental exposures when expressed using different components of individual activity spaces.," *Health Place*, vol. 33, pp. 195–202, May 2015.

[189] B. Kang, A. V Moudon, P. M. Hurvitz, and B. E. Saelens, "Differences in behavior, time, location, and built environment between objectively measured utilitarian and recreational walking," *Transp. Res. PART D-TRANSPORT Environ.*, vol. 57, pp. 185–194, Dec. 2017.

[190] C. Harding, "Activity space geometry and its effect on mode choice," *92nd Annu. Meet. Transp. Res. Board,* 2013.

[191] N. Chen and G. Akar, "Effects of neighborhood types & socio-demographics on activity space," *J. Transp. Geogr.*, vol. 51, pp. 112–121, 2016.

[192] D. McMinn, N. M. Oreskovic, M. J. Aitkenhead, D. W. Johnston, S. Murtagh, and D. A. Rowe, "The physical environment and health-enhancing activity during the school commute: global positioning system, geographical information systems and accelerometry.," *Geospat. Health*, vol. 8, no. 2, pp. 569–572, May 2014.

[193] M. Helbich, "Children's school commuting in the Netherlands: Does it matter how urban form is incorporated in mode choice models?," *Int. J. Sustain. Transp.*, vol. 11, no. 7, pp. 507–517, 2017.

[194] M. Helbich, M. J. Z. van Emmichoven, M. J. Dijst, M.-P. Kwan, F. H. Pierik, and S. I. de Vries, "Natural and built environmental exposures on children's active school travel: A Dutch global positioning system-based cross-sectional study," *Health Place*, vol. 39, pp. 101–109, May 2016.

[195] D. T. Duncan *et al.*, "Walk score, transportation mode choice, and walking among french adults: A GPS, accelerometer, and mobility survey study," *Int. J. Environ. Res. Public Health*, vol. 13, no. 6, pp. 1–14, Jun. 2016.

[196] B. Chaix *et al.*, "A GPS-Based Methodology to Analyze Environment-Health Associations at the Trip Level: Case-Crossover Analyses of Built Environments and Walking," *Am. J. Epidemiol.*, vol. 184, no. 8, pp. 579–589, Oct. 2016.

[197] S. L. Bell, C. Phoenix, R. Lovell, and B. W. Wheeler, "Using GPS and geo-narratives: a methodological approach for understanding and situating everyday green space encounters," *AREA*, vol. 47, no. 1, pp. 88–96, Mar. 2015.

[198] L. Meijering and G. Weitkamp, "Numbers and narratives: Developing a mixed-methods approach to understand mobility in later life," *Soc. Sci. Med.*, vol. 168, no. SI, pp. 200–206, Nov. 2016.

[199] S. Milton *et al.*, "A qualitative geographical information systems approach to explore how

older people over 70 years interact with and define their neighbourhood environment.," *Health Place*, vol. 36, pp. 127–133, Nov. 2015.

[200] S. L. Bell, C. Phoenix, R. Lovell, and B. W. Wheeler, "Seeking everyday wellbeing: The coast as a therapeutic landscape.," *Soc. Sci. Med.*, vol. 142, pp. 56–67, Oct. 2015.

[201] C. Perchoux, Y. Kestens, F. Thomas, A. Van Hulst, B. Thierry, and B. Chaix, "Assessing patterns of spatial behavior in health studies: Their socio-demographic determinants and associations with transportation modes (the RECORD Cohort Study)," *Soc. Sci. Med.*, vol. 119, pp. 64–73, Oct. 2014.

[202] K. Hasanzadeh, A. Broberg, and M. Kytta, "Where is my neighborhood? A dynamic individual-based definition of home ranges and implementation of multiple evaluation criteria," *Appl. Geogr.*, vol. 84, pp. 1–10, Jul. 2017.

[203] B. W. Wheeler, A. R. Cooper, A. S. Page, and R. Jago, "Greenspace and children's physical activity: a GPS/GIS analysis of the PEACH project.," *Prev. Med. (Baltim).*, vol. 51, no. 2, pp. 148–152, Aug. 2010.

[204] K. Villanueva *et al.*, "How far do children travel from their homes? Exploring children's activity spaces in their neighborhood.," *Health Place*, vol. 18, no. 2, pp. 263–273, Mar. 2012.

[205] C. Babb, D. Olaru, C. Curtis, and D. Robertson, "Children's active travel, local activity spaces and wellbeing: A case study in Perth, WA," *Travel Behav. Soc.*, vol. 9, pp. 81–94, Oct. 2017.

[206] D. Vine, L. Buys, and R. Aird, "The use of amenities in high density neighbourhoods by older urban Australian residents," *Landsc. Urban Plan.*, vol. 107, no. 2, pp. 159–171, Aug. 2012.

[207] S. Yoo and D. H. Kim, "Perceived urban neighborhood environment for physical activity of older adults in Seoul, Korea: A multimethod qualitative study," *Prev. Med. (Baltim).*, vol. 103, no. S, pp. S90–S98, Oct. 2017.

[208] G. O'Donovan, I. M. Lee, M. Hamer, and E. Stamatakis, "Association of 'weekend warrior' and other leisure time physical activity patterns with risks for all-cause, cardiovascular disease, and cancer mortality," *JAMA Intern. Med.*, vol. 177, no. 3, pp. 335–342, 2017.

[209] G. O'Donovan, O. L. Sarmiento, and M. Hamer, "The Rise of the 'Weekend Warrior,'" *J. Orthop. Sport. Phys. Ther.*, vol. 48, no. 8, pp. 604–606, Jul. 2018.

[210] M. Petticrew and H. Roberts, *Systematic Reviews in the Social Sciences: A Practical Guide*. 2006.

[211] B. R. Noon, "Radio Tracking and Animal Populations," *Auk*, 2002.

[212] K. B. Rogers and G. C. White, "Analysis of movement and habitat use from telemetry data," *Anal. Interpret. Freshw. Fish. data*, 2007.

[213] G. Aarts, M. MacKenzie, B. McConnell, M. Fedak, and J. Matthiopoulos, "Estimating space-use and habitat preference from wildlife telemetry data," *Ecography*. 2008.

[214] K. A. Phillips, C. R. Elvey, and C. L. Abercrombie, "Applying GPS to the study of primate ecology: A useful tool?," *Am. J. Primatol.*, 1998.

[215] I. A. R. Hulbert and J. French, "The accuracy of GPS for wildlife telemetry and habitat mapping," *J. Appl. Ecol.*, 2001.

[216] M. G. Mey and H. Ter Heide, "Towards spatiotemporal planning: Practicable analysis of day-to-day paths through space and time," *Environ. Plan. B Plan. Des.*, 1997.

[217] D. A. Fennell, "A tourist space-time budget in the Shetland Islands," *Ann. Tour. Res.*, vol. 23, no. 4, pp. 811–829, Oct. 1996.

[218] R. L. Simons, C. H. Burt, A. B. Barr, M. K. Lei, and E. Stewart, "Incorporating routine activities, activity spaces, and situational definitions into the social schematic theory of crime," *Criminology*, 2014.

[219] P.-O. H. Wikström, V. Ceccato, B. Hardie, and K. Treiber, "Activity Fields and the Dynamics of Crime," in *Crime Opportunity Theories*, 2018.

[220] V. Ceccato and P. O. H. Wikström, "Tracking social life and crime," in *The Urban Fabric of Crime and Fear*, 2012.

[221] B. Chaix *et al.*, "An interactive mapping tool to assess individual mobility patterns in neighborhood studies," *Am. J. Prev. Med.*, vol. 43, no. 4, pp. 440–450, 2012.

[222] M. J. Duncan, H. M. Badland, and W. K. Mummery, "Applying GPS to enhance understanding of transport-related physical activity," *Journal of Science and Medicine in Sport*, vol. 12, no. 5. pp. 549–556, Sep-2009.

[223] S. Duncan *et al.*, "Portable global positioning system receivers: Static validity and environmental conditions," *Am. J. Prev. Med.*, 2013.

[224] P. Stopher, C. FitzGerald, and J. Zhang, "Search for a global positioning system device to measure person travel," *Transp. Res. Part C Emerg. Technol.*, 2008.

[225] T. Kos, I. Markezic, and J. Pokrajcic, "Effects of multipath reception on GPS positioning performance," *Elmar, 2010 Proc.*, 2010.

[226] J. Kerr, S. Duncan, and J. Schipperjin, "Using global positioning systems in health research: A practical approach to data collection and processing," *Am. J. Prev. Med.*, vol. 41, no. 5, pp. 532–540, Nov. 2011.

[227] A. Cetateanu, B. A. Luca, A. A. Popescu, A. Page, A. Cooper, and A. Jones, "A novel methodology for identifying environmental exposures using GPS data," *Int. J. Geogr. Inf. Sci.*, 2016.

[228] J. Auld, C. Williams, A. Mohammadian, and P. Nelson, "An automated GPS-based prompted recall survey with learning algorithms," *Transp. Lett.*, 2009.

[229] M. Lin, W. J. Hsu, and Z. Q. Lee, "Detecting modes of transport from unlabelled positioning sensor data," *J. Locat. Based Serv.*, 2013.

[230] G. H. Cho, D. A. Rodríguez, and K. R. Evenson, "Identifying walking trips using GPS data," *Med. Sci. Sports Exerc.*, 2011.

[231] S. Tsui and A. Shalaby, "Enhanced System for Link and Mode Identification for Personal Travel Surveys Based on Global Positioning Systems," *Transp. Res. Rec.*, vol. 1972, pp. 38–45, 2006.

[232] J. Wolf, W. Bachman, M. S. Oliveira, J. Auld, A. (Kouros) Mohammadian, and P. Vovsha, *Applying GPS Data to Understand Travel Behavior, Volume I: Background, Methods, and Tests*. Washington, DC: The National Academies Press, 2014.

[233] P. Stopher, Q. Jiang, and C. FitzGerald, "Processing GPS data from travel surveys," in *28th Australasian Transport Research Forum, ATRF 05*, 2005.

[234] C. Lawson, C. Chen, and H. Gong, "Advanced Applications of Person-based GPS in an Urban Environment," Albany, 2010.

[235] N. Schuessler and K. Axhausen, "Identifying trips and activities and their characteristics from GPS raw data without further information," *Transp. Res. Rec. J. Transp. Res. Board.*, 2008.

[236] S. Kwon *et al.*, "A closer look at the relationship among accelerometer-based physical activity metrics: ICAD pooled data," *Int. J. Behav. Nutr. Phys. Act.*, 2019.

[237] E. K. Howie and L. M. Straker, "Rates of attrition, non-compliance and missingness in randomized controlled trials of child physical activity interventions using accelerometers: A brief methodological review," *Journal of Science and Medicine in Sport*. 2016.

[238] K. R. Evenson, F. Wen, J. S. Metzger, and A. H. Herring, "Physical activity and sedentary behavior patterns using accelerometry from a national sample of United States adults," *Int. J. Behav. Nutr. Phys. Act.*, vol. 12, p. 20, Feb. 2015.

[239] D. D. Polsby and R. Popper, "The Third Criterion: Compactness as a Procedural Safeguard Against Partisan Gerrymandering," *SSRN Electron. J.*, 2017.

[240] C. C. Harris, "A scientific method of districting.," *Behav. Sci.*, 1964.

[241] K. Manaugh and A. M. El-Geneidy, "What makes travel 'local': Defining and understanding local travel behavior," *J. Transp. Land Use*, 2012.

[242] C. G. Victora, J. P. Habicht, and J. Bryce, "Evidence-Based Public Health: Moving Beyond Randomized Trials," *American Journal of Public Health*. 2004.

[243] J. Panter and D. Ogilvie, "Theorising and testing environmental pathways to behaviour change: Natural experimental study of the perception and use of new infrastructure to promote walking and cycling in local communities," *BMJ Open*, 2015.

[244] C. Pope, "Qualitative research in health care: Analysing qualitative data," *BMJ*, 2000.

[245] A. Goodman, "Walking, Cycling and Driving to Work in the English and Welsh 2011 Census: Trends, Socio-Economic Patterning and Relevance to Travel Behaviour in General," *PLoS One*, vol. 8, no. 8, 2013.

[246] J. Panter, S. Griffin, A. Jones, R. Mackett, and D. Ogilvie, "Correlates of time spent walking and cycling to and from work: baseline results from the commuting and health in Cambridge study.," *Int. J. Behav. Nutr. Phys. Act.*, vol. 8, p. 124, 2011.

[247] C. H. D. Jones and D. Ogilvie, "Motivations for active commuting: a qualitative investigation of the period of home or work relocation," *Int. J. Behav. Nutr. Phys. Act.*, 2012.

[248] J. M. Kesten, C. Guell, S. Cohn, and D. Ogilvie, "From the concrete to the intangible: Understanding the diverse experiences and impacts of new transport infrastructure," *Int. J. Behav. Nutr. Phys. Act.*, 2015.

[249] E. Heinen, J. Panter, A. Dalton, A. Jones, and D. Ogilvie, "Sociospatial patterning of the use of new transport infrastructure: Walking, cycling and bus travel on the Cambridgeshire guided busway," *J. Transp. Heal.*, 2014.

[250] A. M. Dalton, A. P. Jones, J. Panter, and D. Ogilvie, "Are GIS-modelled routes a useful proxy for the actual routes followed by commuters?," *J. Transp. Heal.*, vol. 2, no. 2, pp. 219–229, Jul. 2015.

[251] E. Heinen, J. Panter, R. Mackett, and D. Ogilvie, "Changes in mode of travel to work: A natural experimental study of new transport infrastructure," *Int. J. Behav. Nutr. Phys. Act.*, 2015.

[252] J. Panter, E. Heinen, R. Mackett, and D. Ogilvie, "Impact of New Transport Infrastructure on

Walking, Cycling, and Physical Activity," *Am. J. Prev. Med.*, vol. 50, no. 2, pp. e45–e53, 2016.

[253] L. Foley, J. Panter, E. Heinen, R. Prins, and D. Ogilvie, "Changes in active commuting and changes in physical activity in adults: A cohort study," *Int. J. Behav. Nutr. Phys. Act.*, vol. 12, no. 1, 2015.

[254] S. Costa, D. Ogilvie, A. Dalton, K. Westgate, S. Brage, and J. Panter, "Quantifying the physical activity energy expenditure of commuters using a combination of global positioning system and combined heart rate and movement sensors," *Prev. Med. (Baltim).*, 2015.

[255] R. G. Prins, J. Panter, E. Heinen, S. J. Griffin, and D. B. Ogilvie, "Causal pathways linking environmental change with health behaviour change: Natural experimental study of new transport infrastructure and cycling to work," *Prev. Med. (Baltim).*, 2016.

[256] M. D. Fetters, L. A. Curry, and J. W. Creswell, "Achieving integration in mixed methods designs - Principles and practices," *Health Serv. Res.*, 2013.

[257] A. O'Cathain, E. Murphy, and J. Nicholl, "Three techniques for integrating data in mixed methods studies," *BMJ*, 2010.

[258] B. Clark, K. Chatterjee, and S. Melia, "Changes to commute mode: The role of life events, spatial context and environmental attitude," *Transp. Res. Part A Policy Pract.*, vol. 89, pp. 89–105, Jul. 2016.

[259] S. Sahlqvist, A. Goodman, A. R. Cooper, and D. Ogilvie, "Change in active travel and changes in recreational and total physical activity in adults: Longitudinal findings from the iConnect study," *Int. J. Behav. Nutr. Phys. Act.*, 2013.

[260] D. A. Freedman, "Statistical models for causation: What inferential leverage do they provide?," *Eval. Rev.*, 2006.

[261] J. Seawright, "Regression-Based Inference: A Case Study in Failed Causal Assessment," in *Rethinking Social Inquiry*, 2010.

[262] P. C. Meijer, N. Verloop, and D. Beijaard, "Multi-method triangulation in a qualitative study on teachers' practical knowledge: An attempt to increase internal validity," *Qual. Quant.*, 2002.

[263] L. Scott, L. R. Mobley, and D. Il'Yasova, "Geospatial analysis of inflammatory breast cancer and associated community characteristics in the United States," *Int. J. Environ. Res. Public Health*, 2017.

[264] J. X. Moore *et al.*, "Mapping hot spots of breast cancer mortality in the United States: place matters for Blacks and Hispanics," *Cancer Causes Control*, 2018.

[265] Y. Liu *et al.*, "Investigation of space-time clusters and geospatial hot spots for the occurrence of tuberculosis in Beijing," *Int. J. Tuberc. Lung Dis.*, 2012.

[266] S. Agarwal, D. T. Nguyen, L. D. Teeter, and E. A. Graviss, "Spatial-temporal distribution of genotyped tuberculosis cases in a county with active transmission," *BMC Infect. Dis.*, 2017.

[267] J. F. Mosha *et al.*, "Hot spot or not: A comparison of spatial statistical methods to predict prospective malaria infections," *Malar. J.*, 2014.

[268] M. Kwan and J. Lee, "Geovisualization of Human Activity Patterns Using 3D GIS : A Time-Geographic Approach," in *Spatially integrated social science*, 2003.

[269] T. Hägerstrand, "Reflections on 'what about people in regional science?,'" *Pap. Reg. Sci.*,

1989.

[270] H. J. Miller, “Measuring space-time accessibility benefits within transportation networks: Basic theory and computational procedures,” *Geogr. Anal.*, vol. 31, no. 1, pp. 1–26, Jan. 1999.

[271] S. A. Prince, G. P. Butler, D. P. Rao, and W. Thompson, “Where are children and adults physically active and sedentary? – A rapid review of location-based studies,” *Heal. Promot. Chronic Dis. Prev. Canada*, 2019.

[272] M. Sahebkar *et al.*, “Geographical Patterning of Physical Activity Prevalence in Iran: Spatial Analysis of 4 Pooled National Health Surveys Among 119,560 Adults,” *J. Phys. Act. Heal.*, 2019.

[273] R. Mitra, R. N. Buliung, and G. E. J. Faulkner, “Spatial clustering and the temporal mobility of walking school trips in the Greater Toronto Area, Canada.,” *Health Place*, vol. 16, no. 4, pp. 646–655, Jul. 2010.

[274] L. Huang, D. G. Stinchcomb, L. W. Pickle, J. Dill, and D. Berrigan, “Identifying Clusters of Active Transportation Using Spatial Scan Statistics,” *Am. J. Prev. Med.*, 2009.

[275] C. Buck, T. Kneib, T. Tkaczick, K. Konstabel, and I. Pigeot, “Assessing opportunities for physical activity in the built environment of children: interrelation between kernel density and neighborhood scale.,” *Int. J. Health Geogr.*, vol. 14, p. 35, Dec. 2015.

[276] A. I. Robinson, F. Carnes, N. M. Oreskovic, R. A.I., and C. F., “Spatial analysis of crime incidence and adolescent physical activity,” *Prev. Med. (Baltim).*, vol. 85, pp. 74–77, Apr. 2016.

[277] H. J. Miller *et al.*, “Public transit generates new physical activity: Evidence from individual GPS and accelerometer data before and after light rail construction in a neighborhood of Salt Lake City, Utah, USA,” *Heal. Place*, 2015.

[278] B. E. Saelens, A. V. Moudon, B. Kang, P. M. Hurvitz, and C. Zhou, “Relation between higher physical activity and public transit use,” *Am. J. Public Health*, 2014.

[279] M. P. Kwan, “Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: A methodological exploration with a large data set,” *Transp. Res. Part C Emerg. Technol.*, 2000.

[280] J. Chen, S.-L. Shaw, H. Yu, F. Lu, Y. Chai, and Q. Jia, “Exploratory data analysis of activity diary data: a space-time GIS approach,” *J. Transp. Geogr.*, vol. 19, no. 3, SI, pp. 394–404, May 2011.

[281] H. Yu, “Visualizing and analyzing activities in an integrated space-time environment - Temporal geographic information system design and implementation,” *Transp. Res. Rec.*, no. 2024, pp. 54–62, 2007.

[282] A. Valjarević, D. Živković, D. Valjarević, V. Stevanović, and J. Golijanin, “GIS Analysis of Land Cover Changes on the Territory of the Prokuplje Municipality,” *Sci. World J.*, 2014.

[283] A. Geymen and I. Baz, “Monitoring urban growth and detecting land-cover changes on the Istanbul metropolitan area,” in *Environmental Monitoring and Assessment*, 2008.

[284] S. Brage, N. Brage, P. W. Franks, U. Ekelund, and N. J. Wareham, “Reliability and validity of the combined heart rate and movement sensor actiheart,” *Eur. J. Clin. Nutr.*, 2005.

[285] S. Brage *et al.*, “Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity,” *J. Appl. Physiol.*, 2007.

[286] P. S. Royall, R. P. Troiano, M. A. Johnson, H. W. Kohl, and J. E. Fulton, “Appendix 1. Translating Scientific Evidence About Total Amount and Intensity of Physical Activity Into Guidelines,” *2008 Physical Activity Guidelines for Americans. United States Department of Health and Human Services.*, 2008. .

[287] M. Jetté, K. Sidney, and G. Blümchen, “Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity,” *Clin. Cardiol.*, 1990.

[288] T. Lindsay *et al.*, “Descriptive epidemiology of physical activity energy expenditure in UK adults (The Fenland study),” *Int J Behav Nutr Phys Act*, vol. 16, no. 1, p. 126, 2019.

[289] ESRI, “What is raster data?,” 2016. [Online]. Available: https://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/what-is-raster-data.htm.

[290] M. J. Widener, L. M. Minaker, J. L. Reid, Z. Patterson, T. K. Ahmadi, and D. Hammond, “Activity space-based measures of the food environment and their relationships to food purchasing behaviours for young urban adults in Canada,” *Int. J. Middle East Stud.*, 2018.

[291] J. Muenchow, P. Schratz, and A. Brenning, “RQGIS: Integrating R with QGIS for statistical geocomputing,” *R J.*, 2017.

[292] GeoPandas developers, “GeoPandas 0.7.0.” [Online]. Available: https://geopandas.org/#. [Accessed: 02-Jun-2020].

[293] S. Sahlqvist, Y. Song, and D. Ogilvie, “Is active travel associated with greater physical activity? The contribution of commuting and non-commuting active travel to total physical activity in adults,” *Prev. Med. (Baltim).*, 2012.

[294] D. Humphreys, J. Panter, S. Sahlqvist, A. Goodman, and D. Ogilvie, “Changing the environment to improve population health: a framework for considering exposure in natural experimental studies.,” *J Epidemiol Community Heal.*, vol. 70, no. 9, pp. 941–946, 2016.

[295] M. Pratt *et al.*, “Can Population Levels of Physical Activity Be Increased? Global Evidence and Experience,” *Prog. Cardiovasc. Dis.*, 2015.

[296] F. F. Sniehotta, V. Araújo-Soares, J. Brown, M. P. Kelly, S. Michie, and R. West, “Complex systems and individual-level approaches to population health: a false dichotomy?,” *The Lancet Public Health*. 2017.

[297] F. J. M. Mölenberg, J. Panter, A. Burdorf, and F. J. Van Lenthe, “A systematic review of the effect of infrastructural interventions to promote cycling: Strengthening causal inference from observational data,” *International Journal of Behavioral Nutrition and Physical Activity*, vol. 16, no. 1. BioMed Central Ltd., 26-Oct-2019.

[298] D. Ogilvie *et al.*, “Interventions to promote walking: Systematic review,” *Br. Med. J.*, 2007.

[299] L. Yang, S. Sahlqvist, A. McMinn, S. J. Griffin, and D. Ogilvie, “Interventions to promote cycling: Systematic review,” *BMJ (Online)*. 2010.

[300] B. Giles-Corti and R. J. Donovan, “The relative influence of individual, social and physical environment determinants of physical activity.,” *Soc. Sci. Med.*, vol. 54, no. 12, pp. 1793–1812, 2002.

[301] D. Ogilvie *et al.*, “Using natural experimental studies to guide public health action: turning the evidence-based medicine paradigm on its head,” *J. Epidemiol. Community Health*, 2019.

[302] L. Foley, D. Dumuid, A. J. Atkin, T. Olds, and D. Ogilvie, “Patterns of health behaviour

associated with active travel: A compositional data analysis," *Int. J. Behav. Nutr. Phys. Act.*, 2018.

[303] S. S. Intille, J. Lester, J. F. Sallis, and G. Duncan, "New Horizons in sensor development," *Med. Sci. Sports Exerc.*, 2012.

[304] J. Bort-Roig, N. D. Gilson, A. Puig-Ribera, R. S. Contreras, and S. G. Trost, "Measuring and influencing physical activity with smartphone technology: A systematic review," *Sports Medicine*. 2014.

[305] R. E. Rice and J. E. Katz, "Comparing internet and mobile phone usage: Digital divides of usage, adoption, and dropouts," *Telecomm. Policy*, 2003.

[306] M. Pratt *et al.*, "The implications of megatrends in information and communication technology and transportation for changes in global physical activity," *The Lancet*. 2012.

[307] G. Wilson, J. Bryan, K. Cranston, J. Kitzes, L. Nederbragt, and T. K. Teal, "Good enough practices in scientific computing," *PLoS Comput. Biol.*, 2017.

[308] S. Harper, "A future for observational epidemiology: Clarity, credibility, transparency," *American Journal of Epidemiology*. 2019.

[309] D. S. Procter *et al.*, "An open-source tool to identify active travel from hip-worn accelerometer, GPS and GIS data," *Int. J. Behav. Nutr. Phys. Act.*, 2018.

[310] K. Lee and M. P. Kwan, "Physical activity classification in free-living conditions using smartphone accelerometer data and exploration of predicted results," *Comput. Environ. Urban Syst.*, 2018.

[311] R. Brondeel, B. Pannier, and B. Chaix, "Using GPS, GIS, and Accelerometer Data to Predict Transportation Modes," *Med. Sci. Sports Exerc.*, vol. 47, no. 12, pp. 2669–2675, Dec. 2015.

[312] T. Feng and H. J. P. Timmermans, "Transportation mode recognition using GPS and accelerometer data," *Transp. Res. Part C Emerg. Technol.*, 2013.

APPENDIX

A. SUPPLEMENTAL MATERIAL FOR CHAPTER 2

CHARACTERISTICS OF THE ENVIRONMENT AND PHYSICAL ACTIVITY IN MIDLIFE: FINDINGS FROM UK BIOBANK

A1. ADDITIONAL RESULTS

Table A.1: Adjusted cross-sectional associations between environmental characteristics and physical activity outcomes (Model 1)

		Recorded measures			Reported measures					
		Mean acceleration	MVPA		MVPA		Total walking		Walking for pleasure	
		β (95% CI)	Middle tertile RRR (95% CI)	Upper tertile RRR (95% CI)	Middle tertile RRR (95% CI)	Upper tertile RRR (95% CI)	Middle tertile RRR (95% CI)	Upper tertile RRR (95% CI)	Middle tertile RRR (95% CI)	Upper tertile RRR (95% CI)
Spaces for PA	**Facilities for PA** (ref: none)	†	n.s	†	*n.i*	*n.i*	*	*	n.s	*
	One or more	0.19 (0.05, 0.33)	1.04 (1.00, 1.09)	1.06 (1.01, 1.11)			1.03 (1.01, 1.05)	1.03 (1.01, 1.05)	1.01 (0.99, 1.03)	1.03 (1.01, 1.05)
	Parks (ref: none)	*n.i*	n.s	n.s	*n.i*	*n.i*	n.s	n.s	†	n.s
	One or more		0.99 (0.95, 1.05)	1.01 (0.96, 1.06)			1.00 (0.98, 1.03)	0.99 (0.97, 1.02)	1.02 (1.00, 1.04)	1.01 (0.99, 1.03)
Walk-ability	**Walkability** (ref: lowest)	**	**	**	**	**	**	**	**	n.s
	Q2	0.04 (-0.15, 0.22)	1.01 (0.96, 1.07)	1.03 (0.98, 1.10)	1.01 (0.98, 1.03)	1.03 (1.01, 1.06)	1.01 (0.98, 1.03)	1.04 (1.02, 1.07)	0.98 (0.95, 1.00)	0.98 (0.95, 1.00)
	Q3	0.12 (-0.08, 0.32)	1.06 (1.00, 1.13)	1.08 (1.01, 1.15)	1.02 (1.00, 1.05)	1.07 (1.04, 1.10)	1.05 (1.02, 1.08)	1.07 (1.04, 1.10)	0.97 (0.94, 1.00)	0.96 (0.93, 0.99)
	Q4	0.45 (0.23, 0.67)	1.15 (1.07, 1.23)	1.28 (1.20, 1.38)	1.09 (1.06, 1.12)	1.12 (1.09, 1.15)	1.16 (1.13, 1.20)	1.14 (1.10, 1.17)	1.07 (1.04, 1.11)	1.03 (1.00, 1.06)
Disturbance	**NO_X** (ref: lowest)	**	†	**	**	**	n.s	*	*	**
	Highest	-0.57 (-0.84, -0.30)	0.92 (0.85, 1.00)	0.86 (0.79, 0.94)	0.93 (0.89, 0.96)	0.84 (0.81, 0.88)	1.00 (0.97, 1.04)	0.93 (0.90, 0.97)	0.93 (0.90, 0.97)	0.85 (0.81, 0.88)
	Noise pollution (ref: lowest)	*n.i*	*n.i*	*n.i*	*n.i*	*n.i*	n.s	†	n.s	n.s
	Highest						1.01 (0.99, 1.03)	1.02 (1.00, 1.04)	1.00 (0.98, 1.01)	1.02 (1.00, 1.03)
	Distance to major road (ref: closest)	*n.i*	*n.i*	*n.i*	*n.i*	*n.i*	n.s	n.s	*n.i*	*n.i*
	Furthest						1.00 (0.99, 1.01)	0.99 (0.98, 1.00)		
Natural environment	**Terrain** (ref: mean slope <3°)	*n.i*	*n.i*	*n.i*	*n.i*	*n.i*	n.s	†	**	**
	Mean slope ≥3°						0.99 (0.98, 1.01)	1.02 (1.01, 1.04)	1.05 (1.04, 1.07)	1.08 (1.06, 1.10)
	Greenness (ref: least)	*n.i*	*n.i*	*n.i*	n.s	**	**	**	n.s	**
	Q2				1.04 (1.01, 1.06)	1.10 (1.07, 1.12)	1.03 (1.01, 1.06)	1.06 (1.03, 1.08)	1.01 (0.99, 1.04)	1.02 (1.00, 1.05)
	Q3				1.02 (1.00, 1.04)	1.09 (1.06, 1.11)	1.05 (1.02, 1.07)	1.06 (1.04, 1.09)	1.01 (0.99, 1.04)	1.02 (1.00, 1.05)
	Q4				1.00 (0.98, 1.03)	1.08 (1.05, 1.10)	1.09 (1.06, 1.12)	1.14 (1.11, 1.17)	1.00 (0.98, 1.03)	1.08 (1.05, 1.11)
Sociodemographic	**Urban-rural status** (ref: urban)	**	n.s	**	**	**	n.s	**	**	**
	Fringe	0.30 (0.05, 0.54)	0.98 (0.91, 1.06)	1.07 (0.99, 1.15)	1.05 (1.01, 1.08)	1.08 (1.05, 1.12)	1.04 (1.00, 1.07)	1.09 (1.05, 1.13)	1.16 (1.12, 1.21)	1.26 (1.22, 1.30)
	Rural	0.83 (0.53, 1.12)	1.00 (0.92, 1.10)	1.18 (1.07, 1.30)	1.09 (1.04, 1.14)	1.21 (1.16, 1.27)	1.01 (0.97, 1.06)	1.12 (1.07, 1.17)	1.11 (1.06, 1.17)	1.25 (1.20, 1.31)
	Area-level deprivation (ref: least deprived)	**	**	**	**	†	†	**	**	**
	Q2	-0.06 (-0.24, 0.13)	0.96 (0.90, 1.01)	0.96 (0.90, 1.01)	1.00 (0.97, 1.03)	1.00 (0.97, 1.03)	1.03 (1.00, 1.05)	1.06 (1.03, 1.09)	0.98 (0.95, 1.00)	0.99 (0.96, 1.01)
	Q3	-0.16 (-0.36, 0.03)	0.90 (0.85, 0.95)	0.90 (0.85, 0.96)	0.98 (0.95, 1.01)	1.01 (0.98, 1.04)	1.02 (0.99, 1.05)	1.07 (1.04, 1.09)	0.91 (0.88, 0.93)	0.90 (0.88, 0.92)
	Q4	-0.42 (-0.63, -0.22)	0.86 (0.81, 0.92)	0.87 (0.81, 0.92)	0.95 (0.92, 0.97)	1.00 (0.97, 1.02)	1.03 (1.00, 1.06)	1.09 (1.06, 1.12)	0.85 (0.83, 0.88)	0.82 (0.80, 0.84)
	Most deprived	-0.82 (-1.06, -0.59)	0.80 (0.74, 0.86)	0.80 (0.74, 0.86)	0.89 (0.86, 0.92)	0.96 (0.93, 0.99)	1.03 (1.00, 1.06)	1.09 (1.06, 1.13)	0.77 (0.74, 0.79)	0.74 (0.72, 0.76)

Model adjusted for age, sex, ethnicity, education, income, car ownership, assessment centre, housing tenure, employment status, children in household, urban-rural status, area-level deprivation plus significant environmental characteristics from univariate analyses (Model 0). Walkability components have been substituted for walkability summary score
**p<0.001 *p<0.01 †p<0.05 indicates test for trend. β – regression coefficient; RRR – relative risk ratio; CI – confidence interval; *n.i* – not included in model

Table A.2: Adjusted cross-sectional associations between environmental characteristics, including walkability components as separate variables, and physical activity outcomes

		Recorded measures			Reported measures					
		Mean acceleration	MVPA		MVPA		Total walking		Walking for pleasure	
		β (95% CI)	Middle tertile RRR (95% CI)	Upper tertile RRR (95% CI)	Middle tertile RRR (95% CI)	Upper tertile RRR (95% CI)	Middle tertile RRR (95% CI)	Upper tertile RRR (95% CI)	Middle tertile RRR (95% CI)	Upper tertile RRR (95% CI)
Spaces for PA	**Facilities for PA** (ref: none)	†	n.s	†	*n.i*	*n.i*	*	†	n.s	n.s
	One or more	0.16 (0.01, 0.30)	1.04 (0.99, 1.08)	1.05 (1.01, 1.10)			1.03 (1.01, 1.05)	1.03 (1.01, 1.05)	1.00 (0.98, 1.02)	1.02 (1.00, 1.04)
	Parks (ref: none)	*n.i*	n.s	n.s	*n.i*	*n.i*	n.s	n.s	n.s	n.s
	One or more		0.99 (0.94, 1.04)	1.00 (0.95, 1.06)			1.00 (0.98, 1.03)	0.99 (0.97, 1.01)	1.02 (0.99, 1.04)	1.00 (0.98, 1.03)
Walkability	**Street connectivity** (ref: least)	*	n.s	**	*	†	n.s	n.s	**	*
	Q2	0.06 (-0.12, 0.24)	1.03 (0.97, 1.09)	1.06 (1.00, 1.12)	1.02 (1.00, 1.05)	1.03 (1.00, 1.05)	1.02 (0.99, 1.04)	1.01 (0.99, 1.03)	1.01 (0.98, 1.03)	1.00 (0.98, 1.03)
	Q3	0.11 (-0.07, 0.29)	1.04 (0.98, 1.10)	1.10 (1.03, 1.16)	1.03 (1.00, 1.06)	1.06 (1.03, 1.09)	1.01 (0.98, 1.04)	1.01 (0.99, 1.04)	1.03 (1.00, 1.05)	1.03 (1.00, 1.05)
	Q4	0.32 (0.13, 0.52)	1.06 (0.99, 1.14)	1.19 (1.11, 1.27)	1.04 (1.01, 1.07)	1.02 (0.99, 1.05)	1.02 (0.99, 1.05)	0.98 (0.95, 1.01)	1.08 (1.05, 1.12)	1.05 (1.02, 1.08)
	Dwelling density (ref: lowest)	*n.i*	n.s	n.s	n.s	n.s	**	*	**	**
	Q2		1.03 (0.97, 1.09)	0.94 (0.88, 1.00)	0.97 (0.94, 0.99)	0.94 (0.91, 0.96)	1.05 (1.02, 1.07)	1.01 (0.98, 1.03)	0.99 (0.96, 1.01)	0.94 (0.91, 0.96)
	Q3		1.00 (0.93, 1.07)	0.89 (0.83, 0.95)	0.97 (0.94, 1.00)	0.95 (0.92, 0.98)	1.06 (1.03, 1.09)	1.04 (1.01, 1.07)	0.93 (0.90, 0.96)	0.87 (0.84, 0.89)
	Q4		1.08 (1.00, 1.17)	1.03 (0.95, 1.11)	1.02 (0.98, 1.05)	0.99 (0.96, 1.03)	1.15 (1.12, 1.19)	1.05 (1.02, 1.09)	0.94 (0.91, 0.98)	0.85 (0.83, 0.88)
	Land use mix (ref: lowest)	**	*	**	*	**		**	**	**
	Q2	0.12 (-0.06, 0.30)	1.01 (0.96, 1.07)	1.02 (0.96, 1.08)	1.01 (0.98, 1.03)	1.03 (1.01, 1.06)	1.02 (1.00, 1.05)	1.04 (1.02, 1.07)	1.03 (1.00, 1.05)	1.03 (1.01, 1.06)
	Q3	0.21 (0.03, 0.40)	1.05 (0.99, 1.11)	1.04 (0.98, 1.10)	1.00 (0.98, 1.03)	1.03 (1.00, 1.05)	1.03 (1.00, 1.05)	1.07 (1.04, 1.10)	1.03 (1.00, 1.05)	1.05 (1.02, 1.08)
	Q4	0.34 (0.15, 0.52)	1.08 (1.02, 1.15)	1.11 (1.05, 1.18)	1.03 (1.01, 1.06)	1.08 (1.05, 1.11)	1.10 (1.07, 1.13)	1.16 (1.13, 1.19)	1.08 (1.05, 1.11)	1.14 (1.11, 1.17)
Disturbance	**NO$_x$** (ref: lowest)	**	†	**	**	**	n.s	†	†	**
	Highest	-0.55 (-0.82, -0.28)	0.92 (0.85, 1.00)	0.88 (0.81, 0.96)	0.93 (0.90, 0.97)	0.86 (0.83, 0.90)	1.00 (0.96, 1.04)	0.95 (0.91, 0.99)	0.95 (0.91, 0.99)	0.88 (0.85, 0.91)
	Noise pollution (ref: lowest)	*n.i*	*n.i*	*n.i*	*n.i*	*n.i*	n.s	*	n.s	n.s
	Highest						1.01 (1.00, 1.03)	1.03 (1.01, 1.05)	0.99 (0.97, 1.01)	1.00 (0.99, 1.02)
	Distance to major road (ref: closest)	*n.i*	*n.i*	*n.i*	*n.i*	*n.i*	n.s	†	*n.i*	*n.i*
	Furthest						1.00 (0.99, 1.01)	0.99 (0.98, 1.00)		
Natural environment	**Terrain (**ref: mean slope <3°)	***n.i***	*n.i*	*n.i*	*n.i*	*n.i*	**n.s**	**n.s**	**	**
	Mean slope ≥3°						1.00 (0.98, 1.02)	1.02 (1.00, 1.04)	1.04 (1.03, 1.06)	1.06 (1.04, 1.08)
	Greenness (ref: least)	*n.i*	*n.i*	*n.i*	n.s	**	**	**	n.s	**
	Q2				1.04 (1.01, 1.06)	1.10 (1.07, 1.12)	1.04 (1.01, 1.06)	1.06 (1.03, 1.09)	1.01 (0.99, 1.04)	1.02 (1.00, 1.05)
	Q3				1.02 (0.99, 1.04)	1.09 (1.06, 1.11)	1.04 (1.02, 1.07)	1.07 (1.05, 1.10)	1.02 (0.99, 1.04)	1.04 (1.02, 1.07)
	Q4				1.00 (0.98, 1.03)	1.08 (1.06, 1.11)	1.09 (1.06, 1.12)	1.15 (1.12, 1.18)	1.01 (0.99, 1.04)	1.10 (1.07, 1.13)

Sociodemographic	**Urban-rural status (ref: urban)**	**	**n.s**	*	**	**	**n.s**	**	**	**
	Fringe	0.25 (0.01, 0.50)	0.97 (0.90, 1.05)	1.03 (0.95, 1.12)	1.04 (1.00, 1.07)	1.05 (1.02, 1.09)	1.05 (1.01, 1.08)	1.07 (1.03, 1.11)	1.14 (1.10, 1.18)	1.20 (1.16, 1.24)
	Rural	0.76 (0.46, 1.07)	1.00 (0.91, 1.10)	1.12 (1.01, 1.24)	1.07 (1.02, 1.12)	1.16 (1.11, 1.21)	1.02 (0.97, 1.06)	1.08 (1.03, 1.13)	1.08 (1.03, 1.14)	1.16 (1.11, 1.21)
	Area-level deprivation (ref: least)	**	**	**	**	n.s	n.s	**	**	**
	Q2	-0.06 (-0.25, 0.12)	0.96 (0.90, 1.01)	0.96 (0.90, 1.02)	1.00 (0.97, 1.03)	1.00 (0.98, 1.03)	1.02 (0.99, 1.05)	1.06 (1.03, 1.08)	0.98 (0.95, 1.01)	0.99 (0.97, 1.02)
	Q3	-0.17 (-0.36, 0.03)	0.90 (0.84, 0.95)	0.91 (0.86, 0.97)	0.98 (0.96, 1.01)	1.02 (0.99, 1.05)	1.01 (0.98, 1.04)	1.06 (1.03, 1.09)	0.91 (0.89, 0.94)	0.91 (0.89, 0.94)
	Q4	-0.41 (-0.61, -0.21)	0.86 (0.80, 0.92)	0.87 (0.82, 0.93)	0.95 (0.92, 0.97)	1.01 (0.98, 1.03)	1.02 (0.99, 1.05)	1.09 (1.06, 1.12)	0.86 (0.84, 0.89)	0.84 (0.82, 0.86)
	Q5	-0.78 (-1.02, -0.55)	0.80 (0.74, 0.86)	0.80 (0.74, 0.87)	0.89 (0.86, 0.92)	0.97 (0.94, 1.01)	1.02 (0.98, 1.05)	1.10 (1.06, 1.14)	0.78 (0.76, 0.81)	0.77 (0.74, 0.79)

Model adjusted for age, sex, ethnicity, education, income, car ownership, assessment center, housing tenure, employment status, children in household, urban-rural status, area-level deprivation plus significant environmental characteristics from univariate analyses (Model 0).

**p<0.001 *p<0.01 †p<0.05 indicates test for trend. β – regression coefficient; RRR – relative risk ratio; CI – confidence interval; *n.i* – not included in model

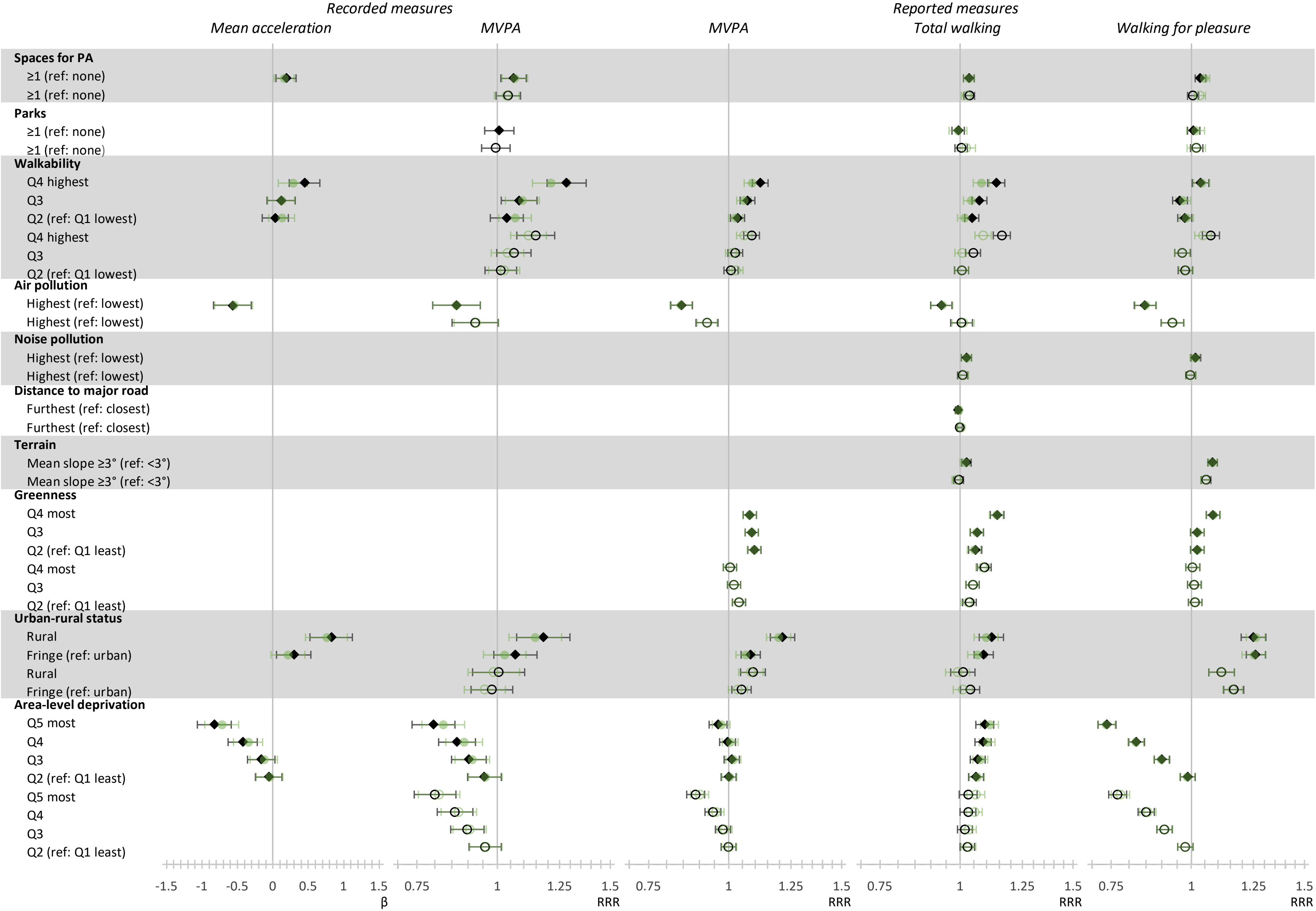


Figure A.1 Adjusted associations between environmental characteristics and activity outcomes

Outcome variables: ■ Continuous data; ◆ Upper tertile; ○ Middle tertile; ⊢⊣ 95% Confidence interval. Results of original analyses (Model 1: 1 km neighbourhood measures) shown in black; Results of sensitivity analyses (0.5 km neighbourhood measures) shown in green. White space is where variables have not been included in adjusted model
β = regression coefficient presented on linear scale; RRR = relative risk ratio presented on log scale; MVPA = moderate-to-vigorous physical activity

APPENDIX

B. SUPPLEMENTAL MATERIAL FOR CHAPTER 3

ACTIVITY SPACES IN STUDIES OF THE ENVIRONMENT AND PHYSICAL ACTIVITY: A REVIEW AND SYNTHESIS OF IMPLICATIONS FOR CAUSALITY

B1. ADDITIONAL RESULTS

LIST OF INCLUDED STUDIES

[112] A. G. Rundle *et al.*, "Using GPS Data to Study Neighborhood Walkability and Physical Activity.," *Am. J. Prev. Med.*, vol. 50, no. 3, pp. e65–e72, Mar. 2016.

[159] S. N. Zenk *et al.*, "Activity space environment and dietary and physical activity behaviors: a pilot study.," *Health Place*, vol. 17, no. 5, pp. 1150–1161, Sep. 2011.

[163] M. Sanchez *et al.*, "Predictors of daily mobility of adults in peri-urban south India," *Int. J. Environ. Res. Public Health*, vol. 14, no. 7, p. 783, Jul. 2017.

[164] M. Kamruzzaman and J. Hine, "Analysis of rural activity spaces and transport disadvantage using a multi-method approach," *Transp. POLICY*, vol. 19, no. 1, pp. 105–120, Jan. 2012.

[165] D. A. Rodriguez *et al.*, "Influence of the Built Environment on Pedestrian Route Choices of Adolescent Girls," *Environ. Behav.*, vol. 47, no. 4, pp. 359–394, May 2015.

[166] D. A. Rodríguez *et al.*, "Out and about: Association of the built environment with physical activity behaviors of adolescent females," *Health Place*, vol. 18, no. 1, pp. 55–62, 2012.

[167] J. A. Hirsch, M. Winters, M. C. Ashe, P. J. Clarke, and H. A. McKay, "Destinations That Older Adults Experience Within Their GPS Activity Spaces: Relation to Objectively Measured Physical Activity," *Environ. Behav.*, vol. 48, no. 1, SI, pp. 55–77, Jan. 2016.

[168] J. A. Hirsch, M. Winters, P. Clarke, and H. McKay, "Generating GPS activity spaces that shed light upon the mobility habits of older adults: a descriptive analysis.," *Int. J. Health Geogr.*, vol. 13, p. 51, Dec. 2014.

[169] T. Franke, M. Winters, H. McKay, and H. Chaudhury, "A grounded visualization approach to explore sociospatial and temporal complexities of older adults' mobility," *Soc. Sci. Med.*, vol. 193, pp. 59–69, 2017.

[170] G. Vich, O. Marquet, and C. Miralles-Guasch, "Suburban commuting and activity spaces: using smartphone tracking data to understand the spatial extent of travel behaviour," *Geogr. J.*, vol. 183, no. 4, pp. 426–439, Dec. 2017.

[171] M. Kamruzzaman, J. Hine, B. Gunay, and N. Blair, "Using GIS to visualise and evaluate student travel behaviour," *J. Transp. Geogr.*, vol. 19, no. 1, pp. 13–32, Jan. 2011.

[172] P. A. Plazier, G. Weitkamp, and A. E. van den Berg, "'Cycling was never so easy!' An analysis of e-bike commuters' motives, travel behaviour and experiences using GPS-tracking and interviews," *J. Transp. Geogr.*, vol. 65, pp. 25–34, Dec. 2017.

[173] N. Colabianchi, C. J. Coulton, J. D. Hibbert, S. M. McClure, C. E. Ievers-Landis, and E. M. Davis, "Adolescent self-defined neighborhoods and activity spaces: spatial overlap and relations to physical activity and obesity.," *Health Place*, vol. 27, pp. 22–29, May 2014.

[174] B. Chaix, D. Duncan, J. Vallee, A. Vernez-Moudon, T. Benmarhnia, and Y. Kestens, "The 'Residential' Effect Fallacy in Neighborhood and Health Studies Formal Definition, Empirical Identification, and Correction," *EPIDEMIOLOGY*, vol. 28, no. 6, pp. 789–797, Nov. 2017.

[175] S. Kosaka, M. Umezaki, M. Ishikawa, and C. Watanabe, "Physical activity and the neighborhood environment in a heavy snowfall area in Japan: The role of 'Gangi-dori,'" *Landsc. Urban Plan.*, vol. 123, pp. 124–133, Mar. 2014.

[176] C. P. Tribby, H. J. Miller, B. B. Brown, K. R. Smith, and C. M. Werner, "Geographic regions for assessing built environmental correlates with walking trips: A comparison using different metrics and model designs," *Heal. Place*, vol. 45, pp. 1–9, 2017.

[177] T. Burgoine, A. P. Jones, R. J. Namenek Brouwer, and S. E. Benjamin Neelon, "Associations between

BMI and home, school and route environmental exposures estimated using GPS and GIS: do we see evidence of selective daily mobility bias in children?," *Int. J. Health Geogr.*, vol. 14, p. 8, Feb. 2015.

[178] N. Chen, C.-H. Wang, and G. Akar, "Geographically Weighted Regression Approach to Investigate Spatial Variations in Activity Space," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2671, pp. 40–50, Jan. 2017.

[179] N. C. Lee, C. Voss, A. D. Frazer, J. A. Hirsch, H. A. McKay, and M. Winters, "Does Activity Space Size Influence Physical Activity Levels of Adolescents? - A GPS study of an urban environment.," *Prev. Med. reports*, vol. 3, pp. 75–78, 2016.

[180] Y. Fan, "The built environment, activity space, and time allocation: An activity-based framework for modeling the land use and travel connection," The University of North Carolina at Chapel Hill, Ann Arbor, 2007.

[181] J. E. Loebach and J. A. Gilliland, "Free Range Kids? Using GPS-Derived Activity Spaces to Examine Children's Neighborhood Activity and Mobility," *Environ. Behav.*, vol. 48, no. 3, pp. 421–453, Apr. 2016.

[182] D. A. Quistberg *et al.*, "The relationship between objectively measured walking and risk of pedestrian-motor vehicle collision," *Am. J. Epidemiol.*, vol. 185, no. 9, pp. 810–821, 2017.

[183] D. Houston, "Implications of the modifiable areal unit problem for assessing built environment correlates of moderate and vigorous physical activity," *Appl. Geogr.*, vol. 50, pp. 40–47, Jun. 2014.

[184] K. Larsen, J. Gilliland, P. Hess, P. Tucker, J. Irwin, and M. He, "The Influence of the Physical Environment and Sociodemographic Characteristics on Children's Mode of Travel to and From School.," *Am J Public Heal.*, vol. 99, no. 3, pp. 520–526, 2009.

[185] C. Hand, D. Rudman, S. Huot, J. Gilliland, Ma. MA, and R. Pack, "Toward Understanding Person-Place Transactions in Neighborhoods: A Qualitative-Participatory Geospatial Approach.," *Gerontologist*, vol. 58, no. 1, pp. 89–100, 2018.

[186] K. Larsen, R. N. Buliung, and G. Faulkner, "School travel route measurement and built environment effects in models of children's school travel behavior," *J. Transp. Land Use*, vol. 9, no. 2, pp. 5–23, 2016.

[187] J. Loebach and J. Gilliland, "Neighbourhood play on the endangered list: examining patterns in children's local activity and mobility using GPS monitoring and qualitative GIS," *Child. Geogr.*, vol. 14, no. 5, pp. 573–589, Oct. 2016.

[188] T. van Heeswijck, C. Paquet, Y. Kestens, B. Thierry, C. Morency, and M. Daniel, "Differences in associations between active transportation and built environmental exposures when expressed using different components of individual activity spaces.," *Health Place*, vol. 33, pp. 195–202, May 2015.

[189] B. Kang, A. V Moudon, P. M. Hurvitz, and B. E. Saelens, "Differences in behavior, time, location, and built environment between objectively measured utilitarian and recreational walking," *Transp. Res. PART D-TRANSPORT Environ.*, vol. 57, pp. 185–194, Dec. 2017.

[190] C. Harding, "Activity space geometry and its effect on mode choice," *92nd Annu. Meet. Transp. Res. Board,* 2013.

[191] N. Chen and G. Akar, "Effects of neighborhood types & socio-demographics on activity space," *J. Transp. Geogr.*, vol. 51, pp. 112–121, 2016.

[192] D. McMinn, N. M. Oreskovic, M. J. Aitkenhead, D. W. Johnston, S. Murtagh, and D. A. Rowe, "The physical environment and health-enhancing activity during the school commute: global positioning system, geographical information systems and accelerometry.," *Geospat. Health*, vol. 8, no. 2, pp. 569–572, May 2014.

[193] M. Helbich, "Children's school commuting in the Netherlands: Does it matter how urban form is incorporated in mode choice models?," *Int. J. Sustain. Transp.*, vol. 11, no. 7, pp. 507–517, 2017.

[194] M. Helbich, M. J. Z. van Emmichoven, M. J. Dijst, M.-P. Kwan, F. H. Pierik, and S. I. de Vries, "Natural

and built environmental exposures on children's active school travel: A Dutch global positioning system-based cross-sectional study," *Health Place*, vol. 39, pp. 101–109, May 2016.

[195] D. T. Duncan *et al.*, "Walk score, transportation mode choice, and walking among french adults: A GPS, accelerometer, and mobility survey study," *Int. J. Environ. Res. Public Health*, vol. 13, no. 6, pp. 1–14, Jun. 2016.

[196] B. Chaix *et al.*, "A GPS-Based Methodology to Analyze Environment-Health Associations at the Trip Level: Case-Crossover Analyses of Built Environments and Walking," *Am. J. Epidemiol.*, vol. 184, no. 8, pp. 579–589, Oct. 2016.

[197] S. L. Bell, C. Phoenix, R. Lovell, and B. W. Wheeler, "Using GPS and geo-narratives: a methodological approach for understanding and situating everyday green space encounters," *AREA*, vol. 47, no. 1, pp. 88–96, Mar. 2015.

[198] L. Meijering and G. Weitkamp, "Numbers and narratives: Developing a mixed-methods approach to understand mobility in later life," *Soc. Sci. Med.*, vol. 168, no. SI, pp. 200–206, Nov. 2016.

[199] S. Milton *et al.*, "A qualitative geographical information systems approach to explore how older people over 70 years interact with and define their neighbourhood environment.," *Health Place*, vol. 36, pp. 127–133, Nov. 2015.

[200] S. L. Bell, C. Phoenix, R. Lovell, and B. W. Wheeler, "Seeking everyday wellbeing: The coast as a therapeutic landscape.," *Soc. Sci. Med.*, vol. 142, pp. 56–67, Oct. 2015.

[201] C. Perchoux, Y. Kestens, F. Thomas, A. Van Hulst, B. Thierry, and B. Chaix, "Assessing patterns of spatial behavior in health studies: Their socio-demographic determinants and associations with transportation modes (the RECORD Cohort Study)," *Soc. Sci. Med.*, vol. 119, pp. 64–73, Oct. 2014.

[202] K. Hasanzadeh, A. Broberg, and M. Kytta, "Where is my neighborhood? A dynamic individual-based definition of home ranges and implementation of multiple evaluation criteria," *Appl. Geogr.*, vol. 84, pp. 1–10, Jul. 2017.

[203] B. W. Wheeler, A. R. Cooper, A. S. Page, and R. Jago, "Greenspace and children's physical activity: a GPS/GIS analysis of the PEACH project.," *Prev. Med. (Baltim).*, vol. 51, no. 2, pp. 148–152, Aug. 2010.

[204] K. Villanueva *et al.*, "How far do children travel from their homes? Exploring children's activity spaces in their neighborhood.," *Health Place*, vol. 18, no. 2, pp. 263–273, Mar. 2012.

[205] C. Babb, D. Olaru, C. Curtis, and D. Robertson, "Children's active travel, local activity spaces and wellbeing: A case study in Perth, WA," *Travel Behav. Soc.*, vol. 9, pp. 81–94, Oct. 2017.

[206] D. Vine, L. Buys, and R. Aird, "The use of amenities in high density neighbourhoods by older urban Australian residents," *Landsc. Urban Plan.*, vol. 107, no. 2, pp. 159–171, Aug. 2012.

[207] S. Yoo and D. H. Kim, "Perceived urban neighborhood environment for physical activity of older adults in Seoul, Korea: A multimethod qualitative study," *Prev. Med. (Baltim).*, vol. 103, no. S, pp. S90–S98, Oct. 2017.

Table B.1: Characteristics of included studies

Study reference	Population	Specific research question(s)	Types of environmental features assessed[b]	Measures of activity space	Assessment of physical activity	Assessment of health
Studies which investigated features of an activity space						
[163]	Adults	1) What are the individual, external, and contextual predictors of mobility dimensions (attributes of the activity space)?	Street connectivity Land use types and mix	Shape Size		
[164][a]	Adults	1) Are there patterns of transport disadvantage in different rural settings identified through the activity space concept? 2) Can findings be validated with qualitative evidence from disadvantaged groups?	Street connectivity Land use types and mix	Size Shape		
[168]	Older adults	1) Is the size/compactness of activity spaces associated with sociodemographics and available resources?	Destinations Provision for walking and cycling	Shape Size		
[170]	Adults	1) What personal and environmental factors affect the spatial extent of daily mobility?	Street connectivity	Size	Travel mode	
[171]	Adults	1) How does home location and socioeconomic characteristics affect the size of the activity space? 2) How does the activity space vary between those who live near a demand responsive transport route and those who do not?	Intervention Street connectivity	Size		
[178]	Adults	1) How do the effects of the built environment on activity spaces vary spatially across the study region?	Street connectivity Land use types and mix	Size		
[179]	Adolescents	1) Is the size of activity spaces associated with MVPA?		Size	MVPA	
[180]	All	1) What effect do the built environment, traffic conditions, and weather conditions have in individual travel behaviour?	Street connectivity Natural environment	Size		
[181]	Children	1) How do individual, perceptual, or environmental factors affect the size or time children spend in neighbourhood activity space?	Street connectivity Land use types and mix	Size	Active travel	
[190]	Adults	1) What is the relationship between activity space size and travel mode?		Size Shape	Travel mode	
[191]	Adults	1) What are the relationships between socio-demographic characteristics, travel time, the built environment and resulting average activity spaces for all activities and non-work activities separately?	Land use types and mix	Size		

[202]	Adults	1) Is size of the home range (as an indicator of physical activity) associated with perceived health?		Size		Perceived health
[204]	Children	1) What are the built environmental, social–cultural and individual factors influencing the size of a child's activity space?	Destinations Provision for walking and cycling Natural environment Land use types and mix	Size		
[205][a]	Children	1) How do realised and potential activity spaces compare? 2) What features do children want/not want to use locally?		Size	Travel mode	
Studies which investigate features within an activity space						
[112]	Adults	1) What are the size and characteristics of neighbourhoods utilised and not utilised? 2) Does walkability influence the size of the utilised area? 3) Is walkability associated with physical activity in the activity space and the neighbourhood?	Provision for walking and cycling		MPA	
[159]	All	1) Are features within the activity space associated with physical activity and diet?	Destinations Natural environment		MVPA	
[165]	Adolescents	1) Which environmental characteristics are associated with a higher likelihood of choosing a walking route? 2) How does this compare across two locations (urban/rural)?	Destinations Provision for walking and cycling Aesthetics and safety		Walking	
[166]	Adolescents	1) What are the associations between the outdoor built environment and MVPA? 2) How does this compare for weekends and weekdays and from one year to the next?	Destinations		LPA MVPA	
[167]	Older adults	1) Are destinations associated with physical activity?	Destinations		Total physical activity	
[169][a]	Older adults	1) Where are older adults active and why? 2) How does this change over time?	Provision for walking and cycling		Steps	
[172][a]	Adults	1) Under what conditions can e-bikes substitute motorised commuting? 2) Which role do travel experiences play in the daily commute by e-bike?	Provision for walking and cycling Street connectivity Aesthetics and safety		Cycling	
[174]	Adults	1) Is residential accessibility to services associated with walking? 2) Is there residential effect fallacy and confounding and can it be corrected for by using non-residential accessibility measures?	Destinations		Walking	

[175]	Adults	1) What is the role of the gangi-dori (covered walkway) for physical activity?	Intervention	Steps LPA MVPA	
[177]	Children	1) Are home, school and journey exposures to food, physical activity and built environments associated with BMI?	Destinations Provision for walking and cycling Street connectivity Aesthetics and safety Natural environment		BMI
[182]	Adults	1) Is walking associated with pedestrian collision risk of a collision?	Street connectivity	Walking	
[183]	Adults	1) What is the relationship between built environment factors on walking routes and MVPA?	Provision for walking and cycling Natural environment Land use types and mix	Walking MVPA	
[184]	Children	1) Are characteristics of the physical and social environment associated with the journey to and from school?	Provision for walking and cycling Street connectivity Natural environment Land use types and mix	Travel mode	
[185][a]	Older adults	1) How do older adults connect within and with their neighbourhoods?	Destinations	Walking	
[186]	Children	2) Are characteristics of the built environment on the school commutes related to travel mode?	Provision for walking and cycling Street connectivity Aesthetics and safety Natural environment	Walking	
[187]	Adults	1) What characteristics are associated with transport related physical activity?	Destinations Street connectivity Natural environment Land use types and mix	Active travel	
[189]	Adults	1) How do recreational and utilitarian walking behaviours differ spatially and temporally?	Destinations Street connectivity Natural environment Land use types and mix	Walking	

[192]	Children	1) Are physical environmental factors on the school commute associated with physical activity?	Provision for walking and cycling Street connectivity Natural environment	MVPA	
[193]	Children	1) Are physical environmental factors on the school commute associated with walking, cycling, and being chauffeured?	Street connectivity Aesthetics and safety Land use types and mix	Travel mode	
[194]	Children	1) Are the natural and built environmental features on the school commute associates with active travel?	Destinations Street connectivity Aesthetics and safety Natural environment Land use types and mix	Active travel	
[195]	Adults	1) What are the associations between walkability, transportation mode choice, and walking?	Provision for walking and cycling	Walking Steps	
[196]	Adults	1) What are the associations between built environments and walking at the trip level? 2) How do trip-level influences differ from residential environmental influences?	Destinations Natural environment	Walking Steps	
[197][a]	Adults	1) How are greenspace experiences shaped by everyday individual agency, life circumstances, and past place experiences?	Natural environment		Wellbeing
[198][a]	Older adults	1) What do everyday mobility practices look like for older adults?	Destinations	Travel mode Trips	Wellbeing
[199][a]	Older adults	1) How do older adults interact with and define their neighbourhood? With social factors such as friends and family, community activities, places or facilities?	Destinations Street connectivity		Perceived health Wellbeing
[200][a]	Adults	1) How do wellbeing experiences relate to different green and blue space interactions, life circumstances and transitions, and personal identities?	Natural environment		Wellbeing
[203]	Children	1) Is children's physical activity associated with greenspace, outdoors in non-greenspace, and indoors? 2) Is MVPA more likely in greenspace than non-greenspace after school?	Natural environment	MVPA	

[206][a]	Older adults	1) How do participants experience the built environment and what factors facilitate or inhibit physical activity?	Destinations Provision for walking and cycling Street connectivity Land use types and mix		Physical activity Walking	
[207][a]	Older adults	1) What environments encourage physical activity for older adults?	Provision for walking and cycling Street connectivity Aesthetics and safety		Physical activity Trips	
Studies which investigate both features of and features within an activity space						
[173]	Adolescents	1) Is the size of the self-defined neighbourhood associated with physical activity and weight status? 2) Are the facilities within the self-defined neighbourhood associated with physical activity and weight status? 3) Do physical activity spaces overlap with neighbourhoods?	Destinations	Size	Total physical activity	BMI
[176]	Adults	1) Does a change in the neighbourhood environment (complete streets intervention) influence residents' walking trips, self-defined neighbourhoods, and walking activity spaces?	Intervention Destinations Provision for walking and cycling Street connectivity Aesthetics and safety	Shape Size	Walking	
[181][a]	Children	1) What are the patterns in children's primary activities and settings, independent mobility levels, and perception and use of neighbourhood affordances?	Destinations	Size	Play	
[201]	Adults	1) Which characteristics are related to spatial behaviour/dimensions? 2) Does spatial behaviour/dimensions relate to transport modes?	Destinations	Size Shape	Travel mode	

LPA = light physical activity, M(V)PA = moderate (to vigorous) physical activity, BMI = body mass index

[a]Studies include a qualitative assessment

[b]Typical features included in broad groups: **Destinations** = healthcare, community, food outlets, parks, schools, physical activity facilities, **Provision for walking and cycling** = walk score, footpath provision, cycle path provision, **Street connectivity =** connectivity, road or intersection density, public transport, **Aesthetics and safety** = aesthetics, road safety, crime safety, **Natural environment** = greenspace, bluespace, trees, slope, **Land use types and mix** = recreational, institutional, residential, commercial, urbanicity, land use mix

APPENDIX

C. SUPPLEMENTAL MATERIAL FOR CHAPTER 4
GPS DATA CLEANING AND PREPARATION

C1. ADDITIONAL METHODS

EXAMPLES OF FUNCTIONS WRITTEN IN PYTHON TO PROCESS GPS DATA

Example Python function

Demonstrates logic and syntax required in a basic Python function and how to call the function and use it on data in ArcGIS

```
def ExampleFunction(var_a, var_b):
  i = var_a + var_b
  return i
```

Define function with a name and number of required parameters in brackets

Parameters used inside each function to perform an operation

Value is returned for each row of data

```
ExampleFunction(!a!, !b!)
```

Use ArcGIS Field Calculator tool to call function using its name and populate rows of data field c

Information passed into function must match number of required parameters. Field names surrounded by exclamation marks

a	b	c
2	1	3
3	2	5
4	3	7

PopulateIf Function

Populates field based on an input value
Useful for labelling week and weekend days and jumps in the data

```
def PopulateIf(timeDiff):
  i = 0
  if timeDiff > '00:02:00' :
    i = 1
  if timeDiff > '01:00:00' :
    i = 2
  return i
```

Create a default value inside function. This will be created for each row of data

Update value using 'if' statement based on information passed into function

PreviousRow function

Populates the input row of data with a value from the previous row
Useful for calculating spatial and temporal differences between points and identifying jumps in the data

```
prev_val = None
def SequentialDif(curr_val):
  global prev_val
  if (prev_val is None):
    prev_val = curr_val
  new = prev_val
  prev_val = curr_val
  return new
```

Create default value outside function. This will be created once and updated inside the function for each row of data

First row of data will have no value, it is therefore given the same value from the input row of data

The value from the input row is stored to be used as the previous row value in the next iteration of the function on the following row of data

The value from the previous row is returned

CreateIndex Function

Creates a new value each time the input value is different to the previous row of data
Useful for creating index numbers for days of wear

```
prevDate = None
i = 1
def createIndex(day):
  global prevDate
  global i
  if prevDate is None:
    prevDate = day
  if prevDate == day:
    i = i
  else:
    i = i+1
  prevDate = day
  return i
```

First row of data will have no value, it is therefore given the same value from the input row of data

If values are the same as the previous row, i remains the same. If the value differs from the previous row, an incremental value is added to i

The previous row value is updated based on the present row of data and used in the subsequent iteration of the function on the next row

SegmentTotal function

Sums the total number of points in each spatial segment (where a spatial segment represents consecutive points without a spatial jump in the data)
Useful for identifying clusters of signal stray and calculating wear time

The ID of the spatial segment the input row is in is provided as a parameter

An array of ID numbers for each spatial segment of points is created for the whole file. E.g. 1111222222222222333333333

```
filename = "%Name%"
listOfPointIds = arcpy.da.TableToNumPyArray(filename, 'SpatialSegmentID')
sumOfIds = 0
def countPoints(id):
  global sumOfIds
  sumOfIds = sum(listOfPointIds[listOfPointIds['SpatialSegmentID'] == id]['SpatialSegmentID'])
  countOfPoints = sumOfIds/id
  return  countOfPoints
```

The sum of IDs is divided by the ID number itself to calculate the total number of points with the input ID.

The numbers in the array that match the input parameter are summed. Using the example array, if id = 1, all 1s in the array will be summed and a value of 4 returned

APPENDIX

D. SUPPLEMENTAL MATERIAL FOR CHAPTER 5
GPS DATA CLEANING AND PREPARATION

D1. ADDITIONAL RESULTS

RESULTS FOR CHANGES IN ACTIVITY SPACE SIZE FOR WEEKDAY AND WEEKEND DATA

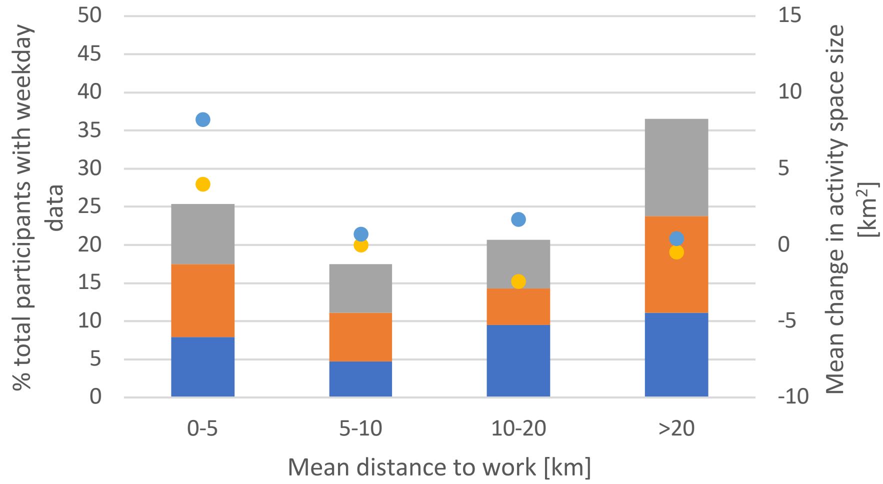


Figure D.1: Change in weekday activity space size by mean distance to work between study phases

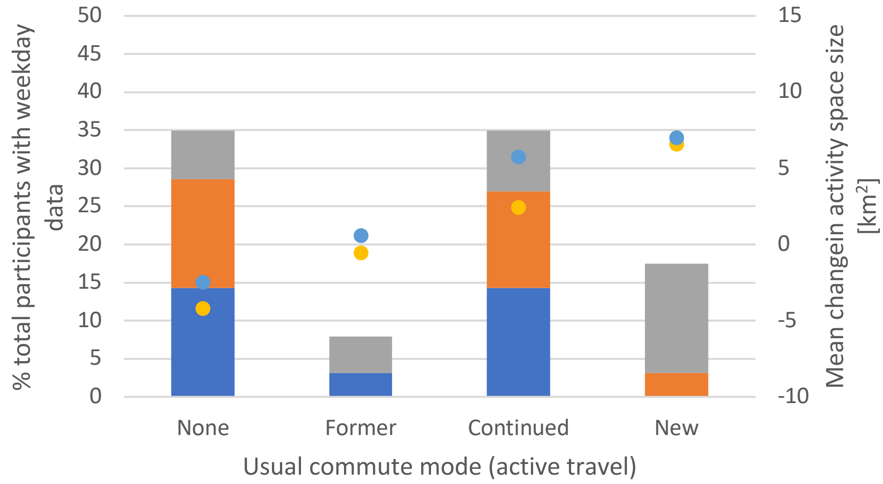


Figure D.2: Change in weekday activity space size by whether participants actively commuted

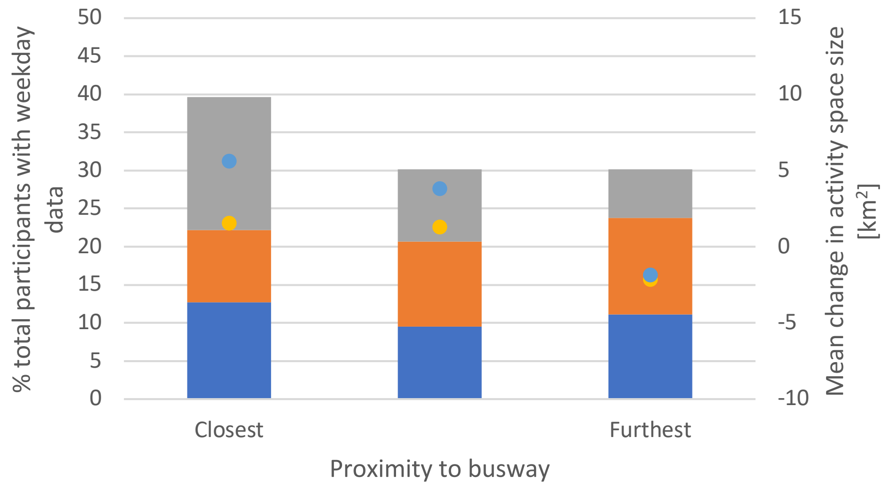


Figure D.3: Change in weekday activity space size by mean proximity from home address to busway between phases 2 and 4

Change in activity space size: Decrease No change Increase

Mean change in activity space size: All sample Non-movers only

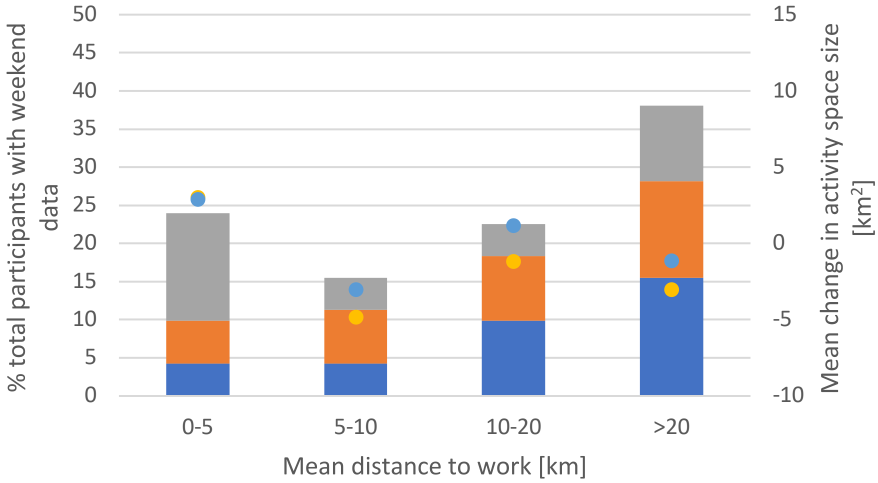


Figure D.4: Change in weekend activity space size by mean distance to work between study phases

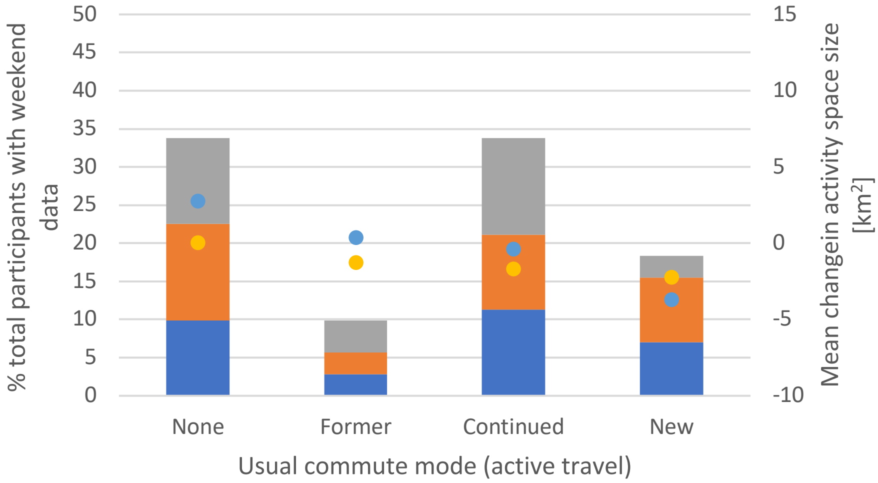


Figure D.5: Change in weekend activity space size by whether participants actively commuted

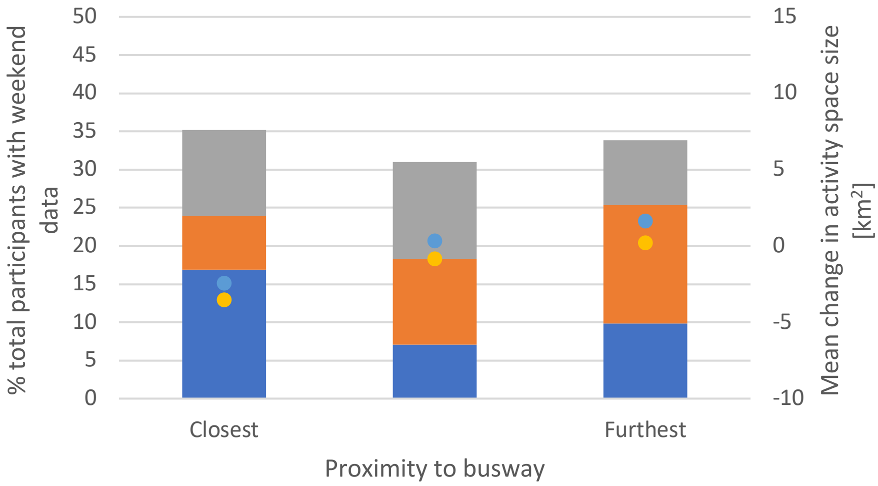


Figure D.6: Change in weekend activity space size by mean proximity from home address to busway between phases 2 and 4

Change in activity space size: Decrease No change Increase
Mean change in activity space size: All sample Non-movers only

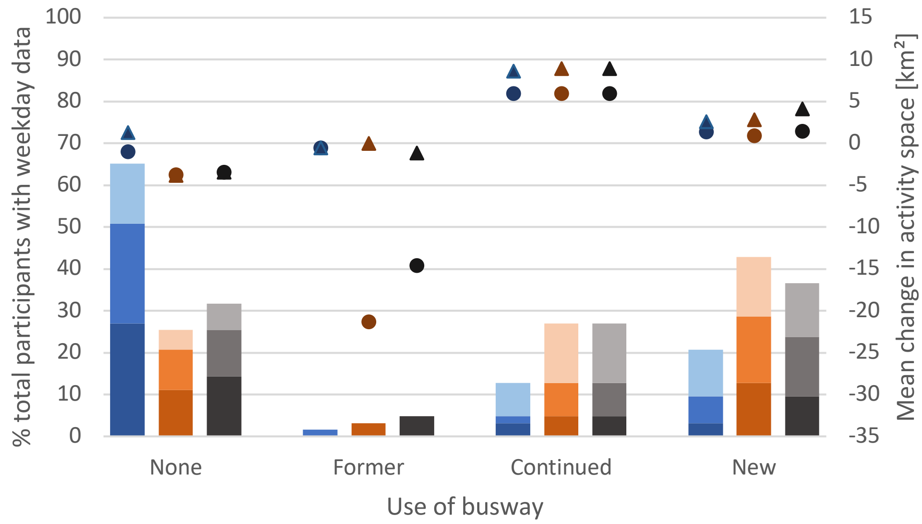


Figure D.7: Change in weekday activity space size by use of busway

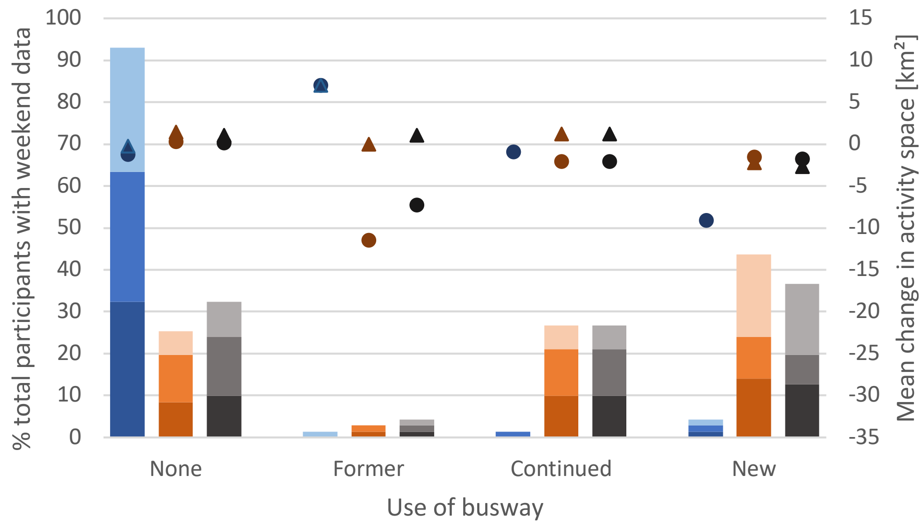


Figure D.8: Change in weekend activity space size by use of busway

GPS measure of use:	Decrease	No change	Increase
Self-reported use:	Decrease	No change	Increase
Self-reported walk or cycle:	Decrease	No change	increase
Mean change in activity space size:	○ All sample	△ Non-movers only	

Table D.1: Adjusted associations between sociodemographic and geographic characteristics and exposure to the busway with change in activity space size (for non-movers only)

	Week		Weekday		Weekend	
	Decrease RRR (95% CI)	Increase RRR (95% CI)	Decrease RRR (95% CI)	Increase RRR (95% CI)	Decrease RRR (95% CI)	Increase RRR (95% CI)
Urban rural status (ref: urban)	*n.i*	*n.i*	*n.i*	*n.i*		*
Rural					0.28 (0.05, 1.55)	**0.14 (0.03, 0.75)**
Proximity to busway (ref: closest)				*		
[square root of mean distance]	0.81 (0.45, 1.47)	0.84 (0.49, 1.42)	0.77 (0.43, 1.36)	**0.51 (0.27, 0.96)**	0.63 (0.33, 1.20)	0.83 (0.46, 1.50)

Model adjusted for age, sex, and significant variables from univariate analyses.
Bold text indicates statistical significance (**p<0.001 *p<0.01 †p<0.05)
RRR – relative risk ratio; CI – confidence interval; n.i – not included in model

RESULTS FOR CHANGES IN ACTIVITY SPACE SHAPE FOR WEEKDAY AND WEEKEND DATA

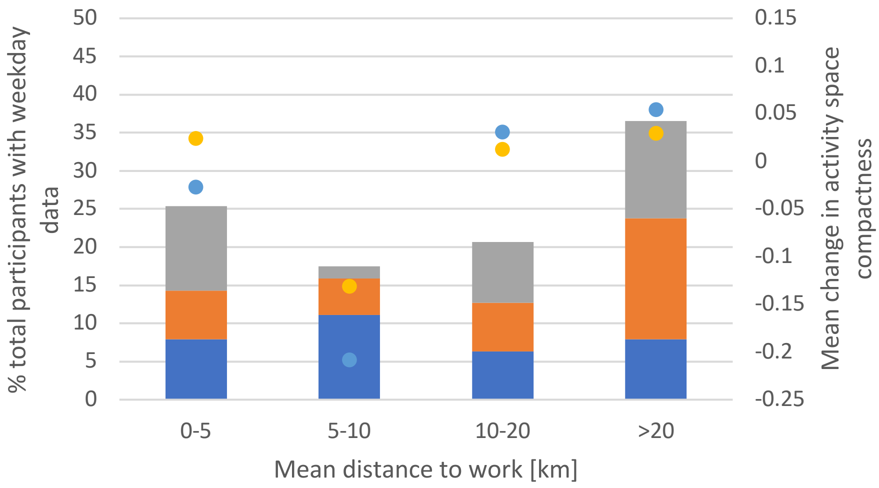


Figure D.9: Change in weekday activity space shape by mean distance to work between study phases

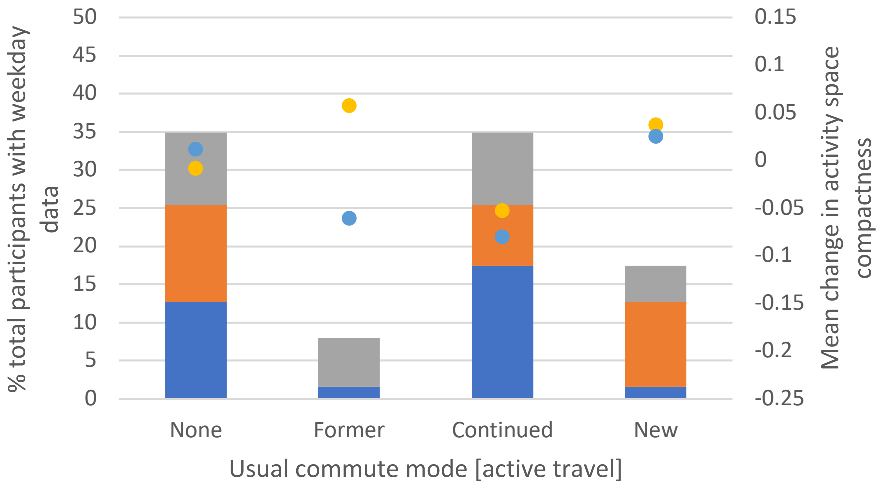


Figure D.10: Change in weekday activity space shape by whether participants actively commuted

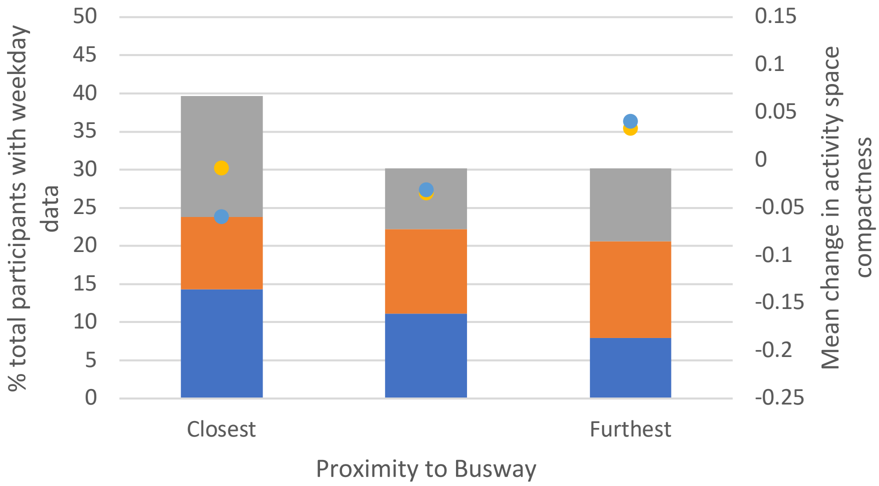


Figure D.11: Change in weekday activity space shape by mean proximity from home address to busway between phases 2 and 4

Change in activity space size: Decrease No change Increase

Mean change in activity space size: All sample Non-movers only

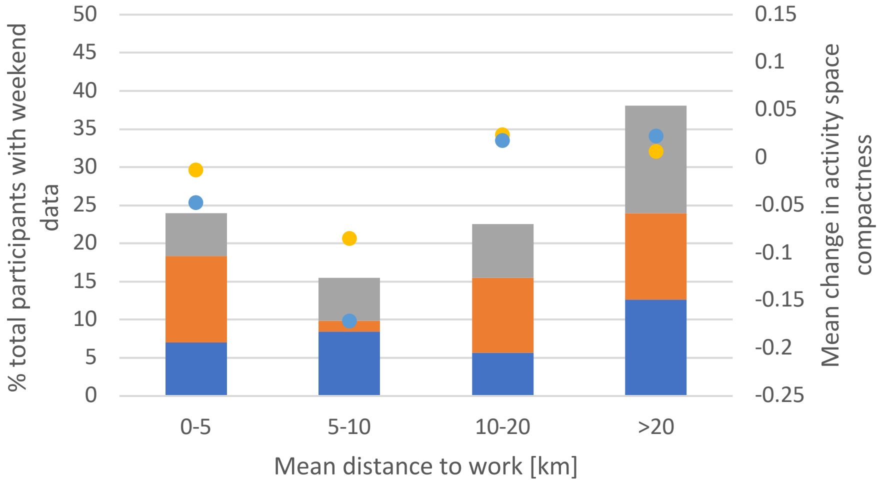


Figure D.12: Change in weekend activity space shape by mean distance to work between study phases

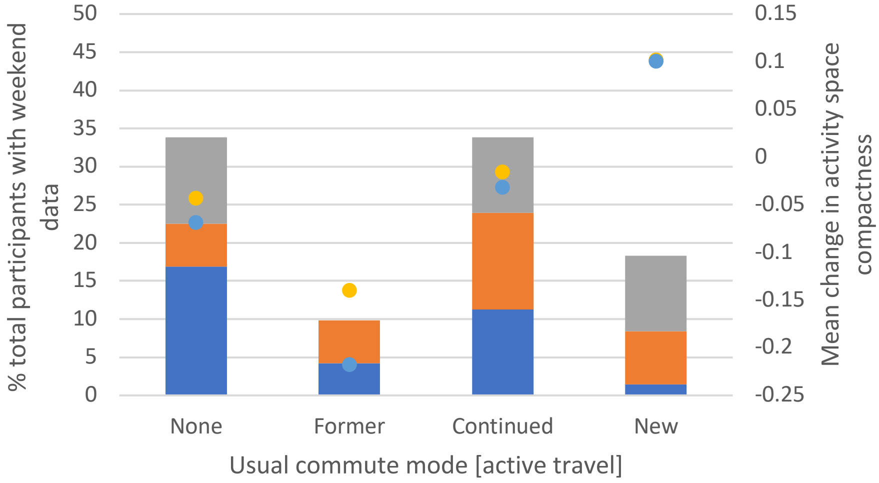


Figure D.13: Change in weekend activity space shape by whether participants actively commuted

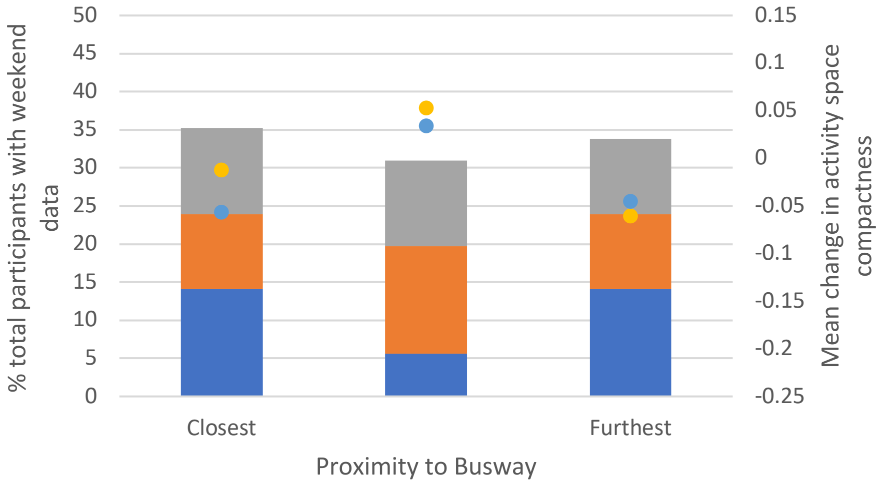


Figure D.14: Change in weekend activity space shape by mean proximity from home address to busway between phases 2 and 4

Change in activity space size: Decrease No change Increase
Mean change in activity space size: All sample Non-movers only

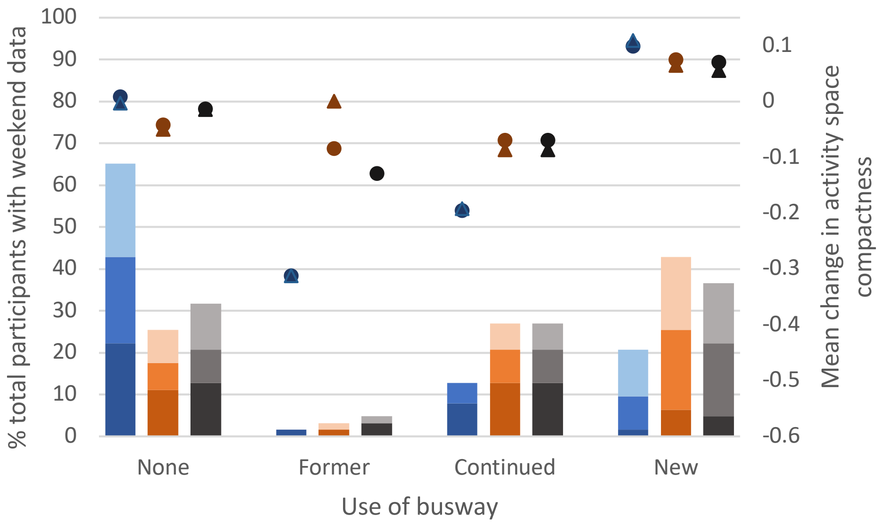


Figure D.15: Change in weekday activity space shape by use of busway

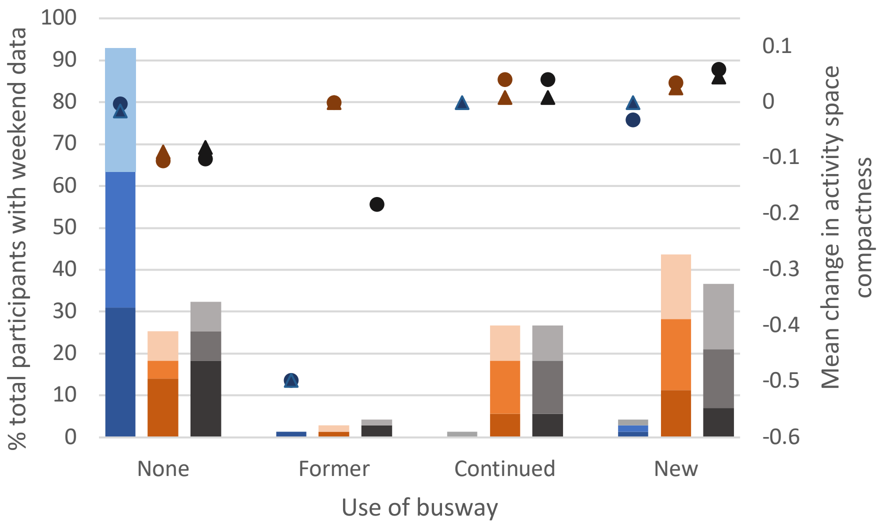


Figure D.16: Change in weekend activity space shape by use of busway

GPS measure of use:	Decrease	No change	Increase
Self-reported use:	Decrease	No change	Increase
Self-reported walk or cycle:	Decrease	No change	increase
Mean change in activity space size:	○ All sample	△ Non-movers only	

Table D.2: Adjusted associations between sociodemographic and geographic characteristics and exposure to the busway with change in activity space shape (for non-movers only)

	Week		Weekday		Weekend	
	Less compact RRR (95% CI)	More compact RRR (95% CI)	Less compact RRR (95% CI)	More compact RRR (95% CI)	Less compact RRR (95% CI)	More compact RRR (95% CI)
Car ownership (ref: none)	*n.i*	*n.i*			*n.i*	*n.*
One or more cars			0.36 (0.07, 1.78)	1.72 (0.30, 9.98)		
Proximity to busway (ref: closest)						
[square root of mean distance]	1.01 (0.58, 1.79)	1.05 (0.58, 1.89)	0.69 (0.38, 1.25)	0.79 (0.43, 1.44)	0.90 (0.53, 1.51)	1.05 (0.61, 1.80)

Model adjusted for age, sex, and significant variables from univariate analyses.
Bold text indicates statistical significance (**p<0.001 *p<0.01 †p<0.05)
RRR – relative risk ratio; CI – confidence interval; n.i – not included in model