Evaluating the impact of urban green space interventions on physical activity and other wellbeing behaviours: developing and implementing improved natural experimental methods

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Biology, Medicine and Health.

2020

JACK S. BENTON

Division of Psychology and Mental Health

School of Health Sciences

Contents

Word count: 47,233

List of Tables

List of Figures

List of Appendices

Abbreviations

AIC	Akaike's information criterion
CC	Cormac Coulter
CI	Confidence interval
DF	David French
ESRC	Economic and Social Research Council
GHIA	Green infrastructure and the Health and wellbeing Influences on an Ageing population
GIS	Geographic information system
GM	Greater Manchester
GPS	Global positioning system
ICC	Intraclass correlation coefficient
IJBNPA	International Journal of Behavioral Nutrition and Physical Activity
IMD	Index of multiple deprivation
IQR	Interquartile range
IRR	Incidence rate ratio
JA	Jamie Anderson
JB	Jack Benton
LSOA	Lower layer super output area
MOHAWk	Method for Observing pHysical Activity and Wellbeing
MP	Margaret Pulis
MRC	Medical Research Council
NDVI	Normalised difference vegetation index
NEF	New Economics Foundation
NIHR	National Institute for Health Research
RCT	Randomised controlled trial
SD	Standard deviation
SK	Sophie King
SOPARC	System for Observing Play and Recreation in Communities
TIDieR	Template for Intervention Description and Replication
TREND	Transparent Reporting of Evaluations with Nonrandomized Designs

UK	United Kingdom
US	United States
WHO	World Health Organization
VAT	Value added tax
VM	Vanessa Macintyre

Thesis Abstract

Evidence suggests that interventions to improve or create new urban green spaces have positive effects on physical activity and other behaviours important for wellbeing. However, the evidence base is scarce, especially in Europe, and often of poor quality. Environmental interventions are rarely amenable to randomisation, so a natural experimental study is the optimal approach to evaluate the causal effects of urban green space interventions. My recent review of twelve natural experimental studies of environmental interventions on physical activity found major methodological weaknesses causing high risk of bias and identified eight recommendations for future research. This PhD aimed to: (1) address these recommendations by developing new methods to improve the internal validity of natural experimental studies; (2) develop and validate a new systematic observation tool for unobtrusively assessing physical activity and two other wellbeing behaviours in urban spaces; (3) implement these new methods in two natural experimental studies of urban green space interventions.

Three studies combined to develop and validate MOHAWk (Method for Observing pHysical Activity and Wellbeing): a systematic observation tool for assessing physical activity and two other evidence-based wellbeing behaviours (social interactions and taking notice of the environment) in urban spaces. From 156 hours of observations using six observers across five urban spaces, evidence is provided that MOHAWk is reliable and valid.

New natural experimental methods were developed and implemented in two studies of urban green space interventions. Key methodological improvements included a new process for systematically identifying matched comparison sites; appropriate adjustment for confounders to minimise the risk of confounding; publication of study protocols with a priori analyses specified (reporting any deviations); sample size calculations; process measures; and clear reporting in line with standardised checklists.

The first natural experimental study assessed the impact of low-cost changes (e.g. tree planting) to four urban amenity green spaces on older adults' and adults' wellbeing behaviours in Greater Manchester, UK. There was no evidence that the interventions increased observable wellbeing behaviours or green space use. A nested qualitative study suggested the interventions were not substantial enough to be noticed compared to other recent neighbourhood changes and therefore unlikely to influence behaviour change.

The second natural experimental study assessed the impact of new walking infrastructure and green space improvements along an urban canal on canal usage and wellbeing behaviours among adults in Greater Manchester, UK. There was evidence that the intervention significantly increased the total number of people using the canal path compared to the comparison sites at all follow-ups. There was some evidence that the intervention brought about increases in walking and vigorous physical activity, social interactions, and people taking notice of the environment. A process evaluation suggested that there was some displacement of activity, but the intervention also encouraged existing users to use the canal more often.

These natural experimental studies provide exemplars of how to use methods with substantially lower risk of bias than previous research. MOHAWk is an unobtrusive and inexpensive outcome measure that will enable more robust natural experimental studies, particularly in Europe where there is a dearth of evidence. More robust natural experimental studies like these are now needed to better inform policy and practice recommendations on the (in)effectiveness of a wider range of urban green space interventions. There is a need for better theory to understand how urban green space interventions bring about their effects, especially focusing on the physical and social contextual factors that influence whether interventions are likely to work or not.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

Funding

This PhD was funded by the Economic and Social Research Council as part of the North-West Social Science Doctoral Training Partnership (ES/P000665/1). Funding was also received from the University of Manchester President's Doctoral Scholarship Award. The first natural experimental study was carried out as part of the Green infrastructure and the Health and wellbeing Influences on an Ageing population (GHIA) project. Funders: Natural Environment Research Council, the Arts and Humanities Research Council and the Economic and Social Research Council under the Valuing Nature Programme (NE/N013530/1). The funders had no role in the conceptualisation, design, data collection, analysis, publication, or preparation of this thesis. The views expressed in this thesis are those of the authors and not necessarily those of the funders or wider project partners.

The Author

The author of this thesis has a BSc (Hons) 1st Class degree in Psychology from the University of Sheffield. The author also has an MSc (Distinction) in Clinical and Health Psychology from the University of Manchester. The author has worked as a Research Assistant at the University of Manchester on a variety of projects. Most recently, and most relevant to this thesis, the author worked as part of the Green infrastructure and the Health and wellbeing Influences on an Ageing population (GHIA) project, which was set up to investigate how urban green space influences the health and wellbeing of older adults. It enabled researchers from a diverse range of academic specialisms to work with a variety of partners and organisations in Greater Manchester, including Manchester City Council (MCC), Red Rose Forest (City of Trees), The Canal and River Trust, Manchester Climate Change, Greater Manchester Centre for Voluntary Organisation (GMCVO) and Public Health Manchester.

Acknowledgements

I would like to thank my very supportive supervisors: Prof David French, Dr Jamie Anderson and Dr Sarah Cotterill. Thank you for giving me this opportunity in the first place, and for all the time and effort you have given to providing feedback and encouragement throughout my PhD. I feel incredibly lucky to have had such a knowledgeable team, and you have made this a very enjoyable and personally enriching experience. I would like to say a special thanks to David, as my primary supervisor. I would not be where I am today without your time, guidance, advice and support.

I thank Manchester City Council (Dave Barlow, Barbara Drummond and Matt Doran), Southway Housing Trust (Philippa Reece) and Peel Land & Property Trust (Joanne Holden, Charlotte Culshaw and Louise Morrissey). This PhD research would not have been possible without the support of these organisations.

I acknowledge the financial support provided through the Economic and Social Research Council as part of the North-West Social Science Doctoral Training Partnership, for which I am very grateful. I would also like to thank the University of Manchester President's Doctoral Scholarship Award for further financial support.

I would like to thank my fellow PhD students who have shared this journey with me, including Dr Laura McGowan, Hannah Long, Michael Loughran, Stephanie Lyons, Stephanie Mace, Lauren Rockcliffe, and others – thank you for your friendship, advice and motivation. Also, thank you to the wider Manchester Centre for Health Psychology team who have provided a supportive environment.

Finally, I would like to offer special thanks to my wonderful family and friends for your support, even if none of you really understood what I was doing! In particular, I would like to thank my girlfriend Laurie and dog Claude for providing vital support and distraction away from PhD work.

Rationale for submitting the thesis in journal format

It was envisaged from the start that the studies conducted during this PhD would be disseminated in peer-reviewed journals. Given the progress made in the publication of research papers, the journal format was deemed to be most appropriate.

List of publications derived from this PhD

Published

Benton JS, Anderson J, Cotterill S, Dennis M, Lindley SJ, French DP. Evaluating the impact of improvements in urban green space on older adults' physical activity and wellbeing: protocol for a natural experimental study. BMC Public Health. 2018;18(1):923.

Benton JS, Anderson J, Cotterill S, Dennis M, French DP. The impact of new walking infrastructure and changes to green space along an urban canal on physical activity and wellbeing: protocol for a natural experimental study. Open Sci Framew. 2018; https://osf.io/zcm7v/

Benton JS, Anderson J, Pulis M, Cotterill S, Hunter RF, French DP. Method for Observing pHysical Activity and Wellbeing (MOHAWk): validation of an observation tool to assess physical activity and other wellbeing behaviours in urban spaces. Cities & Health. 2020.

Under peer review

Benton JS, Cotterill S, Anderson J, Macintyre VG, Gittins M, Dennis M, Lindley SJ, French DP. Impact of a low-cost urban green space intervention on wellbeing behaviours in older adults: a natural experimental study. Wellbeing, Space & Society. Under review.

Benton JS, Cotterill S, Anderson J, Macintyre VG, Gittins M, Dennis M, French DP. A natural experimental study of improvements along an urban canal: impact on canal usage, physical activity and other wellbeing behaviours. International Journal of Behavioral Nutrition and Physical Activity. Under review.

Published (not included in this PhD)

Macintyre VG, Cotterill S, Anderson J, Phillipson C, **Benton JS**, French DP. "I Would Never Come Here Because I've Got My Own Garden": Older Adults' Perceptions of Small Urban Green Spaces. International Journal of Environmental Research and Public Health. 2019;16(11):1994.

Contributions of the candidate and co-authors

The candidate (Jack Benton) undertook the conception and design of all studies in this PhD and wrote the first drafts for all papers. The candidate conducted the majority of data collection and analyses. The supervisors of this PhD (Prof David French, Dr Jamie Anderson and Dr Sarah Cotterill) inputted into the conceptualisation and design of the studies, advised on data analyses, contributed to revising the papers, and approved the final versions before submission for publication. Prof David French and Dr Jamie Anderson conducted some observations. Dr Jamie Anderson originally conceptualised the observation tool, which was developed and named by the candidate. Dr Sarah Cotterill conducted the statistical analyses for the power calculations in the two natural experimental studies, with the assistance of the candidate, and also inputted into the design and writing of analysis plans. Margaret Pulis, who was a co-author on the observation tool development paper, led the conceptualisation, design and data collection in the first validation study for the observation tool. Dr Ruth Hunter, who was a co-author on the observation tool development paper, provided input into the design and conduct of the third validation study for the observation tool. Dr Matthew Dennis, who was a co-author on two natural experimental studies and two study protocols, identified potential neighbourhoods using spatial analysis to help select potential comparison sites for the two natural experimental studies. Dr Matthew Gittins, who was a co-author on two natural experimental studies, provided expert advice on the statistical analyses. Prof Sarah Lindley, who was a co-author on one natural experimental study, provided input into the design of the study and revision of the manuscript. Vanessa Macintyre, who was a co-author on two natural experimental studies, led the nested qualitative study in the first natural experimental study and assisted with data collection, entry and analyses for both natural experimental studies.

Introduction to chapters

This thesis is presented as a series of papers, with each paper in its own chapter. Chapters 2 and 3 have been published in peer-reviewed journals (Cities & Health and BMC Public Health respectively), Chapter 5 has been registered on the Open Science Framework, and Chapters 4 and 6 are currently under peer review (Wellbeing, Space & Society and International Journal of Behavioral Nutrition and Physical Activity respectively).

Chapter 1 provides an overview of the main topics included in this thesis. The chapter discusses the use of natural experiments to study the effect of urban green space interventions on physical activity. This chapter reviews the limitations of existing natural experimental studies, including eight recent recommendations that I made in a pre-PhD review to address major methodological weaknesses causing high risk of bias in previous natural experimental studies of environmental interventions on physical activity. This chapter also highlights the potential of using systematic observation to assess a range of behaviours important for wellbeing by using the Five Ways to Wellbeing.

Chapter 2 reports on findings from three studies that aimed to develop and validate MOHAWk (Method for Observing pHysical Activity and Wellbeing): a newly developed systematic observation tool for assessing physical activity and two other evidence-based wellbeing behaviours (social interactions and taking notice of the environment) in urban spaces. Two studies were conducted in Greater Manchester (England) and one study in Belfast (Northern Ireland) using six observers across five urban spaces. From 156 hours of observations, evidence is provided that MOHAWk is reliable and valid, including inter-rater reliability, reliability of shortened observation schedules and criterion-related validity.

Chapter 3 is a protocol paper describing the planned methods and analyses for the first natural experimental study. This was a study of the impact of low-cost improvements (e.g. tree planting) to four urban amenity green spaces on older adults' and adults' wellbeing behaviours in Greater Manchester, UK. We developed a new five-step process to identify eight well-matched comparison sites using ten key variables associated with physical activity at the neighbourhood (e.g. population density) and street level (e.g. street lighting). Unobtrusive observations using MOHAWk were conducted at baseline, 6- and 12-months, with a nested qualitative study. We used a difference in differences analysis to control for known confounders (day, time of day, precipitation). The methods and analyses used to reduce confounding in this study are vast improvements on many previous natural experimental studies in this area, which have often used poorly matched comparison groups and poor control of confounding variables.

Chapter 4 reports on the findings of this first natural experimental study. There was no evidence that the interventions increased observable wellbeing behaviours or usage of green spaces in older adults or adults. The nested qualitative study suggested the interventions were not substantial enough to be noticed compared to other recent neighbourhood changes and therefore unlikely to influence behaviour change. These null findings are inconsistent with current UK policy and practice recommendations for planning and designing urban green spaces to increase physical activity. These findings suggest the need for more nuanced policy and practice recommendations that specify key contextual factors which may influence the effectiveness of urban green space interventions (e.g. type of green space, scale, local population characteristics).

Chapter 5 is a protocol paper describing the planned methods and analyses for the second natural experimental study. This was a study of the impact of new walking infrastructure and green space improvements along an urban canal on canal usage and wellbeing behaviours in adults and older adults in Greater Manchester, UK. Two well-matched comparison sites were identified using the same comparison site matching process used in the first natural experimental study. Unobtrusive observations using

MOHAWk were conducted at baseline and 7-, 12- and 24-month follow-ups, supplemented with intercept surveys at each follow-up.

Chapter 6 reports on the findings of this second natural experimental study. We found evidence that the intervention significantly increased the total number of people using the canal path compared to the comparison sites at all follow-ups. There was some evidence that the intervention brought about significant increases in walking and vigorous physical activity, social interactions, and people taking notice of the environment. The process evaluation suggested that there was some displacement of activity, but that the intervention also encouraged existing users to use the canal more often. Using robust natural experimental methods, these findings demonstrate the potential of relatively small-scale interventions designed to improve access to urban canals can bring about positive changes in green space usage. Small-scale interventions are rarely formally evaluated but offer a more feasible and less costly approach compared to whole-network interventions, especially in densely populated urban areas where opportunities for large-scale changes are limited.

Finally, Chapter 7 discusses the extent to which this PhD successfully addressed my eight recommendations to improve natural experimental studies. Suggestions for future research are then made, followed by a discussion of the implications for research, policy and practice.

Overview of thesis structure

Section 1: Background

- **Chapter 1** - Overview of the key topics addressed in this thesis, namely natural experimental methods to evaluate urban green space interventions, and eight recent recommendations to improve natural experimental studies of environmental interventions.

Section 2: Observation tool development and validation

- **Chapter 2** - Paper reporting on three studies that aimed to develop and validate MOHAWk (Method for Observing pHysical Activity and Wellbeing): a newly developed observation tool for assessing physical activity and two other evidence-based wellbeing behaviours in urban spaces (published).

Section 3: Natural experimental study 1

- **Chapter 3** - Protocol paper for the first natural experimental study of low-cost improvements to urban amenity green spaces (published).
- **Chapter 4** - Paper reporting the findings of the first natural experimental study (under peer review).

Section 4: Natural experimental study 2

- **Chapter 5** - Protocol paper for the second natural experimental study of improvements along an urban canal (published).
- **Chapter 6** - Paper reporting the findings of the second natural experimental study (under peer review).

Section 5: General Discussion

- **Chapter 7** - Overall discussion from the whole PhD, including methodological reflections, suggestions for future research and the wider implications for research, policy and practice.

SECTION 1. Background

Section overview

This section contains one chapter (Chapter 1) that provides an overview of the main topics included in this thesis. The chapter discusses the use of natural experiments to study the effect of urban green space interventions on physical activity. This chapter reviews the limitations of existing natural experimental studies, including eight recent recommendations that I made in a pre-PhD review to address major methodological weaknesses causing high risk of bias in previous natural experimental studies of environmental interventions on physical activity. This chapter also highlights the potential of using systematic observation to assess a range of behaviours important for wellbeing by using the Five Ways to Wellbeing.

Chapter 1. General Introduction

1.1 Physical activity and health

Physical activity can provide many physical, psychological and social health benefits for all age groups. Regular physical activity can prevent numerous chronic diseases (e.g. cardiovascular disease, type 2 diabetes, and many types of cancer) [1] and contributes to the prevention of other key risk factors for chronic diseases, such as hypertension and obesity [2]. Physical activity can prevent or delay functional decline and loss of independence, especially in people with chronic diseases and disabilities [3]. It has also been associated with improved wellbeing, mental health and quality of life [4], as well as social benefits (e.g. opportunities for social interaction) [5].

Despite the well-documented health benefits of physical activity, estimates from 2010 suggest that 23% of adults worldwide do not meet the World Health Organization (WHO) global recommendations on physical activity [6] i.e. to engage in 150 minutes of moderate-intensity activity (or equivalent) per week, such as brisk walking, gardening or household chores [7]. Physical inactivity is the fourth leading risk factor for global mortality, and approximately 3.2 million deaths are attributable to physical inactivity each year [8]. Therefore, physical inactivity presents a global public health burden, comparable to the impact of smoking and obesity [9].

1.2 Population-level interventions to increase physical activity

Much research evidence on tackling physical inactivity has focused on developing and evaluating individually focused (or ‘downstream’) interventions; that is, interventions tailored and delivered towards high-risk individuals who are the most physically inactive. Examples of individually focused interventions include counselling, social support programs and self-monitoring using pedometers. Individually focused interventions are more amenable to traditional evaluation research, such as randomised controlled trials

(RCT), resulting in rigorous evidence. However, individually focused interventions that aim to increase physical activity levels only yield small to moderate improvements [10,11]; are usually resource-intensive and thus difficult to scale up to reach large numbers of people; and tend to show little effectiveness on physical activity beyond 15 months [12]: all of which means individually focused interventions are not optimal in producing population-level change in physical activity [13].

Given the prevalence of physical inactivity across the population, small increases in physical activity in many individuals would have substantial population impact [14]. Therefore, prevention strategies should also include population-level (or ‘upstream’) interventions that affect large populations (e.g. environmental change, policies) to complement individually focused interventions. Population-level interventions attempt to address the underlying causes that make physical inactivity common across the population, such as inadequate walking infrastructure or the financial cost of cycling to work. Population-level interventions have the potential to reach a large proportion of the population over sustained periods of time, which means they are often more cost-effective than individually focused interventions [13,15]. Research efforts to preventing physical inactivity are now beginning to focus on population-level interventions to produce large and sustainable increases in physical activity across the population [16].

1.3 Urban green spaces and physical activity

Creating new or improving existing urban green spaces has the potential to promote physical activity in the population. For the purposes of this PhD, urban green space is defined as:

> “…any area of grass, trees or other vegetation, which in towns and cities is deliberately reserved for recreational, aesthetic or environmental purposes; this

term therefore covers a range of green urban features, including parks, sports pitches and streetscape greenery" ([17]: p.2).

Another common term applied in urban planning is 'green infrastructure', which is often used to refer to networks of green space [18]. Urban green spaces can also be located around blue space (e.g. canals, rivers, lakes), known as 'green-blue space' or 'green-blue infrastructure' [19].

Urban green spaces can provide people with attractive, accessible and functional space for engaging in a range of physical activities such as walking, cycling and sports, at low or no cost. Urban green spaces are often easier and less expensive to modify than other aspects of the built environment, such as street layouts or public transport [20]. However, due to increasing urbanisation, more people are living in built-up areas with limited green space. By 2050, it is predicted that two-thirds of the world's population will be living in urban areas [21]. If it is the case that the availability, accessibility and quality of green space in people's living environment influence their physical activity levels, then reductions in green space within urban areas can have serious ramifications for population levels of physical activity. Conversely, improving or creating new green space presents a valuable opportunity to intervene in the physical inactivity pandemic.

1.4 The evidence on urban green spaces and physical activity

Many studies have investigated the association between urban green spaces (including measures of the availability, accessibility and use of urban green spaces) and physical activity levels. Yet findings are generally mixed: reviews have found some studies show positive associations between urban green space and physical activity, whilst some studies have shown mixed, weak or no associations [19,22–28]. Reasons for inconsistent findings have been attributed to the considerable heterogeneity across the evidence base, such as differences in the types of green space evaluated (e.g. parks, woodlands, trails), exposure assessments (e.g. subjective versus objective measures) and target populations (e.g. different age groups) [23,27,29].

Moreover, most studies to date that have investigated the link between urban green space and physical activity are cross-sectional [30–32]. Cross-sectional studies cannot determine causality for a few reasons. First, and most importantly, the exposure (urban green space) and outcome (physical activity) are measured at the same point in time, which makes it difficult to know if the exposure temporally precedes the outcome [33]. Second, comparison groups are not part of the design in cross-sectional studies, but comparison groups enable researchers to construct a useful counterfactual inference; that is, comparing what would have happened to the outcomes in the absence of the intervention [34]. Third, cross-sectional studies are more likely to be biased by residential self-selection: people may choose to live in neighbourhoods based on their physical activity abilities, needs and preferences, thus confounding the relationship between exposure and outcome [35]. Therefore, whilst it would seem plausible that creating new or improving existing urban green spaces (hereafter referred to as ‘interventions’) will cause increases in physical activity, cross-sectional studies alone cannot provide strong evidence for causality.

1.5 Using natural experiments to strengthen causal inferences

RCTs are often considered the ‘highest’ grade of evidence to assess the causal effects of an intervention under controlled conditions [36]. Randomisation in RCTs reduces the risk of unevenly distributed confounding variables across intervention and comparison groups, meaning that any change in the outcome can be attributed to the intervention rather than some other known or unknown confounding variable. However, the use of randomisation to examine the effects of urban green space interventions on physical activity is rarely feasible due to practical, financial and ethical reasons [37]. Further, the relationship between urban green space interventions and physical activity is likely to be complex, as there are many interrelated physical, socio-economic, cultural and political factors that influence this relationship [38,39]. Therefore, using tightly controlled RCTs may mask important contextual factors integral to intervention effectiveness. For these reasons, RCTs are usually not the optimal approach for evaluating the causal effects of urban green space interventions.

There have been increasing calls for exploiting opportunities presented by natural experiments [37]. In line with United Kingdom (UK) Medical Research Council (MRC) guidance [40], natural experiments are broadly defined as any event or intervention that cannot be manipulated by the researcher and divides a population into exposed and unexposed groups. Examples of urban green space natural experiments include the development of new greenways, improvements to park facilities and greening of vacant lots. Natural experiments are therefore real-world events or interventions discovered by researchers, rather than a type of study design [41]. Researchers can design quasi-experimental studies around natural experiments: a study evaluating a natural experiment is referred to as a 'natural experimental study' [40]. Some researchers define natural experiments more narrowly to only include events where the process that defines exposure is random or 'as-if' random [42]. However, a more broad and inclusive definition is preferred because truly random assignment is rare in public health, and there are ambiguities in defining 'as-if' randomisation [43].

One approach to evaluating natural experiments is to measure physical activity levels before and after an intervention in an exposed intervention group and unexposed comparison groups. Another common type of natural experimental study design includes measuring physical activity levels before and after people relocate to a new neighbourhood i.e. residential relocation studies (e.g. [44]). However, relocating to a new neighbourhood is often associated with a myriad of other life events and changes that can confound physical activity levels, such as changes in marital status, income and family size [45]. For this reason, this chapter focuses on natural experiments involving urban green space interventions, rather than residential relocation.

Findings from natural experimental studies provide more robust evidence for causality than cross-sectional studies because they can establish temporality i.e. whether changes in exposure (urban green space) subsequently leads to changes in the outcome (physical activity) [33]. The temporal order of exposure and outcome also reduces the risk of bias due to residential self-selection. Because of the aforementioned issues associated with RCTs, controlled before and after natural experimental studies are the most rigorous study design that is feasible to evaluate the causal effects of urban green space interventions on physical activity. This is why many researchers are increasingly encouraging the use of natural experimental studies to study the effects of urban green space interventions [31,38].

Aside from investigating causality, using natural experiments offers many other advantages. Natural experiments are real-world interventions, which means that studies generate practice-based evidence for decision makers on the effectiveness (rather than efficacy) of interventions. Findings from natural experimental studies can shed light on potential health inequalities that are not amenable to evaluation in tightly controlled experimental conditions [37]. Also, researchers are not in control of natural experiments, which encourages collaboration between researchers, policymakers and practitioners: good relationships between researchers and policymakers are critical to achieving evidence-informed policy and practice [46].

Although natural experimental studies can provide stronger evidence of causality compared to cross-sectional studies, natural experimental studies are potentially at higher risk of bias than RCTs due to a lack of true randomisation [40]. A bias is a systematic error in the estimation of a causal effect due to specific flaws in the design, conduct, analysis or reporting within a study, rather than due to random error or chance [47].

Random assignment of units (i.e. people, groups, places) in RCTs distributes both known and unknown variables at random across intervention and comparison groups, thus reducing the risk that a third confounding variable explains the observed intervention effect i.e. reducing the risk of bias due to confounding [37]. However, randomisation is not feasible in natural experimental studies because the researcher lacks control of the intervention exposure or event.

The risk of bias in natural experimental studies is further exacerbated by the practical and methodological obstacles associated with evaluating interventions that the researcher cannot control. For example, delays in the timing of the intervention can affect follow-up data collection [48]. These difficulties, as well as the difficulties in discovering natural experiment opportunities, help explain why there is a dearth of natural experimental studies of urban green space interventions on physical activity.

1.5.1 High risk of bias in existing natural experimental studies

The use of natural experiments is still relatively new to public health research. This evidence base is starting to influence the proliferation of urban green space policy and industry statements, strategies and recommendations published in recent years across cities and countries worldwide [49–59]. It is therefore essential to assess the quality of the available evidence that is being used to inform policy and decision makers.

Prior to this PhD, I conducted a review of systematic reviews to assess the risk of bias in natural experimental studies of environmental interventions on physical activity [32]. From three existing exemplar systematic reviews, I identified 12 before-and-after controlled natural experimental studies (15 physical activity outcomes); nine of these natural experiments were urban green space interventions. I found that all outcomes had an overall “critical” (n = 12) or “serious” (n = 3) risk of bias, thus showing that the strongest natural experimental studies conducted to date are weak; a conclusion in line with other recent systematic reviews that have assessed risk of bias in this area [31,60–63].

1.5.2 Eight recommendations to reduce the risk of bias in natural experimental studies

Based on the findings of my pre-PhD review, I made eight recommendations to reduce the risk of bias in future natural experimental studies of environmental interventions on physical activity [32]:

1. Better matching of comparison sites;
2. Use multiple comparison sites;
3. Control for confounders in statistical analyses;
4. Publish study protocols with a priori analyses specified;
5. Use adequate outcome measurements;
6. Conduct sample size calculations;
7. Better reporting of samples and interventions; and,
8. Measure intervention exposure at the individual level.

1.5.2.1 Recommendation 1: Better matching of comparison sites

Bias due to confounding is a common problem in natural experimental studies [40]. While there are no general solutions to fully address the pervasive issue of confounding in natural experimental studies, the use of appropriate comparison (or control) groups is a key design feature that can reduce the risk of bias due to confounding in natural experimental studies.

Due to the absence of randomisation in natural experiments, comparison groups ought to be matched on all important variables that influence the outcome to strengthen internal validity by balancing the intervention and comparison groups at baseline [40]. However, nine of the 12 included studies in my review used poorly matched comparison groups. Studies rarely used objective variables that strongly correlate with physical activity for matching (e.g. population density, street connectivity).

1.5.2.2 Recommendation 2: Use multiple comparison sites

Multiple comparison sites should be used to increase the likelihood of finding balanced comparison groups and ensure any observed differences are attributable to the intervention rather than other site-specific variables [62]. However, six of the studies included in my review did not use multiple comparison sites.

1.5.2.3 Recommendation 3: Control for confounders in statistical analyses

Another key design feature that can reduce the risk of bias due to confounding is to control for important observed and unobserved confounders in statistical analyses. However, many studies in my review lacked appropriate statistical control of important confounders (e.g. weather). Given the difficulties in identifying equally balanced comparison groups on known and unknown confounders [40], appropriate analyses should be carried out that attempt to minimise these differences. MRC guidance provides recommendations on analysis techniques for evaluating natural experiments, such as a 'difference-in-differences' approach that controls for unobserved fixed confounders that cannot be balanced when matching intervention and comparison sites [40].

1.5.2.4 Recommendation 4: Publish study protocols with a priori analyses specified

None of the included 12 studies included in my review referred to a published pre-registered study protocol with a priori analyses specified. Publishing a study protocol increases transparency and thus reduces the risk of publication or reporting biases [64]. It also allows researchers to carefully plan the study and create contingency plans for potential issues [65], which is particularly important for natural experimental studies given that the researcher cannot control the intervention. Pre-registered study protocols can be published at low or no cost on open access platforms (e.g. Open Science Framework [66]) or as journal articles. In recent years since my review was carried out, there have been publications of pre-registered study protocols for natural experimental studies of urban green space interventions (e.g. [67–71]).

1.5.2.5 Recommendation 5: Use adequate outcome measurements

Bias in measurement of outcomes was the domain with the highest risk of bias for all included studies. There were a number of important issues that contributed to this particularly high risk of bias. Many studies relied on outcome measures that have not been validated: seven out of fifteen outcomes were reported using unvalidated outcome measures. Reliable and valid outcome measures are necessary to avoid using measurements that are inaccurate and vary over time, therefore reducing possible measurement bias [72]. Eight studies also relied on single outcome measures, yet triangulating different outcome measures is optimal to strengthen causal inferences. Finally, 11 of the 15 outcomes were conducted at only one follow-up time point, which is unlikely to provide a valid measure of change due to fluctuations in physical activity behaviour across different seasons and over time [31].

1.5.2.6 Recommendation 6: Conduct sample size calculations

None of the included studies reported any attempt to make a sample size calculation. Without an appropriate sample size calculation, studies may be at an increased risk of type II errors due to inappropriately small sample size to detect an effect. Alternatively, studies may have larger numbers of participants than is required to adequately power a study, resulting in overly expensive studies. Although there are a lack of data on urban green space interventions to form the basis of sample size calculations, more recent natural experimental studies evaluating urban green space interventions provide examples of methods for calculating the appropriate sample size (e.g. [68]).

1.5.2.7 Recommendation 7: Better reporting of samples and interventions

Poor reporting was a common issue for many studies in my review, which contributed to the high risk of bias. Clear, complete and transparent reporting of studies is necessary to facilitate replication, synthesis and critical appraisal of evidence [73]. Given the difficulties of eliminating all bias in natural experimental studies, transparent reporting is vital. In the absence of reporting guidelines specifically for natural experimental studies, relevant established guidelines such as TREND statement for non-randomised studies [74] or the STROBE statement for observational studies [75] are appropriate to ensure studies are clearly reported. It is also important that interventions are well-described; the Template for Intervention Description and Replication for population health and policy interventions checklist (TIDieR-PHP) [76] provides a formal tool that ensures reporting of interventions captures the complexity of the context into which interventions are placed [77].

1.5.2.8 Recommendation 8: Measure intervention exposure at the individual level

Measuring individual-level exposure is important to determine how much of the change in physical activity levels is attributable to participants that are exposed to the intervention [30]. Subjective measurement of intervention exposure increases the risk of invalid measurements of participants' exposure to the intervention [78]. Yet in my review, none of the four self-report outcomes that were conducted away from the intervention site measured intervention exposure objectively.

1.5.3 Other issues with existing natural experimental studies

The majority of natural experimental studies to date have been based in the US [19]. Due to the unique urban morphology of the US compared to European cities (e.g. lower density cities [79]), most US studies are of park-based or large-scale redevelopment interventions [31]. As a result, the impact of small-scale replicable green space interventions, such as improving urban canals, greening residential streets or creating informal amenity green spaces, remains largely unknown. This is an important evidence gap because understanding the impact of more affordable and feasible small-scale approaches (sometimes called 'urban acupuncture' approaches [80]) to improving urban green spaces is crucial, especially in more densely populated European urban areas where opportunities for larger-scale interventions are limited.

1.6 Outcome measurement: the potential of systematic observation

A major source of bias that this PhD addresses is in the measurement of outcomes. Reliable and valid measures of physical activity are important for monitoring trends in physical activity within populations; making comparisons between populations; and most importantly for this PhD, monitoring the effects of interventions [81]. However, a fundamental weakness of the evidence on evaluating environmental interventions is a reliance on unvalidated outcome measures [32].

Systematic observation (i.e. direct observations of behaviour using predetermined criteria) is one objective method of assessing physical activity, which has numerous advantages over other measures of physical activity, including self-reported measures (e.g. questionnaires) and device-based measures (e.g. accelerometers). Systematic observation is a potentially unobtrusive measure; that is, observations are intended to be covert, and therefore participants are generally unaware that their behaviour is being measured. This reduces possible reactivity of measurement associated with self-report and device-based measures of physical activity [82]. This also reduces possible recall bias associated with self-report [83]. The removal of participant burden eliminates selection bias due to low response rates associated with self-reported measures [83], with response rates as low as 14% not uncommon in natural experimental studies in this area [17]. Observations can provide reliable and valid estimates of the type and intensity of physical activity behaviour (e.g. sedentary, moderate, vigorous) [84]. Systematic observation can assess large numbers of people within a short period of time.

Using systematic observation is particularly useful when evaluating urban green space interventions for four key reasons. First, observations capture behaviour within the specific environment of interest, thus providing contextually rich data. Second, assuming ethical approval and accessibility, observations can be carried out in almost any setting. Third, the removal of participant burden means that observations can be used to assess subpopulations who may benefit from accessing and using urban green space, but for whom self-reported measures of physical activity are less appropriate e.g. people with cognitive or communication impairments. Finally, systematic observation can be easily carried out by non-academics (e.g. by volunteers from the local community [68]), thus providing a participatory and sustainable outcome measure. For these reasons, systematic observation is a promising measure of physical activity in the context of urban green spaces.

However, systematic observation can be resource-intensive, both in terms of time and money. The main costs derive from training and deploying people to observe, which can be substantial if using multiple observers. The resource-intensive nature of systematic observation can be somewhat mitigated with good planning. For example, preliminary studies before main data collection can inform the researcher how to optimally schedule observations to avoid having larger numbers of observations than is required [85]. A balance between accuracy and cost is necessary.

Another limitation with using systematic observation is that researchers cannot obtain information about individual-level behaviour change and information about green space users, such as whether new people are using the green space or where people came from (e.g. local residents or visitors) [31]. It is therefore important to triangulate observations with intercept surveys or household interviews to provide a more comprehensive assessment of physical activity.

1.6.1 Limitations of existing systematic observation tools

Numerous systematic observation tools have been developed for different settings and purposes [84,86–88]. System for Observing Play and Active Recreation in Communities (SOPARC) [84] is the most common validated observation tool for assessing physical activity [89]. SOPARC uses momentary time sampling techniques (brief observational scans of target areas) to measure the characteristics and physical activity levels of individuals during one-hour observation periods.

Although SOPARC has provided valuable data on a range of environmental interventions in various settings and populations, it has limitations. SOPARC was developed and validated for outdoor environments that attract high numbers of users, specifically neighbourhood and state parks, and therefore struggles to capture data in urban environments that typically have lower numbers of users (e.g. amenity green spaces) or that people pass through (e.g. green corridors). Also, SOPARC was initially developed in California, United States (US) [84], and the majority of subsequent studies using SOPARC have been conducted in the US [89]. Compared to many UK and other European cities, US cities often differ in terms of key variables that affect physical activity, such as climate, street design, population density, transportation networks and physical activity patterns [45,90]; therefore features of the tool may be unsuitable for settings dissimilar to the US. For example, SOPARC recommends that observations are not carried out during inclement weather. However, it is impractical and costly to rearrange observation periods during inclement weather in European cities that have much higher levels of rainfall, including Manchester [91].

1.6.2 Using systematic observation to assess a range of wellbeing behaviours

Urban green spaces not only provide opportunities for physical activity but also offer potential benefits for people’s wellbeing via other behavioural pathways. For example, urban green spaces are where people interact socially with others and get outdoors to take notice of nature [92,93]. It is likely that relationships between urban green spaces and physical activity are influenced by these other behaviours [94]. For example, social interactions can play an important role in explaining why some people engage in physical activity, such as older adults [95] and dog walkers [96]. However, existing validated systematic observation tools (including SOPARC) tend to focus only on physical activity and therefore fail to capture a wide range of behaviours that are important for wellbeing (hereafter referred to as ‘wellbeing behaviours’).

Systematic observation could be used to assess a range of wellbeing behaviours by using the ‘Five Ways to Wellbeing’ [97]. On behalf of the UK Government’s Foresight programme, the New Economics Foundation (NEF) conducted a review of the wellbeing literature. They identified five behaviours for which there was evidence that engaging in these behaviours improved an individual’s wellbeing: (1) Be Active (engage in physical activity); (2) Connect (socially interact with others); (3) Take Notice (awareness of one’s internal and external environment); (4) Keep Learning (learn new activities); and (5) Give (pursue altruistic activities). Three of these wellbeing behaviours can be observed in the context of people using outdoor urban environments: Be Active, Connect and Take Notice. The other two behaviours (Keep Learning and Give) are more difficult to observe in the context of people using outdoor urban environments.

There are many different theories and definitions of wellbeing [98], but using the Five Ways to Wellbeing behaviours as a framework for assessing wellbeing has many advantages. Importantly, there is evidence that each of the three observable Five Ways behaviours are associated with improved wellbeing, including both hedonic wellbeing (functioning well) and eudaimonic wellbeing (feeling good) [99–102]. For example, frequently engaging in physical activity (Be Active), time spent socialising (Connect) and savouring an experience (Take Notice) are all behaviours associated with improved wellbeing. Therefore, any measured changes in these Five Ways behaviours ought to indicate changes in wellbeing in the population. These behaviours also provide an accessible set of actions for people to improve personal wellbeing. The Five Ways to Wellbeing are increasingly being used by UK local and national government authorities, healthcare providers and other organisations (e.g. charities, workplaces) as a framework for promoting wellbeing [103–107]. However, there is a lack of intervention research that has examined how to effectively promote Connect and Take Notice behaviours, especially for environmental interventions [92].

A recent natural experimental study used systematic observation to assess three of the Five Ways behaviours (Be Active, Connect, Take Notice) before and after an urban green space intervention in the centre of Manchester, UK [92]. The researchers observed large increases in usage, Take Notice and Connect behaviours in the intervention site at one-year follow-up compared to a matched comparison site. This study demonstrates the feasibility of using systematic observation to assess wellbeing behaviours in urban green spaces and provides preliminary evidence of sensitivity to change.

1.7 Thesis aims and objectives

This chapter has highlighted major methodological weaknesses which have led to a high risk of bias in natural experimental studies of environmental interventions. The eight recommendations from my pre-PhD review provided clear and pragmatic suggestions to address these weaknesses [32]. The aims of this PhD were to implement these recommendations by developing new methods to improve natural experimental studies of environmental interventions, and to implement these new methods in natural experimental studies to assess the effectiveness of various urban green space interventions on physical activity and other wellbeing behaviours. There were three specific objectives:

1. Develop new methods to reduce the risk of bias in natural experimental studies;
2. Develop and validate a new systematic observation tool for unobtrusively assessing physical activity and other wellbeing behaviours in urban spaces;

3. Implement these new methods in two natural experimental studies of urban green space interventions.

1.8 References

1. World Health Organization. Global Recommendations on Physical Activity for Health. 2010. Available from: http://apps.who.int/iris/bitstream/10665/44399/1/9789241599979_eng.pdf

2. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ. 2006;174(6):801–9.

3. Durstine JL, Painter P, Franklin BA, Morgan D, Pitetti KH, Roberts SO. Physical activity for the chronically ill and disabled. Sports Med. 2000;30(3):207-19.

4. Penedo FJ, Dahn JR. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. Curr Opin Psychiatry. 2005;18(2):189–93.

5. Eime RM, Young JA, Harvey JT, Charity MJ, Payne WR. A systematic review of the psychological and social benefits of participation in sport for adults: informing development of a conceptual model of health through sport. Int J Behav Nutr Phys Act. 2013;10(1):135.

6. World Health Organization. Governance: Development of a draft global action plan to promote physical activity. World Health Organization; 2017. Available from: http://www.who.int/ncds/governance/physical_activity_plan/en/

7. World Health Organization. Global Health Observatory (GHO) data: Prevalence of insufficient physical activity. WHO. World Health Organization; 2016. Available from: http://www.who.int/gho/ncd/risk_factors/physical_activity/en/

8. World Health Organization. Global health risks: mortality and burden of disease attributable to selected major risks. 2009. Available from: http://www.who.int/healthinfo/global_burden_disease/GlobalHealthRisks_report_f

ull.pdf

9. Lee I-M, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet. 2012;380(9838):219–29.

10. Conn VS, Hafdahl AR, Mehr DR. Interventions to increase physical activity among healthy adults: meta-analysis of outcomes. Am J Public Health. 2011;101(4):751–8.

11. Foster C, Hillsdon M, Thorogood M, Kaur A, Wedatilake T. Interventions for promoting physical activity. Cochrane Database of Systematic Reviews. 2005;(1):CD003180.

12. Murray JM, Brennan SF, French DP, Patterson CC, Kee F, Hunter RF. Effectiveness of physical activity interventions in achieving behaviour change maintenance in young and middle aged adults: A systematic review and meta-analysis. Soc Sci Med. 2017;192:125–33.

13. Wu S, Cohen D, Shi Y, Pearson M, Sturm R. Economic Analysis of Physical Activity Interventions. Am J Prev Med. 2011;40(2):149–58.

14. Rose G. Sick individuals and sick populations. Int J Epidemiol. 2001;30(3):427–32.

15. Masters R, Anwar E, Collins B, Cookson R, Capewell S. Return on investment of public health interventions: a systematic review. Public Heal Wales. 2017;71:827-34.

16. Reis RS, Salvo D, Ogilvie D, Lambert E V, Goenka S, Brownson RC. Scaling up physical activity interventions worldwide: stepping up to larger and smarter

approaches to get people moving. Lancet. 2016;388(10051):1337–48.

17. Houlden V, Weich S, Jarvis S. A cross-sectional analysis of green space prevalence and mental wellbeing in England. BMC Public Health. 2017;17:460.

18. Taylor L, Hochuli DF. Defining greenspace: Multiple uses across multiple disciplines. Landsc Urban Plan. 2017;158:25–38.

19. World Health Organization. Urban Green Space Interventions and Health: A review of evidence. Copenhagen; 2017. Available from: http://www.euro.who.int/__data/assets/pdf_file/0005/321971/Urban-green-spaces-and-health-review-evidence.pdf?ua=1

20. Sugiyama T, Carver A, Koohsari MJ, Veitch J. Advantages of public green spaces in enhancing population health. Landsc Urban Plan. 2018;178:12–7.

21. United Nations. World Urbanization Prospects: The 2014 Revision, Highlights. New York; 2015. Available from: https://esa.un.org/unpd/wup/publications/files/wup2014-report.pdf

22. Kaczynski AT, Henderson KA. Environmental Correlates of Physical Activity: A Review of Evidence about Parks and Recreation. Leis Sci. 2007;29(4):315–54.

23. Lachowycz K, Jones AP. Greenspace and obesity: a systematic review of the evidence. Obes Rev. 2011;12(5):e183–9.

24. Bancroft C, Joshi S, Rundle A, Hutson M, Chong C, Weiss CC, et al. Association of proximity and density of parks and objectively measured physical activity in the United States: A systematic review. Soc Sci Med. 2015;138:22–30.

25. Lee ACK, Maheswaran R. The health benefits of urban green spaces: a review of the evidence. J Public Health. 2010;33(2):212–22.

26. Fong KC, Hart JE, James P. A Review of Epidemiologic Studies on Greenness

and Health: Updated Literature Through 2017. Curr Environ Heal Reports. 2018;5(1):77-87.

27. James P, Banay RF, Hart JE, Laden F. A Review of the Health Benefits of Greenness. Curr Epidemiol Rep. 2015;2:131–42.

28. Hartig T, Mitchell R, De Vries S, Frumkin H. Nature and Health. Annu Rev Public Heal. 2014;35:207–28.

29. Koohsari MJ, Mavoa S, Villanueva K, Sugiyama T, Badland H, Kaczynski AT, et al. Public open space, physical activity, urban design and public health: Concepts, methods and research agenda. Health Place. 2015;33:75–82.

30. McCormack GR, Shiell A. 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. 2011;8(1):125.

31. Hunter RF, Christian H, Veitch J, Astell-Burt T, Hipp JA, Schipperijn J. The impact of interventions to promote physical activity in urban green space: A systematic review and recommendations for future research. Soc Sci Med. 2015;124:246–56.

32. Benton JS, Anderson J, Hunter RF, French DP. The effect of changing the built environment on physical activity: a quantitative review of the risk of bias in natural experiments. Int J Behav Nutr Phys Act. 2016;13(1):107.

33. Hill B. The environment and disease: Association or causation? Proc R Soc Med. 1965;58:295–300.

34. Ferraro PJ. Counterfactual Thinking and Impact Evaluation in Environmental Policy. In: Birnbaum M, Mickwitz P, editors. Environmental program and policy evaluation: New Directions for Evaluation. Wiley InterScience; 2009. p. 75–84.

35. Boone-Heinonen J, Gordon-Larsen P, Guilkey DK, Jacobs DR, Popkin BM, Popkin BM. Environment and Physical Activity Dynamics: The Role of Residential Self-selection. Psychol Sport Exerc. 2011;12(1):54–60.

36. Shadish WR, Cook TD, Campbell DT. Experimental and quasi-exerpimental designs for generalized causal inference. Wadsworth Cengage Learning; 2002.

37. Petticrew M, Cummins S, Ferrell C, Findlay A, Higgins C, Hoy C, et al. Natural experiments: an underused tool for public health? Public Health. 2005;119(9):751–7.

38. World Health Organization. Urban Green Space Interventions and Health: A review of impacts and effectiveness. Copenhagen; 2017. Available from: http://www.euro.who.int/__data/assets/pdf_file/0010/337690/FULL-REPORT-for-LLP.pdf?ua=1

39. Rutter H, Cavill N, Bauman A, Bull F. Systems approaches to global and national physical activity plans. Bull World Health Organ. 2019;97(2):162–5.

40. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. J Epidemiol Community Health. 2012;66(12):1182–6.

41. Humphreys DK, Panter J, Ogilvie D. Questioning the application of risk of bias tools in appraising evidence from natural experimental studies: critical reflections on Benton et al., IJBNPA 2016. Int J Behav Nutr Phys Act. 2017;14(1):49.

42. Dunning T. Natural experiments in the social sciences: a design-based approach. Cambridge University Press; 2012. 358 p.

43. Craig P, Vittal Katikireddi S, Leyland A, Popham F. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health

Intervention Research. Annu Rev Public Heal. 2017;38:39–56.

44. Giles-Corti B, Bull F, Knuiman M, McCormack G, Van Niel K, Timperio A, et al. The influence of urban design on neighbourhood walking following residential relocation: Longitudinal results from the RESIDE study. Soc Sci Med. 2013;77:20-30.

45. Transportation Research Board. Does the Built Environment Influence Physical Activity: Examining the Evidence – TRB Special Report 282. 2005. Available from: http://onlinepubs.trb.org/onlinepubs/sr/sr282.pdf

46. Innvær S, Vist G, Trommald M, Oxman A. Health policy-makers' perceptions of their use of evidence: a systematic review. J Health Serv Res Policy. 2002;7(4):239–44.

47. Higgins JP, Altman DG. Chapter 8. Assessing Risk of Bias in Included Studies. In: Cochrane Handbook for Systematic Reviews of Interventions. Chichester, UK: John Wiley & Sons, Ltd; 2008. p. 187–241. Available from: http://doi.wiley.com/10.1002/9780470712184.ch8

48. Veitch J, Salmon J, Giles-Corti B, Crawford D, Dullaghan K, Carver A, et al. Challenges in conducting natural experiments in parks—lessons from the REVAMP study. Int J Behav Nutr Phys Act. 2017;14(1):5.

49. National Institute for Health and Clinical Excellence (NICE). Physical activity and the environment. Manchester; 2008. Available from: https://www.nice.org.uk/guidance/ph8/resources/physical-activity-and-the-environment-55460874949

50. National Institute for Health and Clinical Excellence (NICE). A summary of selected new evidence relevant to NICE public health guidance 8 "Physical

activity and the environment”. Evidence Update 57. Manchester: NICE; 2014. Available from: https://www.nice.org.uk/guidance/indevelopment/gid-phg97

51. Department of Health. Tackling obesities: future choices: project report. 2007. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/287937/07-1184x-tackling-obesities-future-choices-report.pdf

52. City of New York. Active Design Guidelines: Promoting physical activity and health in design. New York; 2010. Available from: https://centerforactivedesign.org/dl/guidelines.pdf

53. The National Physical Activity Plan Alliance (NPAPA). The US National Physical Activity Plan. 2011. Available from: http://www.physicalactivityplan.org/docs/2016NPAP_Finalforwebsite.pdf

54. Australian Medical Association (AMA). AMA Position Statement. Physical Activity - 2014. Kingston; 2014. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0749379704002417

55. Public Health England. Everybody active, every day: An evidence-based approach to physical activity. London; 2014. Available from: www.gov.uk/phe

56. Government Platform. Society of opportunities. 2007. Available from: http://www.stm.dk/publikationer/UK_Regeringsgrundlag2007/

57. Committee on Environmental Health. The Built Environment: Designing Communities to Promote Physical Activity in Children. Am Acad Pediatr. 2009;123(6):1591–8.

58. Department for Environment Food and Rural Affairs. A Green Future: Our 25 Year Plan to Improve the Environment. London; 2018. Available from:

https://www.gov.uk/government/publications/25-year-environment-plan

59. Public Health England. Improving access to greenspace A new review for 2020. London; 2020.

60. Roberts H, McEachan R, Margary T, Conner M, Kellar I. Identifying Effective Behavior Change Techniques in Built Environment Interventions to Increase Use of Green Space: A Systematic Review. Environ Behav. 2018;50(1):28–55.

61. MacMillan F, George ES, Feng X, Merom D, Bennie A, Cook A, et al. Do Natural Experiments of Changes in Neighborhood Built Environment Impact Physical Activity and Diet? A Systematic Review. Int J Environ Res Public Health. 2018;15(2):217.

62. Moore THM, Kesten JM, López-López JA, Ijaz S, McAleenan A, Richards A, et al. The effects of changes to the built environment on the mental health and well-being of adults: Systematic review. Health Place. 2018;53:237–57.

63. Stappers NEH, Van Kann DHH, Ettema D, De Vries NK, Kremers SPJ. The effect of infrastructural changes in the built environment on physical activity, active transportation and sedentary behavior – A systematic review. 2018;53:135–49.

64. Nosek BA, Ebersole CR, DeHaven A, Mellor D. The Preregistration Revolution. Open Science Framework. 2017: https://osf.io/2dxu5

65. Boccia S. Credibility of observational studies: why public health researchers should care? Eur J Public Health. 2015;25(4):554–5.

66. Collaboration OS. An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. Perspect Psychol Sci. 2012;7(6):657–60.

67. Dobbinson SJ, Veitch J, Salmon J, Wakefield M, Staiger PK, MacInnis RJ, et al.

Study protocol for a natural experiment in a lower socioeconomic area to examine the health-related effects of refurbishment to parks including built-shade (ShadePlus). BMJ Open. 2017;7(2):e013493.

68. Tully MA, Hunter RF, Mcaneney H, Cupples ME, Donnelly M, Ellis G, et al. Physical activity and the rejuvenation of Connswater (PARC study): protocol for a natural experiment investigating the impact of urban regeneration on public health. BMC Public Health. 2013;13(774).

69. Huang TTK, Wyka KE, Ferris EB, Gardner J, Evenson KR, Tripathi D, et al. The Physical Activity and Redesigned Community Spaces (PARCS) Study: Protocol of a natural experiment to investigate the impact of citywide park redesign and renovation. BMC Public Health. 2016;16(1):1160.

70. Astell-Burt T, Feng X, Kolt GS. Large-scale investment in green space as an intervention for physical activity, mental and cardiometabolic health: study protocol for a quasi-experimental evaluation of a natural experiment. BMJ Open. 2016;6(4):e009803.

71. Kestens Y, Winters M, Fuller D, Bell S, Berscheid J, Brondeel R, et al. INTERACT: A comprehensive approach to assess urban form interventions through natural experiments. BMC Public Health. 2019;19(1):51.

72. Sterne J, Higgins J, Reeves BC, on behalf of the development group for ACROBAT-NRSI. A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI), Version 1.0.0, 24 September 2014. 2014. Available from: http://www.riskofbias.info

73. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ. 2008;337:a1655.

74. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health. 2004;94(3):361–6.

75. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. PLoS Med. 2007;4(10):e296.

76. Campbell M, Katikireddi SV, Hoffmann T, Armstrong R, Waters E, Craig P. TIDieR-PHP: a reporting guideline for population health and policy interventions. BMJ. 2018;361:k1079.

77. Cotterill S, Knowles S, Martindale AM, Elvey R, Howard S, Coupe N, et al. Getting messier with TIDieR: Embracing context and complexity in intervention reporting. BMC Medical Research Methodology. 2018;18(12):1-10.

78. Chaix B, Merlo J, Evans D, Leal C, Havard S. Neighbourhoods in eco-epidemiologic research: Delimiting personal exposure areas. A response to Riva, Gauvin, Apparicio and Brodeur. Soc Sci Med. 2009;69:1306–10.

79. Patacchini E, Zenou Y, Henderson JV, Epple D. Urban Sprawl in Europe. Brookings-whart Pap Urban Aff. 2009;125–49.

80. Unt AL, Bell S. The impact of small-scale design interventions on the behaviour patterns of the users of an urban wasteland. Urban For Urban Green. 2014;13(1):121–35.

81. Wareham NJ, Rennie KL. The assessment of physical activity in individuals and populations: why try to be more precise about how physical activity is assessed? Int J Obes Relat Metab Disord. 1998;22 Suppl 2:S30-8.

82. Miles LM, Elbourne D, Farmer A, Gulliford M, Locock L, McCambridge J, et al. Bias due to MEasurement Reactions In Trials to improve health (MERIT): protocol for research to develop MRC guidance. Trials. 2018;19(1):653.

83. French DP, Sutton S. Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? Br J Health Psychol. 2010;15(3):453–68.

84. McKenzie TL, Cohen DA, Sehgal A, Williamson S, Golinelli D. System for Observing Play and Recreation in Communities (SOPARC): Reliability and Feasibility Measures. J Phys Act Heal. 2006;3(s1):S208–22.

85. Cohen DA, Setodji C, Evenson KR, Ward P, Lapham S, Hillier A, et al. How Much Observation Is Enough? Refining the Administration of SOPARC. J Phys Act Heal. 2011;8(8):1117–23.

86. McKenzie TL, Marshall SJ, Sallis JF, Conway TL. Leisure-Time Physical Activity in School Environments: An Observational Study Using SOPLAY. Prev Med. 2000;30(1):70–7.

87. Gehl Institute. The Public Life Data Protocol, Version 1.0. 2017.

88. Suminski RR, Dominick GM, Plautz E. Validation of the Block Walk Method for Assessing Physical Activity occurring on Sidewalks/Streets. Int J Environ Res Public Health. 2019;16(11):1927.

89. Evenson KR, Jones SA, Holliday KM, Cohen DA, McKenzie TL. Park characteristics, use, and physical activity: A review of studies using SOPARC (System for Observing Play and Recreation in Communities). Prev Med. 2016;86:153–66.

90. Tucker P, Gilliland J. The effect of season and weather on physical activity: a

systematic review. Public Health. 2007;121(12):909–22.

91. The World Bank. Average precipitation in depth (mm per year) [Data file]. 2019. Available from: https://data.worldbank.org/indicator/AG.LND.PRCP.MM?end=2014&start=2013&view=map

92. Anderson J, Ruggeri K, Steemers K, Huppert F. Lively Social Space, Well-Being Activity, and Urban Design: Findings From a Low-Cost Community-Led Public Space Intervention. Environ Behav. 2017;49(6):685–716.

93. Bowler DE, Buyung-Ali LM, Knight TM, Pullin AS. A systematic review of evidence for the added benefits to health of exposure to natural environments. BMC Public Health. 2010;10(456).

94. Broyles ST, Mowen AJ, Theall KP, Gustat J, Rung AL. Integrating Social Capital Into a Park-Use and Active-Living Framework. 2011;40(5):522–9.

95. Devereux-Fitzgerald A, Powell R, Dewhurst A, French DP. The acceptability of physical activity interventions to older adults: A systematic review and meta-synthesis. Social Science and Medicine. 2016;158:14-23.

96. Westgarth C, Christley R, Marvin G, Perkins E. I Walk My Dog Because It Makes Me Happy: A Qualitative Study to Understand Why Dogs Motivate Walking and Improved Health. Int J Environ Res Public Health. 2017;14(8):936.

97. New Economics Foundation. Five ways to wellbeing: The evidence. London; 2008. Available from: http://b.3cdn.net/nefoundation/8984c5089d5c2285ee_t4m6bhqq5.pdf

98. Huppert FA, Ruggeri K. Controversies in Well-being: Confronting and resolving the challenges. In: Bhugra D, Bhul K, Wong S, Gillman S, editors. Oxford

Textbook of Public Mental Health. Oxford: Oxford University Press. Available from: http://www.ippanetwork.org/wp-content/uploads/2017/02/Huppert-Ruggeri-Chapter-for-Oxford-Textbook-of-Public-Mental-Health.pdf

99. Dolan P, Peasgood T, White M. Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. J Econ Psychol. 2008;29:94–122.

100. New Economics Foundation. Well-being evidence for policy: A review. London; 2012. Available from: https://b.3cdn.net/nefoundation/10b8aabd90c5771ff9_a0m6bvv5a.pdf

101. Government Office for Science. Mental Capital and Wellbeing: Making the most of ourselves in the 21st century. Final Project report. London; 2008. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/292450/mental-capital-wellbeing-report.pdf

102. Helliwell JF, Putnam RD. The Social Context of Well-being. In: Huppert FA, Baylis N, Keverne B, editors. The Science of Well-being. Oxford: Oxford University Press; 2005. p. 435–59.

103. Michaelson J. Forward from Foresight: The uses of the new economics foundation’s well-being work. J Public Ment Health. 2013;12(2):98–102.

104. National Health Service. 5 steps to mental wellbeing. 2019. Available from: https://www.nhs.uk/conditions/stress-anxiety-depression/improve-mental-wellbeing/

105. Keeman A, Näswall K, Malinen S, Kuntz J. Employee wellbeing: Evaluating a wellbeing intervention in two settings. Front Psychol. 2017;8:505.

106. Farrier A, Dooris M, Froggett L. Five Ways to Wellbeing: holistic narratives of public health programme participants. Glob Health Promot. 2017;26(3):71–9.

107. Mind. Five ways to wellbeing. 2020. Available from: https://www.mind.org.uk/workplace/mental-health-at-work/taking-care-of-yourself/five-ways-to-wellbeing/

SECTION 2. Observation tool development and validation

Section overview

This section contains one chapter (Chapter 2) that reports on findings from three studies that aimed to develop and validate MOHAWk (Method for Observing pHysical Activity and Wellbeing): a newly developed systematic observation tool for assessing physical activity and two other evidence-based wellbeing behaviours (social interactions and taking notice of the environment) in urban spaces.

Chapter 2. Method for Observing pHysical Activity and Wellbeing (MOHAWk): validation of an observation tool to assess physical activity and other wellbeing behaviours in urban spaces

This article has been published in Cities & Health:

- Benton JS, Anderson J, Pulis M, Cotterill S, Hunter RF, French DP. Method for Observing pHysical Activity and Wellbeing (MOHAWk): validation of an observation tool to assess physical activity and other wellbeing behaviours in urban spaces. Cities Health. 2020.

2.1 Abstract

Direct observation of behaviour offers an unobtrusive method of assessing physical activity in urban spaces, which reduces biases associated with self-report. However, there are no existing observation tools that: (1) assess other behaviours that are important for people's wellbeing beyond physical activity; (2) are suitable for urban spaces that typically have lower numbers of users (e.g. amenity green spaces) or that people pass through (e.g. green corridors); and (3) have been validated in Europe. MOHAWk (Method for Observing pHysical Activity and Wellbeing) is a new observation tool for assessing three levels of physical activity (Sedentary, Walking, Vigorous) and two other evidence-based wellbeing behaviours (Connect: social interactions; Take Notice: taking notice of the environment) in urban spaces. Across three studies, we provide evidence that MOHAWk is reliable and valid from 156 hours of observation by six observers in five urban spaces in the UK. MOHAWk can be used in policy or practice (e.g. by local authorities or developers), or in more formal institutional based research projects. This new tool is an inexpensive and easy-to-use method of generating wellbeing impact evidence in relation to the urban physical or social environment. A manual providing detailed instruction on how to use MOHAWk is provided.

2.2 Background

Physical activity provides many health benefits for all age groups [1]. Despite this, most of the world's population is not sufficiently active to gain significant health benefits [2]. Characteristics of the urban environment (e.g. green space, street design) can influence population levels of physical activity [3]. Two-thirds of the world's population predicted to be living in urban areas by 2050 [4]. This growing urbanisation highlights the importance of understanding how the urban environment can facilitate (or inhibit) physical activity.

Systematic observation (i.e. direct observations of behaviour using predetermined criteria) is one method for assessing physical activity that offers many advantages. Systematic observation is an unobtrusive method; that is, participants are generally unaware that their behaviour is being assessed. This reduces possible reactivity of measurement associated with self-report and device-based measures of physical activity [5], reducing the risk of social desirability and recall bias. Systematic observation is not susceptible to poor response rates associated with self-reported measures [6], reducing the risk of selection bias. Systematic observation provides contextually rich data by assessing behaviour directly in the environment of interest. Assuming ethical approval, observations can be carried out in almost any publicly accessible urban environment.

There are several existing observation tools designed to assess physical activity in urban environments (e.g. [7–13]). The most widely-used validated observation tool for assessing physical activity in outdoor urban environments is System for Observing Play and Active Recreation in Communities (SOPARC) [7]. SOPARC uses momentary observational scans to assess the characteristics and physical activity behaviours of people in the area being observed. It has provided valuable data on a range of parks in various settings and populations [14].

However, there are three key reasons why a new observation tool is needed. First, SOPARC only assesses physical activity behaviour; it therefore does not capture other behaviours undertaken by people in urban environments, such as social interactions [15]. Second, SOPARC has only been validated in neighbourhood and state parks [7,16,17]. SOPARC uses single observational scans that only last several seconds, which are less likely to capture valid samples in urban spaces that typically have lower numbers of users (e.g. amenity green spaces) or that people pass through (e.g. residential streets). Third, SOPARC was initially developed in California, United States (US), and has since only been validated in the US. Although SOPARC has been used around the world, it has predominantly been used in the US; across two recent reviews of studies using SOPARC, only two studies were conducted in Europe (Turkey and Belgium), compared to 23 unique studies in the US [14,18]. This is an important issue because the US is different to many European cities in terms of contextual variables that affect physical activity (e.g. population density, city design, climate, population characteristics) [20,21]; therefore, features of the tool may be unsuitable for settings dissimilar to the US. One study has used SOPARC in the United Kingdom (UK) [19]; however, the researchers had to modify the tool in several ways (e.g. used continuous scanning throughout the observation period, modified demographic categories) and therefore of uncertain psychometric properties.

Existing observation tools have been used to assess physical activity. However, observation tools could be used to assess a wider range of behaviours that are known to influence wellbeing, beyond physical activity. Such behaviours have been identified in the ‘Five Ways to Wellbeing’ [22]. On behalf of the UK Government’s Foresight programme, the New Economics Foundation (NEF) conducted a review of the wellbeing literature. They identified five behaviours for which there is evidence that engaging in these behaviours improves an individual’s wellbeing, known as the ‘Five Ways to

Wellbeing' (or 'Five Ways'): Be Active (engage in physical activity); Connect (socially interact with others); Take Notice (be aware of the environment); Keep Learning (acquire knowledge or skill in something new); and Give (contribute to the community). Since there is evidence that each of the Five Ways behaviours are associated with improved wellbeing, including both hedonic and eudaimonic wellbeing [15,22–25], each of these behaviours can be used as indicators of wellbeing (hereafter referred to as 'wellbeing behaviours'). Three of the Five Ways behaviours (Be Active, Connect, Take Notice) can be observed and are relevant to urban environments.

We report here on the development and formal testing of a newly developed observation tool: Method for Observing pHysical Activity and Wellbeing (MOHAWk). An early version of this tool was used in a recent study that evaluated the impact of small-scale pocket park improvements in Manchester, UK [26]. The researchers observed significant increases in wellbeing behaviours assessed using MOHAWk at 1-year follow-up compared to a matched comparison site; thus demonstrating the feasibility of using this tool and evidence of sensitivity to change.

2.2.1 Purpose of MOHAWk

MOHAWk is an observation tool for assessing three levels of physical activity (Sedentary, Walking and Vigorous) and two other wellbeing behaviours (Connect: social interactions; Take Notice: taking notice of the environment) in urban spaces. It also measures the overall number of people, their characteristics (gender, age group, ethnicity), and the presence of incivilities in the environment where observations are carried out (e.g. graffiti, broken glass). MOHAWk has been designed to be used in a wide variety of urban spaces, particularly spaces that typically have lower numbers of users or that people pass through; examples of which may include residential streets, amenity green spaces, green corridors, pocket parks, and urban squares.

MOHAWk is different from existing validated observation tools in three key ways: (1) MOHAWk assesses two other wellbeing behaviours that are relevant to the use of urban spaces (social interactions and taking notice of the environment), not just physical activity; (2) MOHAWk observations occur continuously throughout the observation period, rather than using a series of single observational scans; and (3) MOHAWk observations are carried out regardless of weather conditions, rather than cancelling observations during inclement weather (a sensitivity analysis, or including weather as a covariate, can control for the confounding influence of weather).

MOHAWk is freely available for use. The tool consists of an instruction manual, a standardised observation form, and a data summary form – all of which are provided in Appendices A, B and C. An overview of MOHAWk and procedures for using the tool are summarised in Figure 2.1.

2.2.2 Aims of the present research

This paper reports on three studies that aimed to develop MOHAWk and test for evidence of reliability and validity. The specific aims of these studies were to: (1) assess inter-rater reliability for observing people's characteristics, physical activity levels, and additional wellbeing behaviours (Connect; Take Notice); (2) explore the reliability of shortened observation schedules; and (3) test for evidence of criterion-related validity of observing Take Notice behaviours.

Overview of MOHAWk

MOHAWk uses continuous scanning to record the characteristics and behaviours of each person entering a pre-determined boundary ('target area') during hour-long time periods ('observation periods'). Data are recorded using pen and paper. Observers use a standardised observation form (provided in Appendix B) to record the following information for each person that enters the target area during each observation period: gender (Male or Female), age group (Infant, Child, Teen, Adult or Older Adult), ethnicity (White or Non-white), physical activity level (Sedentary, Walking, Vigorous), social interaction (Connect or No Connect), taking notice of the environment (Take Notice or No Take Notice), activity type (Cycling, Using phone, Dog walking, or other pre-determined activities), and if mobility assistance is required (Yes or No).

The presence of the following 'incivilities' in the urban environment are also recorded: general litter, evidence of alcohol use (empty bottles/ cans), evidence of drug taking (e.g. needles, syringes), graffiti, broken glass, vandalism, dog mess, noise. These items were taken from an existing validated tool for assessing the quality of neighbourhood green space [27].

Observations are carried out regardless of weather conditions, unless weather conditions become so extreme that they compromise the observer's safety. To control for potential bias associated with weather, a sensitivity analysis is carried out to assess the impact of weather; specifically, precipitation. The observer records the duration of any precipitation that occurs during an observation period. Observation periods are removed for the sensitivity analysis (or included as a covariate) if the accumulated duration of any precipitation lasts for 50% or more of the observation period i.e. 30 minutes or more. Due to a lack of data on how precipitation affects people's behaviour in urban spaces, this threshold seemed a reasonable cut-off to account for the high variability in frequency and intensity of precipitation that can occur during observation periods.

MOHAWk data can be inputted into a statistical software program (e.g. SPSS, Excel) for analysis.

Coding Take Notice and Connect behaviours

Take Notice behaviours occur when individuals stop or slow down, and appear as if they are making a conscious decision to appreciate their surroundings. Examples of this include extended viewing of the scenery, an intentional pause in activity to look at or photograph something in the vicinity, or a pronounced head swivel to look at a specific object, view or person. Connect behaviours occur when individuals are engaging or interacting with a person or the people around them in some way. The activity must involve either conversing with other users (e.g. talking and listening, using sign language), being physically linked with someone (e.g. holding hands, linking arms), smiling and making eye contact when passing, or participation in a group activity.

Figure 2.1. An overview of MOHAWk and procedures for using the tool.

2.3 Methods

2.3.1 Description of studies

2.3.1.1 Study 1

MOHAWk was used in two sites in central Manchester, UK: All Saints Park (Site 1A) and St Peter's Chaplaincy (Site 1B) – see Figure 2.2. Site 1A is a small park (~0.9 hectares) that is surrounded on three sides by Manchester Metropolitan University. Site 1B is a non-green urban square located near The University of Manchester.

One observer (MP) used MOHAWk at these two sites across eight weekdays (Thursdays and Fridays) over four weeks during March 2017. Observation periods lasted two hours (8-10am and 2-4pm), but data were recorded into 15-minute blocks within each observation period to allow investigations of patterns of data within each observation period. Observations were fully counterbalanced between the two sites to control for week, day of week and time of day. To assess inter-rater reliability, on four days of observation, a second observer (DF or JA) independently conducted observations alongside MP at the same site. Data were collected over a total of 32 hours, including 16 hours using two observers simultaneously.

The differences between the two sites (e.g. benches, vegetation) permitted testing of criterion-related validity of observing Take Notice behaviours i.e. whether there are significantly higher observed counts of Take Notice behaviours in Site 1A, where there are more opportunities for Take Notice behaviours, compared to Site 1B.

The unanticipated presence of a temporary statue within Site 1A (Figure 2.2) on two days of observations also permitted further criterion-related validity testing: by assessing whether there are more Take Notice behaviours on the two days with the statue, compared to the other six days without the statue in the same site.

2.3.1.2 Study 2

Study 2 was a feasibility study for a natural experimental study of changes to urban green spaces on older adults' physical activity and wellbeing [31]. In terms of the development of MOHAWk, this study had five aims: (1) assess inter-rater reliability; (2) test several modifications to the MOHAWk tool following Study 1, including refined coding procedures (e.g. recording precipitation), refined age group categories, and other new codes (e.g. mobility assistance); (3) determine how many days of observation per week and hours per day are needed to provide a reliable estimate of activity in a UK urban environment; (4) determine what times of the day observations should be carried out to capture variation in activity across the course of a day; and (5) explore differences in activity patterns on weekdays compared to weekends.

Two observers (JB, SK) used MOHAWk at the same time at two separate residential streets (adjacent to small amenity green spaces) in South Manchester during July 2017 (Figure 2.3). One site was a residential street where changes in the aesthetic quality of green space were planned but had not yet been implemented at the time of observations (Site 2A). The other site was a residential street in the same neighbourhood, but no such changes were planned (Site 2B).

Observations were conducted 8am-6pm in 50-minute observation periods (e.g. 8–8.50 am, 9–9.50 am etc.). Observation periods lasted 50 minutes, rather than one hour, to provide a 10-minute break for each observer every hour. Data were recorded into three 15-minute blocks and one 5-minute block within each observation period to allow investigations of patterns of data within each observation period. On the first two days (Thursday and Friday), both observers independently conducted observations at Site 2A at the same time to assess inter-rater reliability. Then, one observer (JB) conducted observations for seven consecutive days from Saturday to Friday in Site 2A. At the same time, the second observer (SK) conducted observations at Site 2B for five consecutive days from Monday to Friday.

Both sites were similar apart from the following key differences at Site 2B: two benches, a litter bin, and more diverse vegetation. Site 2B was also rated by two observers (JB, SK) as being more aesthetically pleasing and better maintained than Site 2A using a validated tool for measuring the quality of neighbourhood green space [27]. These differences between Site 2A and 2B permitted testing of criterion-related validity: whether there are significantly higher observed counts of Take Notice behaviours in Site 2B, where there are more opportunities for Take Notice behaviours, compared to Site 2A.

2.3.1.3 Study 3

The aims of Study 3 were to: (1) assess inter-rater reliability; (2) test new coding procedures for recording Take Notice behaviours to improve the reliability of observing these behaviours; (3) test whether MOHAWk can accurately capture Take Notice and Connect behaviours whilst simultaneously collecting other data; and (4) use MOHAWk outside of Manchester i.e. Belfast, Northern Ireland.

Two observers (JB, CC) independently used MOHAWk at the same time for two consecutive days (Monday, which was a national bank holiday, and Tuesday) during August 2018. Observations were carried out at C.S Lewis Civic Square (Figure 2.4): a civic square in east Belfast, located at the intersection of the Connswater and Comber Greenways. This site contains public art (seven bronze art sculptures), a coffee bar, several seating areas, and green space. Observations were conducted using hour-long observation periods between 8am-4pm. Data were recorded in 5-minute blocks within each observation period to allow investigations of patterns of data within each observation period. Data were collected over 8 hours, all of which were conducted using two observers simultaneously.

On the Monday, both observers used MOHAWk between 10am and 12pm to assess inter-rater reliability. Take Notice behaviours (e.g. an intentional pause in activity to look at or photograph something in the vicinity) are momentary behaviours and thus more difficult to observe, particularly whilst recording other data. Therefore, in the afternoon (1-3pm), both observers used MOHAWk as normal. However, one observer coded Take Notice behaviours only when people were stationary, whilst the second observer coded all Take Notice behaviours regardless of whether people were stationary or not. This was to test whether coding Take Notice behaviours only when people are stationary improves inter-rater reliability for observing Take Notice behaviours.

On the Tuesday morning (8-9am and 10-11am), one observer used MOHAWk as normal, whilst the second observer only recorded Connect behaviours, gender and age group for each person. Similarly, in the afternoon (12-1pm and 2-3pm), one observer used MOHAWk as normal, whilst the second observer only recorded Take Notice behaviours, gender and age group for each person.

2.3.2 Observer training

There were six unique observers across the three studies (JB, MP, JA, DF, SK, CC). The majority of observers were aged between 18 – 34 years old ($n = 4$), and two observers were aged between 35 – 50 years old. There were four male observers and two females. All observers were physically active according to World Health Organisation criteria [32].

Study 1 focused on developing the observation procedures used in a recent study [26], which meant there was limited formal training. An instruction manual was developed as a result of Study 1. This instruction manual can be found in Appendix A, which provides detailed descriptions of MOHAWk procedures and coding conventions (e.g. how to distinguish between Walking and Vigorous behaviours, how to distinguish between age groups based on gait, clothing, and other physical attributes etc.).

Observers in Studies 2 and 3 were formally trained by JB using the MOHAWk instruction manual and by practising observations in the study sites. Training focused on becoming familiar with the operational definitions, key coding conventions, how to use the observation form, and how to code site incivilities. All observers in Study 2 and 3 received at least three hours of training and practising observations with feedback and inter-rater reliability assessments. The aim of training was to achieve inter-rater reliability of at least 0.75 (intraclass correlation coefficient (ICC)) for assessing the total number of people, each behaviour (Sedentary, Walking, Vigorous, Take Notice, Connect) and each participant characteristic (gender, age group, ethnicity). Any discrepancies between observers were resolved by discussion. Before each study, observers agreed on the boundaries of the target area in which all observed individuals were recorded.

2.3.3 Analyses

Table 2.1 contains a summary of the methods and analyses used to address each aim across the three studies. The unit of analysis for all analyses is at the level of the observation period i.e. counts per observation period.

Table 2.1. Summary of the methods and analyses used to assess each aim across the three studies

Aim	Studies	Length of observation period for the analysis	Data analysis methods
Inter-rater reliability	1, 2, 3	Study 1 and 2 = 15 minutes; Study 3 = 5 minutes	Two-way mixed, single measure, consistency ICCs
Reliability of a shortened observation schedule	2	50 minutes	Two-way mixed, single measure, consistency ICCs

Criterion-related validity	1, 2	15 minutes	Mann-Whitney U tests
Patterns of activity	2	50 minutes	Two-way mixed, single measure, consistency ICCs

ICC intraclass correlation coefficient

2.3.3.1 Inter-rater reliability

Inter-rater reliabilities were analysed using two-way mixed, single measure, consistency ICCs. ICCs are suitable for discrete data (i.e. count data). Unlike Cohen's [33] kappa, ICCs incorporate the magnitude of disagreement rather than an all-or-nothing agreement [34]. ICCs can be interpreted as < 0.5 = poor; 0.5 – 0.75 = moderate; 0.76 – 0.9 = good; and > 0.9 = excellent [35].

2.3.3.2 Reliability of shortened observation schedules

Two-way mixed, single measure, consistency ICCs were used to calculate the average reliability of overall daily counts of observed people at Site 2A and 2B for different abbreviated schedules. Specifically, ICCs were calculated for all possible abbreviated observation schedules across a week: combinations of 1, 2 or 3 days per week compared to the full 5 days per week (weekdays only). In the same way, ICCs were calculated for all possible abbreviated observation schedules across a day: combinations of 2, 3, or 4 hours per day compared to the full 10 hours per day. Two hours a day collection was defined as 1 hour in the first half of the day (between 8 am-1 pm) and 1 hour in the second half of the day (between 1-6 pm); three times a day was defined as morning (8 am-12 pm), early afternoon (12-3 pm) and late afternoon/ early evening (3-6 pm); and four times a day was defined as early morning (8-10 am), late morning (10 am-12 pm), early afternoon (12-3 pm) and late afternoon/ early evening (3-6 pm). These analyses were conducted separately for each age group: children, teens, adults and older adults.

2.3.3.3 Criterion-related validity

A Mann-Whitney U test was used to compare the number of Take Notice behaviours per observation period between: (i) Site 1A and Site 1B; (ii) two days when there was a temporary statue at Site 1A compared to the other six days without the statue; and (iii) Site 2A and Site 2B. All statistical tests were performed at the $p \leq 0.05$ level.

2.3.3.4 Patterns of activity

Two-way mixed, single measure, consistency ICCs were used to calculate the consistency in overall counts of people on weekdays compared to the weekend.

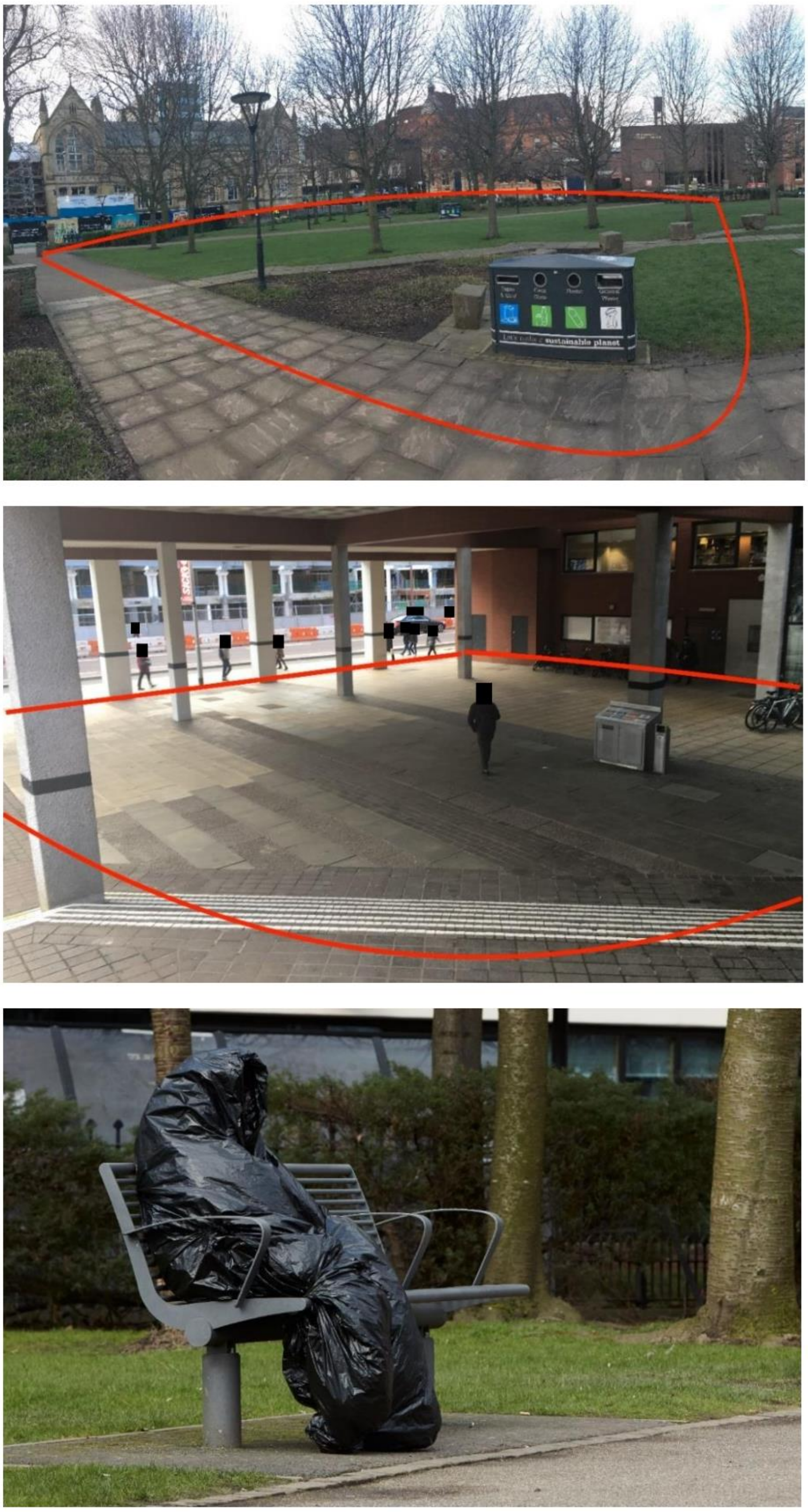


Figure 2.2. Photographs of Sites 1A (top), Site 1B (middle) and the temporary 'Hello Hollow' statue (bottom). *Photographs of Site 1A and 1B taken by Margaret Pulis. Photograph of 'Hello Hollow' was taken by Manchester Metropolitan University (permission approved by the Corporate Marketing and Communications department at Manchester Metropolitan University). All faces have been obscured.*

Jack Benton.

Figure 2.3. Photographs of Site 2A (top) and Site 2B (bottom). *Photographs taken by Jack Benton.*

Figure 2.4. Photographs of Site 3. *Photographs taken by Jack Benton.*

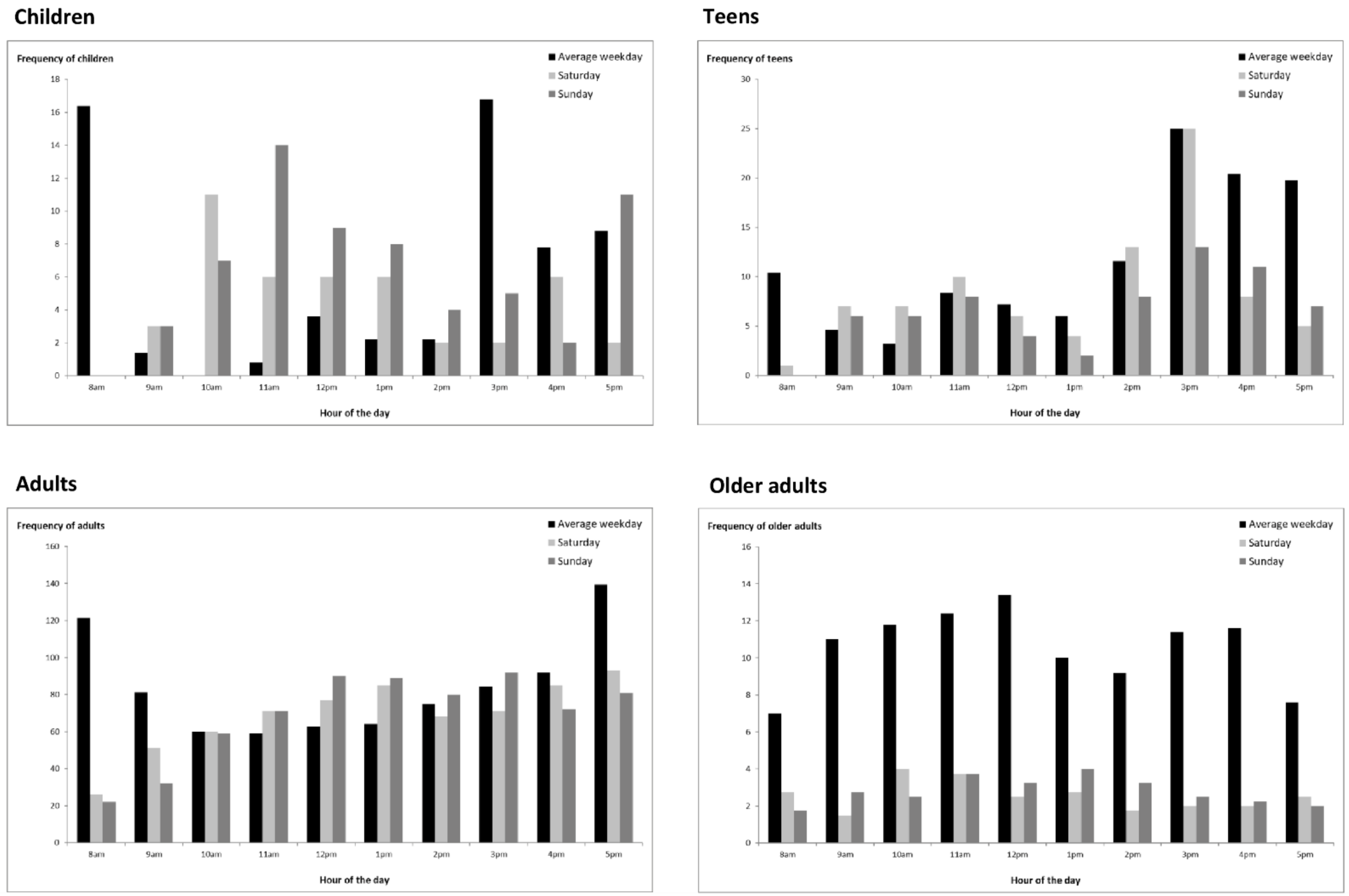


Figure 2.5. Frequencies of children (top left), teens (top right), adults (bottom left) and older adults (bottom right). *Observations were carried out at Site 2A, observed between 8-6 pm on weekdays (average of Monday to Friday), Saturday and Sunday during one week in July 2017.*

Table 2.2. Urban space characteristics and mean behavioural counts per hour for all three studies (five sites)

Study	Site	Type of site	Time of year	Days of observation (total hours)	Mean count of each behaviour per hour (S.D) and total count of behaviours observed per site, *N* [ab]				
					Sedentary	Walking	Vigorous	Take Notice	Connect
Study 1	All Saint's Park, Manchester, UK (Site 1A)	Park	March 2017	Thursday, Friday (16 hours)	6.56 (4.98) *N = 105*	96.56 (39.69) *N = 1497*	2.50 (1.83) *N = 40*	28.25 (22.41) *N = 509*	37.19 (32.65) *N = 595*
	St Peter's Chaplaincy, Manchester, UK (Site 1B)	Urban square	March 2017	Thursday, Friday (16 hours)	14.81 (9.48) *N = 237*	260.50 (95.72) *N = 4168*	7.06 (3.49) *N = 113*	1.75 (1.69) *N = 28*	68.13 (51.07) *N = 1090*
Study 2	Burton Road, Manchester, UK (Site 2A)	Residential street	July 2017	Monday – Sunday (75 hours [c])	2.79 (2.43) *N = 195*	70.53 (19.80) *N = 4937*	24.9 (13.69) *N = 1743*	1.81 (1.57) *N = 127*	22.2 (12.70) *N = 1554*
	Parsonage Road, Manchester, UK (Site 2B)	Residential street	July 2017	Monday – Friday (41 hours and 40 minutes [c])	2.55 (1.84) *N = 107*	62.95 (20.03) *N = 2644*	9.79 (5.13) *N = 411*	9.5 (3.71) *N = 399*	17.17 (16.55) *N = 721*
Study 3	CS Lewis Square, Belfast, UK (Site 3)	Civic square	August 2018	Monday, Tuesday (8 hours [d])	35.00 (5.00) *N = 105*	146.25 (26.58) *N = 439*	39.00 (9.64) *N = 117*	21.25 (11.93) *N = 64*	85.25 (28.73) *N = 256*

[a] Mean counts, standard deviations and total counts for each behaviour are based on the sum of children, teens, adults and older adults;

[b] Each observed person can engage in more than one behaviour per observation period e.g. a person who is walking and talking to a friend would be coded as 'Walking' and 'Connect';

[c] Fifty-minute observation periods;

[d] Mean counts, standard deviations and total counts for each behaviour based on four hours of data

2.4 Results

Table 2.2 displays descriptive statistics for all three studies. Appendix D reports on baseline data from three separate natural experimental studies that recently used MOHAWk [31,36,37]. Appendix E contains details of several key refinements that were made during each study to improve the reliability, validity, and usability of MOHAWk.

2.4.1 Inter-rater reliability

Across the three studies, inter-rater reliability between pairs of observers was mostly ‘good’ or ‘excellent’, with a small number of ‘moderate’ values (range of ICCs in brackets): total number of people (0.98-0.99), gender (0.91-0.99), age group (0.65-1), ethnicity (0.83-1), Sedentary (0.73-0.90), Walking (0.90-0.998), Vigorous (0.92-0.99), Take Notice (0.77-0.87), Connect (0.81-0.99), and mobility assistance required (0.88-1). Appendix F contains a more detailed breakdown of the ICCs and confidence intervals for inter-rater reliability across the three studies.

In Study 3, inter-rater reliability was ‘good’ for recording Take Notice behaviours when one observer only recorded Take Notice behaviours, gender and age group for each person, whilst the second observer used MOHAWk as normal (ICC = 0.80). Inter-rater reliability was ‘excellent’ for recording Connect behaviours when one observer only recorded Connect behaviours, gender and age group for each person, whilst the second observer used MOHAWk as normal (ICC = 0.97).

2.4.2 Reliability of shortened observation schedules

Table 2.3 displays ICCs for all shortened schedules for each age group. On average, observing on one day a week can produce good consistency approaching that obtained by observing five days a week for adults. For teens and older adults, observing on two days a week can produce good consistency approaching that obtained by observing five days a week. For children, observing on three days a week can produce good consistency approaching that obtained by observing five days a week.

On average, observing on two hours a day can produce good consistency approaching that obtained by observing 10 hours a day for adults. For teens and older adults, observing on three hours a day can produce good consistency approaching that obtained by observing 10 hours a day. For children, more than fours a day is required to produce good consistency approaching that obtained by observing 10 hours a day.

Table 2.3. Reliability estimates using the average ICC for shortened observation schedules. *Numbers in bold represent 'good' (ICC > 0.75) or 'excellent' (ICC > 0.9) reliability scores.*

Age group	**Number of days per week**			**Number of hours per day**		
	1	**2**	**3**	**2**	**3**	**4**
Children	.46	.62	**.79***	.48	.58	.65
Teens	.72	**.87***	**.94****	.70	**.82**	**.87***
Adults	**.91****	**.96****	**.98****	**.84***	**.91****	**.93****
Older adults	.64	**.82***	**.91****	.70	**.80***	**.86***

** 'Good' reliability (ICC > 0.75); ** 'Excellent' reliability (ICC > 0.9)*

2.4.3 Criterion-related validity

In Study 1, there were more Take Notice behaviours observed per 15-minute block in Site 1A (median = 7, interquartile range (IQR) = 9) compared to Site 1B (median = 1, IQR = 1) ($p < 0.001$). In Site 1A, there were more Take Notice behaviours observed per 15-minute block on two days when there was a temporary statue (median = 11, IQR = 8.75), compared to the other six days without the statue (median = 1, IQR = 3.75) ($p < 0.001$).

In Study 2, there were more Take Notice behaviours observed per 15-minute block in Site 2B (median = 2, IQR = 3) compared to Site 2A (median = 0, IQR = 1) ($p < 0.001$).

2.4.4 Activity on different days of the week

Figure 2.5 displays patterns of the total number of people across each hour of the day for each age group at Site 2A in Study 2, comparing the average weekday, Saturday and Sunday. There was poor consistency between the average weekday and Saturday (children: ICC = 0.28, teens: ICC = 0.46, adults: ICC = 0.46, older adults: ICC = 0.43) or Sunday (children: ICC = 0.21, teens: ICC = 0.33, adults: ICC = 0.31, older adults: ICC = 0.45) for each age group.

2.5 Discussion

These three studies indicated that MOHAWk is a reliable and valid observation tool. There was high agreement between pairs of observers for recording people's characteristics and their behaviours when using MOHAWk. There was evidence that shortened observation schedules can provide reliable estimates of people using urban spaces. In addition, there was evidence of criterion-related validity of observing Take Notice behaviours. We have provided extensive normative data (means, standard deviations, total counts) on all wellbeing behaviours in a variety of urban spaces to inform sample size calculations when using MOHAWk in natural experimental studies of urban spaces (see Table 2.2 and Appendix D).

2.5.1 Inter-rater reliability

There was good or excellent agreement between pairs of observers (ICC > 0.75) for 93% of observed behaviours and characteristics using six unique observers across three studies. This high agreement suggests that different observers can use MOHAWk and still produce very similar data, thus allowing reliable evaluation of multiple urban spaces at the same time. Inter-rater reliability remained high even when observing in busy urban spaces (e.g. there was an average of 118 observed people per hour in Study 2), suggesting that MOHAWk is robust enough to withstand busy urban spaces. We recommend a minimum of one full day (i.e. eight hours) of training and practice. However, the exact amount of training and practice required will depend on numerous factors, such as previous experience of the observer(s) in using MOHAWk and how busy the target areas are likely to be.

Maintaining high inter-rater reliability was an important consideration when developing the tool; for example, we reduced ethnicity codes into two categories to make it easier for observers to accurately record multiple wellbeing behaviours. Study 3 demonstrated that observers can still achieve high inter-rater reliability when recording the additional Connect (ICC = 0.97) and Take Notice (ICC = 0.80) behaviours. However, agreement on recording Take Notice behaviours tended to be lower than agreement on recording Connect and physical activity behaviours (Sedentary, Walking, Vigorous). This is likely because Take Notice behaviours are more momentary than Connect and physical activity behaviours, which makes Take Notice behaviours harder to observe; for example, someone pausing to look at something in the vicinity (Take Notice) is typically more momentary than someone holding hands (Connect) or cycling (Vigorous). Therefore, training should focus more on improving inter-rater reliability for recording Take Notice behaviours.

2.5.2 Reliability of shortened observation schedules

Study 2 showed that shortened observation schedules can provide reliable estimates of people using an urban space across a week and across a day, albeit not for children. These results are in line with a previous study that found shortened observation schedules using SOPARC can provide reliable estimates of park usage in the US [17]. This provides increased confidence that shortened observation schedules can provide reliable data, therefore reducing the time and cost required for observations. As a general guide, observing at least four hours a day, two days a week is recommended, although other schedules are also reliable (see Table 2.3).

2.5.3 Criterion-related validity

An important difference between MOHAWk and existing observation tools is the addition of two other wellbeing behaviours in MOHAWk: socially interacting with others (Connect) and taking notice of the environment (Take Notice). There is evidence that these additional wellbeing behaviours being assessed are valid and meaningful. Studies 1 and 2 demonstrated that there were significantly higher observed counts of Take Notice behaviours in sites that were hypothesised as offering more opportunities for Take Notice behaviours e.g. "greener" and more aesthetically pleasing sites. This suggests that MOHAWk is accurately capturing Take Notice behaviours since observed behaviours were in line with what we would expect to observe. Further, the frequencies of Take Notice and Connect behaviours varied between different sites and in different age groups (see Table 2.2), as well as between weekdays and weekends, which suggests that MOHAWk is sensitive to change and can therefore be used to measure the effect of interventions. These data build on those from an early version of MOHAWk [26], where the researchers observed significant increases in wellbeing behaviours assessed by MOHAWk after one year following an urban pocket park intervention i.e. demonstrating sensitivity to change.

The codes for observing physical activity in MOHAWk (Sedentary, Walking, Vigorous) are based on previous observation tools [7–10], which have been validated using heart rate monitors [10], pedometers [38] and accelerometers [39]; this suggests the physical activity codes used in MOHAWk are valid.

2.5.4 How MOHAWk compares to existing observation tools

To the best of our knowledge, there are no existing observation tools validated for use in urban spaces that typically have lower numbers of users (e.g. amenity green spaces) or that people pass through (e.g. residential streets). MOHAWk uses continuous scanning to count all people and their activities during one-hour observation periods, thus capturing activity in urban spaces which have lower levels of use that could not be captured by single observational scans used in existing observation tools. This is important because MOHAWk is more likely to produce larger sample sizes and thus better powered studies that require fewer observations due to increased sensitivity.

We have found evidence that MOHAWk is valid in two UK cities. There are no existing observation tools validated for use in Europe, and the vast majority of studies using existing observation tools have been conducted in the US [14,18]. This is an issue because the US is different to many European cities in terms of key variables that influence people's use of urban environments; for example, urban sprawl is much more prominent in the US compared to Europe [40]. Therefore, it is unclear whether existing observation tools are valid outside the US.

Existing tools, such as SOPARC, recommend that observations are not carried out during inclement weather. SOPARC was developed in California (US), which has a climate characterised mainly by mild-to-hot and dry weather. However, it is impractical and costly to use SOPARC procedures and rearrange observation periods during inclement weather in cities that have much higher levels of rainfall, such as Manchester and Belfast [41]. The many issues of rearranging observation periods due to weather have been discussed elsewhere (e.g. ambiguous weather forecasts) [42]. To address this issue in MOHAWk, observations are carried out regardless of weather conditions, but a sensitivity analysis controls for the confounding influence of precipitation (or including precipitation as a covariate). Whilst precipitation did not affect any of the observation periods in the three studies reported here, a recent natural experimental study that used MOHAWk in the UK had 50 hours of observations (out of 264 hours) that were removed for the sensitivity analysis due to precipitation [31]. Rearranging 50 hours of observation would have been costly and would have affected the design of the study, thus potentially introducing bias.

2.5.5 Strengths of MOHAWk

Many reviews have shown there is a scarcity of robust natural experimental studies of the causal effects of the urban environment on physical activity and wellbeing, particularly in Europe [43–49]. MOHAWk is an unobtrusive measure that will allow more robust natural experimental studies in this area by providing a measure that has evidence of reliability and validity, and is validated for use in Europe. MOHAWk is currently being used in three separate natural experimental studies of interventions in different types of urban spaces in Greater Manchester (UK), including residential streets [31], green corridor [36], and a small park [37].

We have demonstrated that MOHAWk is a reliable tool, with evidence of validity of observing Take Notice behaviours. This is important given that previous studies in this field have often relied on outcome measures that have not been validated: a recent review on natural experimental studies of changing the built environment on physical activity found that seven out of fifteen outcomes were reported using unvalidated outcome measures [6]. Results from the present studies suggest that MOHAWk is a reliable and valid outcome measure that can be used in a range of urban environments, thus promoting comparability between studies.

The three studies reported here have provided 156 hours of normative data. To date, there are also a further 172 hours of baseline data from three other natural experimental studies [31,36,37], with data provided in Appendix D. These normative data will help researchers conduct sample size calculations for future natural experimental studies; a lack of sample size calculations is a key weakness of previous natural experimental studies of urban spaces on physical activity [46]. These normative data will also help researchers determine the frequency and timing of observations to obtain accurate estimations of activity. More MOHAWk data are now needed in different urban spaces, settings (especially outside the UK) and populations (especially children and teens).

MOHAWk is a tool that can be used by most people. Existing observation tools have previously been used by non-researchers, such as volunteers from the local community [50]. Therefore, MOHAWk can feasibly be used in policy or practice by stakeholders involved in the planning, design, implementation and maintenance of urban spaces; such as local authorities, public health practitioners or developers. This new tool is an inexpensive and easy-to-use method of generating wellbeing impact evidence in relation to changes in the urban physical or social environment. A manual providing detailed instruction on how to use MOHAWk (Appendix A) and observation forms (Appendices B and C) are freely provided to facilitate its widespread use - please contact the corresponding author for further assistance on how to use MOHAWk.

2.5.6 Future research

Further psychometric testing of MOHAWk is required to increase evidence of validity and reliability, particularly for observing Connect behaviours as there was no validity testing for this in the present studies. For example, researchers could evaluate whether there are significantly more Connect behaviours during events where one would expect more social interactions (e.g. a summer fair) compared to days where there is no such event. Sensitivity to change is currently being tested in three separate natural experimental studies in Greater Manchester [31,36,37], but more natural experimental studies are needed to assess how responsive MOHAWk is to change. We encourage other researchers to use MOHAWk to evaluate environmental interventions - see methods described elsewhere [31,36] for examples of how to use MOHAWk in natural experimental studies. Future research should also look into using MOHAWk in the evening, although this may only be feasible using image/ video-capture devices due to ethical issues associated with deploying observers outside daylight hours.

2.7 Conclusions

The results of 156 hours of MOHAWk data, collected by six unique observers across five different sites, suggest that MOHAWk is a reliable and valid observation tool for assessing physical activity and other wellbeing behaviours in urban spaces. We provide extensive normative data on all wellbeing behaviours to inform sample size calculations when using MOHAWk in natural experimental studies of urban spaces. This new observation tool will allow more robust natural experimental studies, particularly in Europe where there is a dearth of robust evidence in this area. MOHAWk can also be used in policy or practice to generate wellbeing impact evidence in relation to changes in the urban physical or social environment.

2.8 References

1. UK Department of Health. Start active, stay active. A report on physical activity for health from our four home countries' chief medical officer. London; 2011. Avaliable from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216370/dh_128210.pdf

2. Lee I-M, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet. 2012;380(9838):219–29.

3. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJ, Martin BW. Correlates of physical activity: why are some people physically active and others not? Lancet. 2012;380(9838):258–71.

4. United Nations. World Urbanization Prospects: The 2014 Revision, Highlights. New York; 2015. Available from: https://esa.un.org/unpd/wup/publications/files/wup2014-report.pdf

5. French DP, Sutton S. Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? Br J Health Psychol. 2010;15(3):453–68.

6. Benton JS, Anderson J, Hunter RF, French DP. The effect of changing the built environment on physical activity: a quantitative review of the risk of bias in natural experiments. Int J Behav Nutr Phys Act. 2016;13(1):107.

7. McKenzie TL, Cohen DA, Sehgal A, Williamson S, Golinelli D. System for Observing Play and Recreation in Communities (SOPARC): Reliability and Feasibility Measures. J Phys Act Heal. 2006;3(s1):S208–22.

8. McKenzie TL, Marshall SJ, Sallis JF, Conway TL. Leisure-Time Physical Activity in School Environments: An Observational Study Using SOPLAY. Prev Med. 2000;30(1):70–7.

9. McKenzie TL, Sallis JF, Nader PR. SOFIT: System for Observing Fitness Instruction Time. J Teach Phys Educ. 1992;11(2):195–205.

10. Mckenzie TL, Saws JF, Nader PR, Patterson TL, Eider JP, Berry CC, et al. BEACHES: an observational system for assessing children's eating and physical activity behaviors and associated events. J Appl Behav Anal. 1991;1(24):141–51.

11. Gehl J, Svarre B. How to Study Public Life. How to Study Public Life. Island Press; 2013. 1-179 p.

12. Gehl Institute. The Public Life Data Protocol, Version 1.0. 2017.

13. Suminski RR, Dominick GM, Plautz E. Validation of the Block Walk Method for Assessing Physical Activity occurring on Sidewalks/Streets. Int J Environ Res Public Health. 2019;16(11):1927.

14. Evenson KR, Jones SA, Holliday KM, Cohen DA, McKenzie TL. Park characteristics, use, and physical activity: A review of studies using SOPARC (System for Observing Play and Recreation in Communities). Prev Med. 2016;86:153–66.

15. Helliwell JF, Putnam RD. The Social Context of Well-being. In: Huppert FA, Baylis N, Keverne B, editors. The Science of Well-being. Oxford: Oxford University Press; 2005. p. 435–59.

16. Whiting JW, Larson LR, Green GT. Monitoring Visitation in Georgia State Parks Using the System for Observing Play and Recreation in Communities (SOPARC). J Park Recreat Adm Winter. 2012;30(4):21–37. Available from:

https://lrl.people.clemson.edu/WebFiles/Whiting.etal.2012_JPRA-SOPARC.pdf

17. Cohen DA, Setodji C, Evenson KR, Ward P, Lapham S, Hillier A, et al. How Much Observation Is Enough? Refining the Administration of SOPARC. J Phys Act Heal. 2011;8(8):1117–23.

18. Joseph RP, Maddock JE. Observational Park-based physical activity studies: A systematic reviewof the literature. Prev Med. 2016;89:257–77.

19. Gidlow C, Ellis N, Smith G, Fairburn J. Promoting Green Space in Stoke-on-Trent (ProGreSS) Final report for Natural England August 2010. 2010. Available from: http://www.staffs.ac.uk/health_progress

20. Sallis JF, Cerin E, Conway TL, Adams MA, Frank LD, Pratt M, et al. Physical activity in relation to urban environments in 14 cities worldwide: a cross-sectional study. Lancet. 2016;387(10034):2207–17.

21. Tucker P, Gilliland J. The effect of season and weather on physical activity: a systematic review. Public Health. 2007;121(12):909–22.

22. New Economics Foundation. Five ways to wellbeing: The evidence. London; 2008. Available from: http://b.3cdn.net/nefoundation/8984c5089d5c2285ee_t4m6bhqq5.pdf

23. New Economics Foundation. Well-being evidence for policy: A review. London; 2012. Available from: https://b.3cdn.net/nefoundation/10b8aabd90c5771ff9_a0m6bvv5a.pdf

24. Dolan P, Peasgood T, White M. Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. J Econ Psychol. 2008;29:94–122.

25. McEwan K, Richardson M, Sheffield D, Ferguson FJ, Brindley P. A Smartphone

App for Improving Mental Health through Connecting with Urban Nature. Int J Environ Res Public Health. 2019;16(18):3373.

26. Anderson J, Ruggeri K, Steemers K, Huppert F. Lively Social Space, Well-Being Activity, and Urban Design: Findings From a Low-Cost Community-Led Public Space Intervention. Environ Behav. 2017;49(6):685–716.

27. Gidlow CJ, Ellis NJ, Bostock S. Development of the Neighbourhood Green Space Tool (NGST). Landsc Urban Plan. 2012;106(4):347–58.

28. Saelens BE, Frank LD, Auffrey C, Whitaker RC, Burdette HL, Colabianchi N. Measuring Physical Environments of Parks and Playgrounds: EAPRS Instrument Development and Inter-Rater Reliability. J Phys Act Heal. 2016;3(s1):S190–207.

29. Broomhall M, Giles-Corti B, Lange A. Quality of Public Open Space Tool (POST). Perth; 2004.

30. Millstein RA, Cain KL, Sallis JF, Conway TL, Geremia C, Frank LD, et al. Development, scoring, and reliability of the Microscale Audit of Pedestrian Streetscapes (MAPS). BMC Public Health. 2013;13(1):403.

31. Benton JS, Anderson J, Cotterill S, Dennis M, Lindley SJ, French DP. Evaluating the impact of improvements in urban green space on older adults' physical activity and wellbeing: protocol for a natural experimental study. BMC Public Health. 2018;18(1):923.

32. World Health Organization. Global Recommendations on Physical Activity for Health. 2010. Available from: http://apps.who.int/iris/bitstream/10665/44399/1/9789241599979_eng.pdf

33. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20(1):37–46.

34. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutor Quant Methods Psychol. 2012;8(1):23–34.

35. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155–63.

36. Benton J, Anderson J, Cotterill S, Dennis M, French DP. The impact of new walking infrastructure and changes to green space along an urban canal on physical activity and wellbeing: protocol for a natural experimental study. Open Sci Framew. 2018; https://osf.io/zcm7v/

37. Anderson J, French DP, Benton JS, Dennis M, Rothwell J. The effects of the Grow Green urban greening improvements on physical health and wellbeing behaviours in Manchester UK: A natural experimental study. Open Sci Framew. 2020; https://osf.io/zqgcn

38. Rowe P, van der Mars H, Schuldheisz J, Fox S. Measuring Students' Physical Activity Levels: Validating SOFIT for Use with High-School Students. J Teach Phys Educ. 2004;23(3):235–51.

39. McKenzie T, Sallis J, Armstrong C. Association between direct observation and accelerometer measures of children's physical activity during physical education and recess. Med Sci Sports Exerc. 1994;26(143).

40. Patacchini E, Zenou Y, Henderson JV, Epple D. Urban Sprawl in Europe. Brookings-Wharton Pap Urban Aff. 2009;125–49.

41. The World Bank. Average precipitation in depth (mm per year) [Data file]. 2019. Downloaded from: https://data.worldbank.org/indicator/AG.LND.PRCP.MM?end=2014&start=2013&view=map

42. Veitch J, Salmon J, Giles-Corti B, Crawford D, Dullaghan K, Carver A, et al. Challenges in conducting natural experiments in parks—lessons from the REVAMP study. Int J Behav Nutr Phys Act. 2017;14(1):5.

43. Benton JS, Anderson J, Hunter RF, French DP. The effect of changing the built environment on physical activity: A quantitative review of the risk of bias in natural experiments. Int J Behav Nutr Phys Act. 2016;13(1).

44. Roberts H, McEachan R, Margary T, Conner M, Kellar I. Identifying Effective Behavior Change Techniques in Built Environment Interventions to Increase Use of Green Space: A Systematic Review. Environ Behav. 2018;50(1):28–55.

45. MacMillan F, George ES, Feng X, Merom D, Bennie A, Cook A, et al. Do Natural Experiments of Changes in Neighborhood Built Environment Impact Physical Activity and Diet? A Systematic Review. Int J Environ Res Public Health. 2018;15(2):217.

46. Hunter RF, Christian H, Veitch J, Astell-Burt T, Hipp JA, Schipperijn J. The impact of interventions to promote physical activity in urban green space: A systematic review and recommendations for future research. Soc Sci Med. 2015;124:246–56.

47. Kärmeniemi M, Lankila T, Ikäheimo T, Koivumaa-Honkanen H, Korpelainen R. The Built Environment as a Determinant of Physical Activity: A Systematic Review of Longitudinal Studies and Natural Experiments. Ann Behav Med. 2018;52(3):239–51.

48. Houlden V, Weich S, Porto de Albuquerque J, Jarvis S, Rees K. The relationship between greenspace and the mental wellbeing of adults: A systematic review. PLoS One. 2018;13(9):e0203000.

49. Moore THM, Kesten JM, López-López JA, Ijaz S, McAleenan A, Richards A, et al. The effects of changes to the built environment on the mental health and well-being of adults: Systematic review. Health Place. 2018;53:237–57.

50. Tully MA, Hunter RF, Mcaneney H, Cupples ME, Donnelly M, Ellis G, et al. Physical activity and the rejuvenation of Connswater (PARC study): protocol for a natural experiment investigating the impact of urban regeneration on public health. BMC Public Health. 2013;13.

SECTION 3. Natural experimental study 1

Section overview

This section relates to the first natural experimental study of low-cost improvements to four urban amenity green spaces on older adults' and adults' wellbeing behaviours in Greater Manchester, UK. Chapter 3 is the protocol paper describing the planned methods and analyses. Chapter 4 reports on the findings of this study.

Chapter 3. Evaluating the impact of improvements in urban green space on older adults' physical activity and wellbeing: protocol for a natural experimental study

This article has been published in BMC Public Health:

- Benton JS, Anderson J, Cotterill S, Dennis M, Lindley SJ, French DP. Evaluating the impact of improvements in urban green space on older adults' physical activity and wellbeing: protocol for a natural experimental study. BMC Public Health. 2018;18(1):923.

3.1 Abstract

Background: Creating or improving urban green space has the potential to be an effective, sustainable and far-reaching way to increase physical activity and improve other aspects of wellbeing in the population. However, there is a dearth of well-conducted natural experimental studies examining the causal effect of changing urban green space on physical activity and wellbeing. This is especially true in older adults and in the United Kingdom. This paper describes a natural experimental study to evaluate the effect of four small-scale urban street greening interventions on older adults' physical activity and wellbeing over a one-year period, relative to eight matched comparison sites. All sites are located in deprived urban neighbourhoods in Greater Manchester, United Kingdom.

Methods: Components of the interventions include tree and flower planting, and artificial tree decorations. Eight unimproved comparison sites were selected based on a systematic process of matching using several known objective and subjective environmental correlates of physical activity in older adults. The outcome measures are physical activity and two other behavioural indicators of wellbeing (Connect: connecting with other people; Take Notice: taking notice of the environment), collected using a newly developed observation tool. The primary outcome is Take Notice behaviour due to largest effects on this behaviour being anticipated from improvements in the aesthetic quality of green space at the intervention sites. Baseline data collection occurred in September 2017 before the interventions were installed in November 2017. Follow-up data collection will be repeated in February/ March 2018 (6 months) and September 2018 (12 months).

Discussion: The present study permits a rare opportunity to evaluate the causal effects of small-scale changes in urban green space in an understudied population and setting. Although the interventions are expected to have small effects on the outcomes, the present study contributes to developing natural experiment methodology in this field by addressing key methodological weaknesses causing high risk of bias in previous natural experimental

studies. Key improvements to reduce risk of bias in the present study are rigorous matching of multiple comparison sites and appropriate statistical control of key confounders.

Trial registration: Retrospectively registered with study ID NCT03575923. Date of registration: 3 July 2018.

3.2 Background

Engaging in regular physical activity provides many physical, social and psychological health benefits for older adults [1,2]. Despite this, physical activity levels tend to decline as adults become older [3,4]. The rapidly increasing number of older adults (≥ 60 years old) worldwide, due to increased life expectancy, emphasises the need to improve the health of older adults to extend quality of life [5].

Interventions targeted at the individual level to increase physical activity in adults yield modest improvements and often lack long-term effectiveness, and are thus not cost-effective [6] and this lack of long-term effectiveness is also evident in older adults [7]. In contrast, creating a supportive environment at the population level may be a more effective, sustainable and far-reaching approach to increase older adults' physical activity levels by targeting the wider determinants of health [8]. Providing a physical environment that supports physical activity may be particularly important for older adults because they are likely to spend more time in their local living environment [9].

Many studies have now shown an association between the built environment (e.g. street design, land-use mix, aesthetic qualities) and physical activity levels in the adult [10] and older adult population [11]. One type of built environment that is particularly promising to improve population levels of physical activity is urban green space [12]; defined as "all publicly owned and publicly accessible open space with a high degree of cover by vegetation, e.g. parks, woodlands, nature areas and other green space" [13, p. 110] within urban areas. However, there is a lack of evidence on the effects of urban green space specifically in older adults [14,15] and findings from this small evidence base in older adults are mixed [16]. The effects of urban green space on older adults may be different to the effects on younger adults because many of the known moderators of the relationship between urban green space and physical activity are strongly linked to age, such as usage of urban green space, physical activity preferences, health and mobility [12,17].

The quality of the evidence to date is limited in a number of ways. First, most of the evidence on urban green space and physical activity is cross-sectional, which limits our ability to infer causality. To infer causality, evaluations of natural experiments have become a priority in this field [15,18,19]. Natural experiments are 'real world' events or changes that cannot be manipulated or controlled by the researcher and divide a population into exposed and unexposed groups [20]. Natural experimental studies can provide stronger inferences about causality than cross-sectional studies due to the temporal order of exposure (change to environment) and outcome (physical activity). Only a handful of studies have investigated causal effects of changes to urban green space using natural experiments: two recent systematic reviews of studies evaluating the effect of the built environment on physical activity [11] and active travel [21] in older adults found only one natural experimental study.

Second, most research on urban green space has been conducted in the United States (US) and Australia, and there is a lack of studies in the United Kingdom (UK) and elsewhere in Europe [11,19]. There are many differences between the US and UK in terms of key variables that affect physical activity levels, including differences in climate, population density, transportation networks and physical activity patterns [22,23]. These differences make it difficult to generalise findings from the US to Europe.

Third, the methodological rigour of the small number of natural experimental studies evaluating the effect of the built environment on physical activity in adult populations is weak. A recent review by Benton et al. [24] evaluated the risk of bias in natural experiments that have evaluated the effects of changing the built environment on physical activity. They identified 12 natural experimental studies (15 physical activity outcomes) on the basis of having strong experimental designs from three existing systematic reviews; nine of these studies evaluated urban green space interventions. They found that all outcomes had an overall critical (n = 12) or serious (n = 3) risk of bias, thus suggesting that the strongest studies conducted to date are weak, a conclusion in line with other recent systematic reviews in this area that included risk of bias assessments [15,25,26].

Benton et al. provided eight recommendations to improve the rigour of future natural experimental studies in this field, based on Medical Research Council (MRC) guidance for using natural experiments [20] as well as other relevant literature [e.g. 15]; the present study is designed to implement these recommendations:

1. Publishing study protocols with a priori analyses specified;
2. Better matching of comparison sites and more nuanced use of graded exposure;
3. Use of multiple comparison sites;
4. Sample size calculations;
5. Use of adequate outcome measurements;
6. Measuring exposure to the intervention at the individual level;
7. Better reporting of samples and interventions; and;
8. Controlling for confounding domains (in statistical analyses).

Urban green space may improve health and wellbeing by several mechanisms, not just physical activity [18]. There has been increasing attention on the effect of urban green space on a wider set of wellbeing indicators [18,27,28]. For instance, urban green space that is a positive sensory and symbolic resource (e.g. visual and audible information of value to users) may encourage people to access and draw upon that space more often, providing a pleasant setting for activities such as social interaction and mindful cognitions [27,29]. It is therefore important to understand how urban green space influences other indicators of wellbeing, in addition to physical activity.

In common with physical activity research, there is a scarcity of evidence on the causal effects of urban green space on wellbeing (i.e. natural experimental studies) [30]. Studies have also often relied on non-validated measures of wellbeing [31]. However, the science of urban wellbeing is in its infancy. In particular, issues in defining wellbeing may, in part, explain the lack of rigorous studies on the effect of urban green space on wellbeing [18,32].

An objective method of measuring wellbeing is to measure associated behavioural indicators of wellbeing. On behalf of the UK Government's Foresight project, New Economics Foundation (NEF) conducted a review of the wellbeing literature and identified behaviours for which there was good evidence that engaging in these behaviours improved an individual's sustained wellbeing. These behaviours are collectively known as the Five Ways to Wellbeing [33]: Be Active (engage in physical activity); Connect (connect with others); Take Notice (awareness of the environment); Keep Learning (learn new activities); Give (give back to the community). Since there is evidence that these five behaviours are linked to improved wellbeing [33], each of these five behaviours can be used as proxy measures of wellbeing.

A recent natural experimental study used direct observations of behaviour (systematic observation) to measure three of the Five Ways behaviours (Be Active, Connect, Take Notice) that are relevant to use of open spaces. They were measured before and after improvements to an urban green space site in central Manchester, UK; which included installation of shade-tolerant planting, an inner-city lawn and vegetation management [27]. The researchers observed significant increases in usage and wellbeing behaviours in the intervention site at one-year follow-up compared to a matched comparison site. Therefore, systematic observation of behavioural indicators of wellbeing offers a feasible method of objectively measuring and quantifying wellbeing in the context of the environment.

3.2.1 Research aim and objectives

The overall aim of this natural experimental study is to investigate the effect of planned changes in urban green space on older adults' physical activity and wellbeing. The specific objectives are as follows:

1. To examine whether small-scale urban green space interventions increase older adults' Take Notice behaviour in comparison to matched comparison sites where no such changes occur (primary outcome). The primary outcome is Take Notice behaviour due to largest effects on this behaviour being anticipated from improvements in the aesthetic quality of green space at the intervention sites;
2. To examine whether small-scale urban green space interventions increase the number of older adults' using the intervention sites in comparison to matched comparison sites where no such changes occur (secondary outcome);
3. To examine whether small-scale urban green space interventions increase older adults' physical activity levels (sedentary, walking, vigorous activity) or Connect behaviour in comparison to matched comparison sites where no such changes occur (exploratory outcomes).

3.3 Feasibility study

3.3.1 Background

To inform the main natural experimental study, a feasibility study was carried out in July 2017. The specific objectives of the feasibility study were to determine: (1) how many days of observation per week and hours per day are needed to provide a valid estimate of older adults' activity in a UK urban setting; (2) what times of the day should observations be carried out to capture variation in older adults' activity across the course of a day; and (3) any key differences in older adults' activity patterns on weekdays compared to weekends.

The feasibility study employed MOHAWk (Method for Observing pHysical Activity and Wellbeing) [34]: a newly developed tool for systematically observing physical activity and behavioural indicators of wellbeing in small urban green spaces such as pocket parks, tree-lined streets and green corridors along waterways (see Methods in the Main Study for further description of the tool).

3.3.2 Methods

Observations were carried out by two observers using MOHAWk at the same time at two separate sites in Greater Manchester (GM). Observations were conducted 8am-6pm in 50-minute observation periods (e.g. 8-8.50am, 9-9.50am etc.) by one observer for seven consecutive days from Saturday to Friday in Intervention Site 1 (Table 3.1/ Figure 3.6). At the same time, a second observer conducted observations in a different site for five consecutive days from Monday to Friday.

3.3.3 Results

To estimate the average reliability of overall daily counts of older adults, single rater, two-way random effects, consistency measure intraclass correlation coefficients (ICC) were used. ICCs were calculated for all possible abbreviated schedules: combinations of two, three, or four days per week compared to the full five days per week (weekdays only). It was found that, on average, observing on two days a week can produce consistency approaching that obtained by observing five days a week (ICC = .82).

Similarly, ICCs were used to estimate the average reliability of hourly counts of older adults for the following abbreviated schedules: combinations of two, three, or four hours per day compared to the full 10 hours per day. Two hours a day collection was defined as one hour in the first half of the day (8am-1pm) and one hour in the second half of the day (1-6pm); three times a day was defined as morning (8am-12pm), early afternoon (12-3pm) and late afternoon/ early evening (3-6pm); and four times a day was defined as early morning (8-10am), late morning (10am-12pm), early afternoon (12-3pm) and late afternoon/ early evening (3-6pm). It was found that, on average, observing four hours per day can produce consistency approaching that obtained by observing 10 hours per day (ICC = .86).

In general, counts of older adults were higher in the morning (8am-12pm) than early afternoon (12-3pm) or late afternoon/ early evening (3-6pm) (Figure 3.1). This is in line with previous studies suggesting that that older adults are generally more active in the mornings, particularly in the late morning, rather than afternoons and evenings [35-37], with very little activity occurring beyond 6pm [36]. As shown in Figure 3.1, there were fewer older adults observed on weekend days than weekdays.

3.3.4 Discussion

The observation data for older adults in a UK urban setting informed decisions about the frequency and timing of observations in the main study.

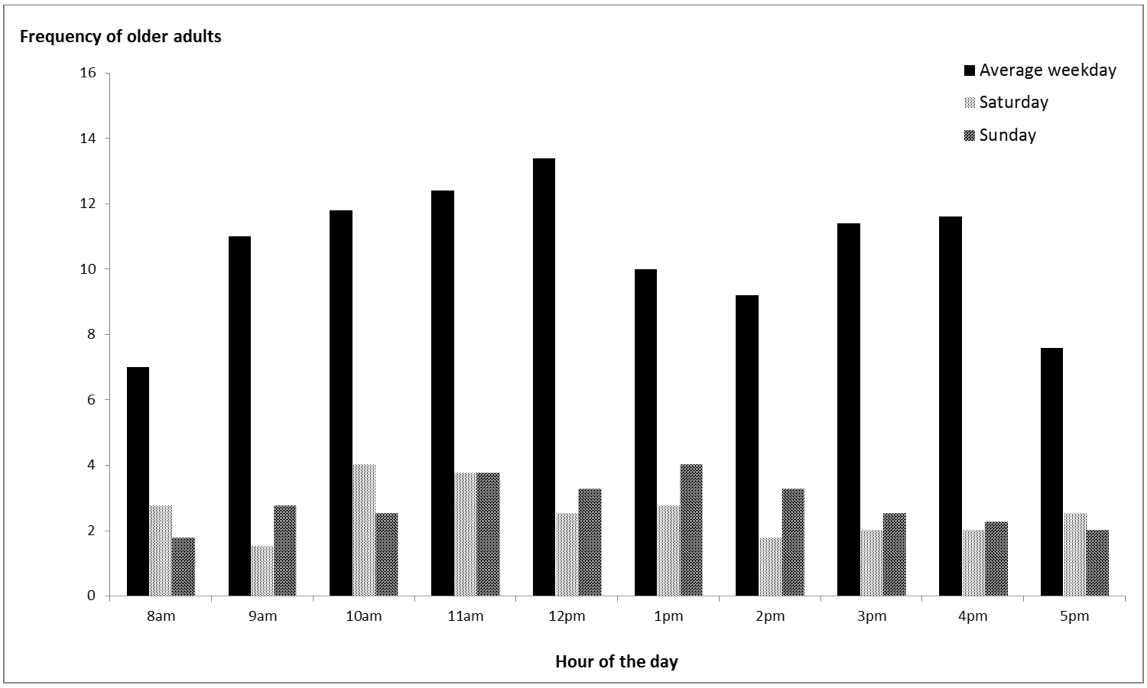


Figure 3.1. Frequency of older adults in Intervention site 1 during the feasibility study. *Counts of older adults were observed between 8am-6pm on weekdays (average of Monday to Friday), Saturday and Sunday during the feasibility study in July 2017.*

3.4 Methods (Main Study)

3.4.1 Study design

This is a prospective controlled before and after natural experimental study of the effects of changes in urban green space, with four intervention and eight matched comparison sites in GM: two comparison sites per intervention site.

3.4.2 Study population

Data will be collected on all individuals (infants, children, teens, adults, older adults) entering the target area during observations. However, this study is focused on older adults and therefore primary analyses will only consider data from older adults. Secondary analyses will consider data from adults.

3.4.3 Procedure

3.4.3.1 Comparison site matching

Due to the absence of randomisation in natural experiments, comparison (or control) groups ought to be matched on all important variables that influence the outcome to strengthen internal validity and improve the accuracy of the estimated intervention effect [20]. However, previous studies in this area have often used poorly matched comparison groups; in particular, an absence of any matching based on objective features of the environment that correlate with physical activity e.g. population density, street connectivity [24]. To address this issue, the present study selected comparison sites that were matched to corresponding intervention sites, using several key objective and subjective environmental variables. Two comparison sites were matched to each intervention site to increase the likelihood of finding comparable comparison sites.

There is an absence of evidence on how characteristics of the built environment may influence wellbeing; accordingly, all variables that comparison sites were matched on were from three recent systematic reviews that have investigated built environmental correlates of older adults' physical activity: one qualitative [38] and two quantitative systematic reviews [11,21]. Figure 3.2 displays the variables that were reported as consistent correlates of older adults' physical activity in at least two of these reviews: these correlates represent the key variables that were used for comparison site matching.

There are no agreed-upon standards for how researchers ought to identify matched comparison sites when studying the effects of the built environment on physical activity [24]; therefore, a five-step process of matching was developed. An overview of this process is displayed in Figure 3.2 and each step is described in more detail in Appendix G.

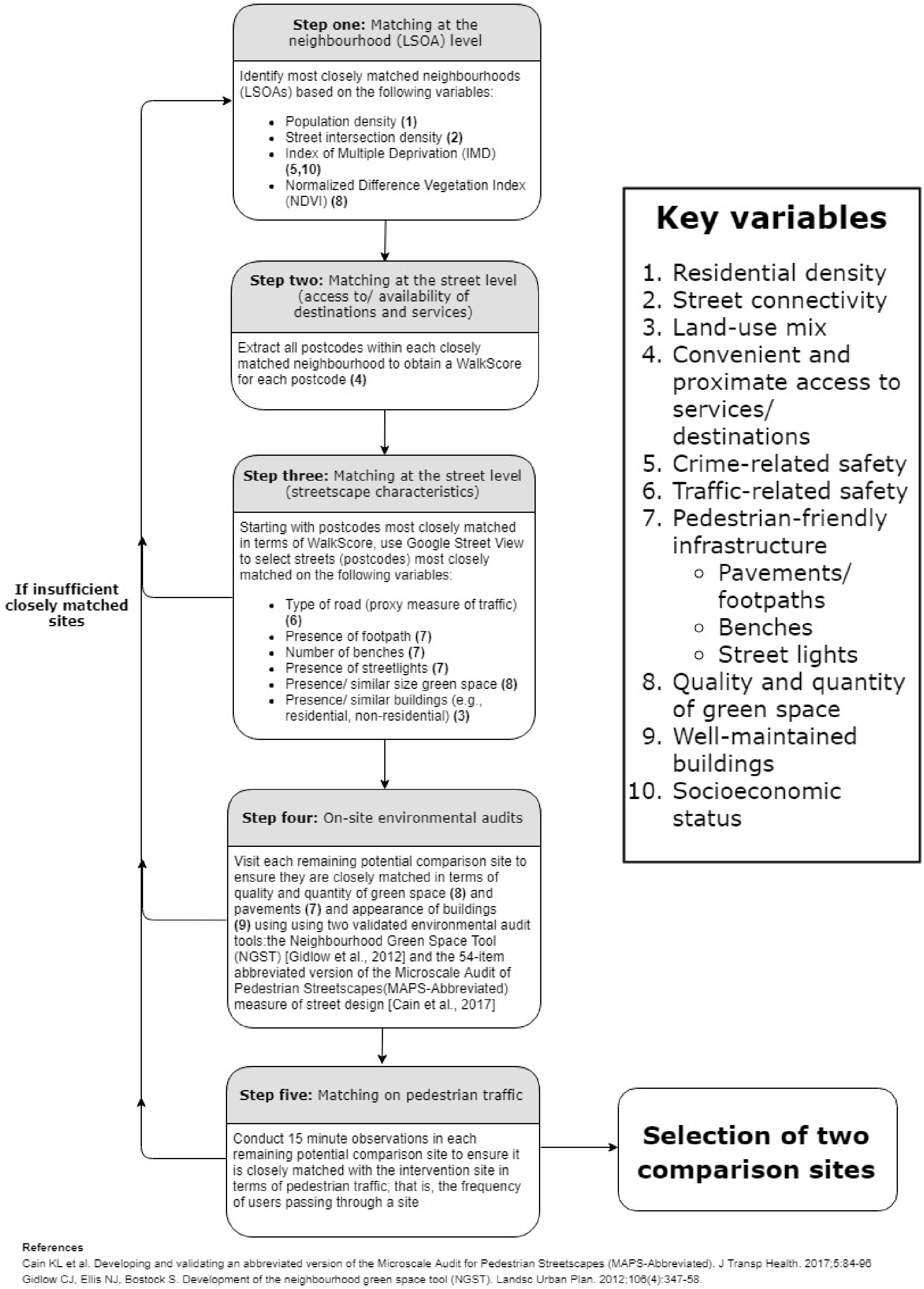


Figure 3.2. Overview of the five steps constituting the comparison site matching and selection process. *Numbers in brackets refer to the key variables used for comparison site matching.*

3.4.3.2 Timing

Baseline data collection was conducted in September 2017 before any changes in urban green space occurred in November 2017. Follow-up will be conducted at two time points: February/ March 2018 (6 months) and September 2018 (12 months). The first follow-up is intended to measure initial short-term effects of the interventions, six months post-baseline and three months after completion of the interventions. The second follow-up will be conducted one year after baseline, at the same time of year to control for seasonal variation.

3.4.3.3 Observation schedule and procedure

Informed by data from the feasibility study, observations will be conducted over two days, four times a day (weekdays only) at each time point, resulting in a total of eight observation periods for each site at each time point. Observations will be conducted at four set observation periods per day: morning (10-11am), lunchtime (12-1pm), afternoon (3-4pm), and evening (5-6pm). These times were found to capture the biggest variation in older adults' activity across the day (Figure 3.1), whilst also providing sufficient time for breaks and possible travel to other sites in between observation periods.

Observations for each intervention site and the two corresponding comparison sites will be spread over two weeks. This will provide a more robust assessment of activity over a longer period rather than observing activity during a single week. Observations will be counterbalanced to control for week, day of week and time of day. The observation schedule used for baseline data collection is displayed in Figure 3.3; this schedule will be replicated at all follow-ups. Any missed observations (e.g. due to illness) will be rescheduled for the same day of the next available week.

The same procedure will be used as set out in the MOHAWk observation manual (Appendix A). All observers will be trained in using MOHAWk and will be required to demonstrate high agreement with the trainer before making observations in the present study. Prior to observations, observers will visit each site to agree on the boundaries of the target area in which all participants will be recorded. Target areas are of similar size between corresponding intervention and comparison sites, and the same target areas will be used at all time points (see Appendix H).

Data will be recorded using pencil and paper. Data will be entered into SPSS at each time point once data collection is completed.

September 2017

	Monday		Tuesday		Wednesday		Thursday		Friday		Saturday	Sunday
	4th		5th		6th		7th		8th		9th	10th
10am			3	4A			4A	4	3A	3B		
12pm			3	4A			4A	4	3A	3B		
3pm			3B	4			4A	4B	3	3B		
5pm			3B	4			4A	4B	3	3B		
	11th		12th		13th		14th		15th		16th	17th
10am	1A	1B	3A	3B	1	-	2A	2	2A	3		
12pm	1A	1B	3A	3B	1	-	2A	2	2A	3		
3pm	1A	1	3A	3	1A	-	2A	2B	2	3A		
5pm	1A	1	3A	3	1A	-	2A	2B	2	3A		
	18th		19th		20th		21st		22nd		23rd	24th
10am	1	-	4	4B	1A	1B	2B	4B	2	2B		
12pm	1	-	4	4B	1A	1B	2B	4B	2	2B		
3pm	1B	-	4A	4B	1	1B	2	4	2A	2B		
5pm	1B	-	4A	4B	1	1B	2	4	2A	2B		

Figure 3.3. Observation schedule for two observers during baseline data collection for all intervention and comparison sites. *The two columns within each day correspond to the two observers i.e. one column for each observer. The numbers/ letters refer to intervention and comparison sites e.g. '1' is Intervention site 1, and '1A' and '1B' are the corresponding comparison sites: Comparison site 1A and Comparison site 1B, respectively. See Table 3.1 for details on each intervention and comparison site.*

3.4.4 Outcome measure: Systematic observation of physical activity and wellbeing

Systematic observation of activity in each of the intervention and comparison sites will be carried out using MOHAWk [34]: an observation tool that measures three levels of physical activity intensity (Sedentary, Walking and Vigorous) and two behavioural indicators of wellbeing (Connect: connecting with other people; and Take Notice: taking notice of the environment). This tool applies interval time sampling techniques using continuous observation of activities and characteristics of all individuals entering predefined target areas during hour-long observation periods. MOHAWk requires observers to record the following characteristics for all observed participants: age (older adults defined as ≥ 60 years of age), gender, ethnicity and whether they require assistance to move. Observers also document weather conditions (precipitation) and incivilities in the target area (e.g. general litter, graffiti). MOHAWk is a newly developed tool, for which there is preliminary evidence of validity [34].

MOHAWk was chosen because existing validated tools for systematic observation of physical activity, such as SOPARC (System for Observing Play and Recreation in Communities) [39], use momentary scans of activity and were developed for outdoor environments that attract consistently high numbers of users or large groups (e.g. large regional parks). The planned study will evaluate small outdoor environments that have lower numbers of users and less consistent usage; thus momentary scans would be unable to reliably capture people's activity within or passing through that space. MOHAWk also measures objective behavioural indicators of wellbeing; wellbeing has predominantly been measured used self-report, which is more susceptible to recall bias and poor response rates [40].

3.4.5 Greening interventions

The following descriptions are in line with the Template for Intervention Description and Replication (TIDieR) checklist [41].

This study is set in GM; a large metropolitan county in North West England with a population of around 2.8 million and containing ten metropolitan boroughs. GM is undergoing rapidly increasing urbanisation. As a result, integrating green space into this highly urbanised county is becoming a political priority for GM [42]. GM is therefore a strong case study for evaluating the effects of changes in urban green space on physical activity and wellbeing.

This study will evaluate four urban green space interventions designed and implemented by Southway Housing Trust: a housing association in GM. Southway Housing Trust state that the aim of the interventions is to increase the number of people actively using specific areas targeted for environmental improvements and improve wellbeing in the local community (P. Reece, personal communication). The intervention sites are located in Old Moat (Figure 3.4): a suburban ward with a population of 14,657 located in the city of Manchester. The population of Old Moat ward is relatively young compared to other wards in Manchester, although there are a high proportion of older adults living in Southway Housing Trust properties [43]. Manchester is ranked as the fifth most deprived local authority in England [44] and Old Moat is ranked as the 22nd most deprived ward in Manchester out of 32 wards [45] (based on the Index of Multiple Deprivation (IMD) Score [46]).

The interventions are located on four publicly accessible sites; the total size of the floor area of green space in each of the intervention sites is small, ranging from 0.09 to 0.35 acres. Components of the interventions include tree and flower planting (expected to bloom by March 2018) and artificial tree decorations such as strings of small electric lights and tree socks (the interventions are hereafter collectively referred to as urban street greening). The total cost for all components across all four intervention sites is approximately £6,000, although this excludes artist fees connected to the project. All components of the interventions were implemented within one week in November 2017 by two arborists, two local artists, staff members at Southway Housing Trust, and local community members from Old Moat and a local school. The Neighbourhood Green Space Tool (NGST) [47], a UK validated environmental audit tool for measuring the quality of green space, will be used to measure the environmental changes at the intervention sites.

A description of the key characteristics and locations of all intervention and comparison sites can be found in Table 3.1 and Figure 3.5, respectively. An example of one of the intervention sites and one of the corresponding comparison sites at baseline are shown in Figure 3.6.

Table 3.1. Key characteristics of all intervention and comparison sites.

Intervention and comparison sites	Location (postcode)	Intervention components	Green space (acres)	LSOA[a]	Population density (persons Ha^{-1})[b]	Intersection density (per $1000m^{-2}$)[c]	IMD[d]	NDVI[e]	Walk Score[f]
Intervention site 1	**M20 3GB**	**2 planted trees; bulb planting**	**0.09**	**Manchester 038C**	**79.50**	**15.37**	**23.08**	**0.38**	**94**
Comparison site 1A	M19 1EN	-	0.05	Manchester 034B	82.58	16.13	29.87	0.47	71
Comparison site 1B	SK2 6DS	-	0.21	Stockport 019D	69.92	16.18	18.02	0.40	63
Intervention site 2	**M20 1FU**	**12 planted trees; bulb planting; string lights**	**0.27**	**Manchester 038A**	**69.77**	**15.54**	**36.84**	**0.37**	**90**
Comparison site 2A	M20 6FE	-	0.24	Manchester 040A	69.85	14.27	38.24	0.44	83
Comparison site 2B	OL6 8HH	-	0.14	Tameside 004C	73.19	16.68	51.14	0.46	54
Intervention site 3	**M20 1GF**	**3 planted trees; string lights; tree socks**	**0.22**	**Manchester 035A**	**80.88**	**15.52**	**47.92**	**0.46**	**92**
Comparison site 3A	M22 9PS	-	0.12	Manchester 050D	78.94	17.36	54.98	0.40	80
Comparison site 3B	OL6 8HW	-	0.19	Tameside 004C	73.19	16.68	51.14	0.46	63
Intervention site 4	**M20 1AQ**	**8 planted trees; bulb planting; string lights; tree socks; information board**	**0.35**	**Manchester 035A**	**80.88**	**15.52**	**47.92**	**0.46**	**48**
Comparison site 4A	M22 9SZ	-	0.17	Manchester 050D	78.94	17.36	54.98	0.40	82
Comparison site 4B	M22 9PU	-	0.13	Manchester 050D	78.94	17.36	54.98	0.40	75

[a] Lower Layer Super Output Area (LSOA): census reporting units containing between 1000 and 3000 individuals
[b] Population density: number of persons per hectare; used as a proxy measure of residential density
[c] Intersection density: the number of 3-way junctions standardised by LSOA area; used as a measure of street connectivity
[d] Index of Multiple Deprivation score (IMD) [46]: an area deprivation score that combines several indicators of deprivation including income, employment, health and crime. Higher scores indicate more deprived areas.
[e] Normalised Difference Vegetation Index (NDVI): a validated normalised scale of healthy vegetation cover; used as a measure for the presence of greenery at the neighbourhood-level. Higher scores indicate areas with more healthy vegetation cover.
[f] WalkScore uses a Google search algorithm to calculate a weighted score (1-100) based on the number and accessibility of amenities (such as shops and parks) within a 1-mile radius of a user-entered postcode, whereby closer amenities with the most accessible walking routes are weighted more strongly; used as a measure of ‘access to/ availability of destinations and services’. Higher scores indicate more ‘walkable’ areas.

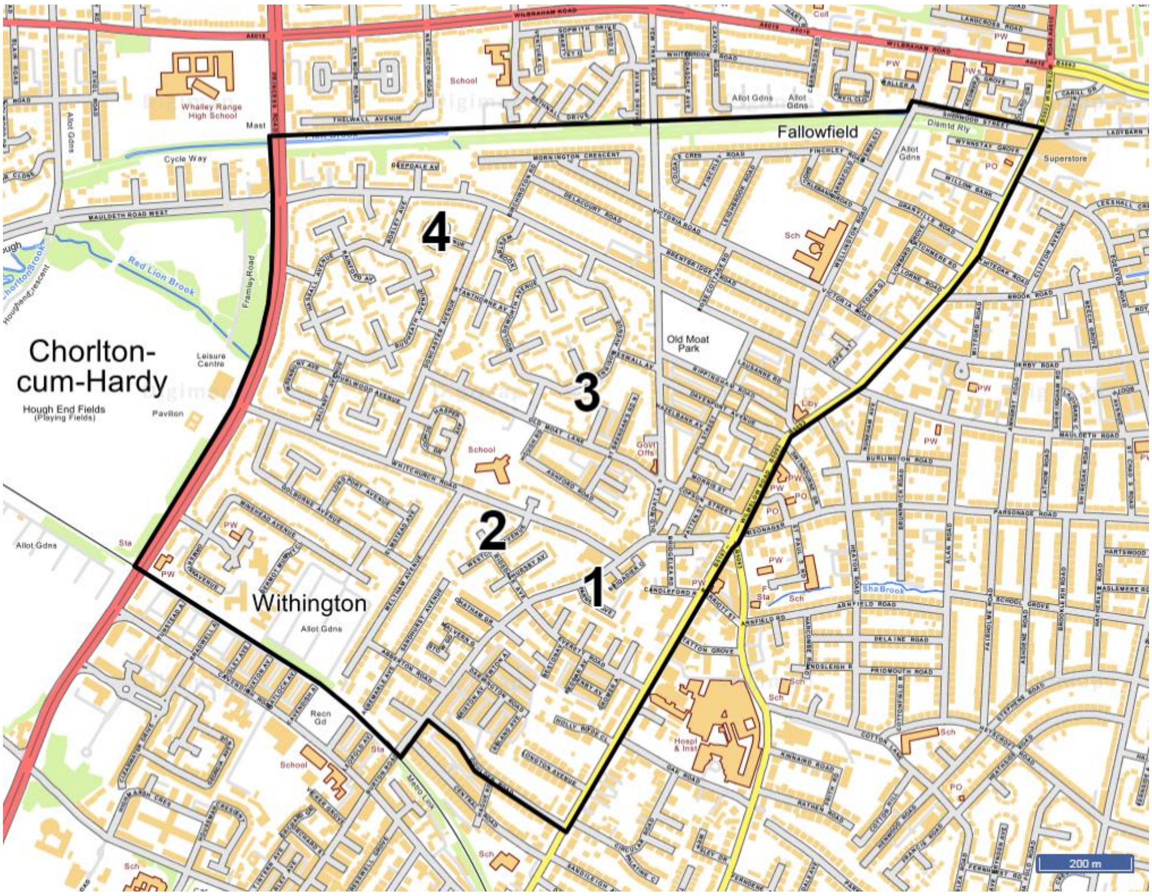


Figure 3.4. Map showing the boundary of Old Moat and location of all intervention sites.

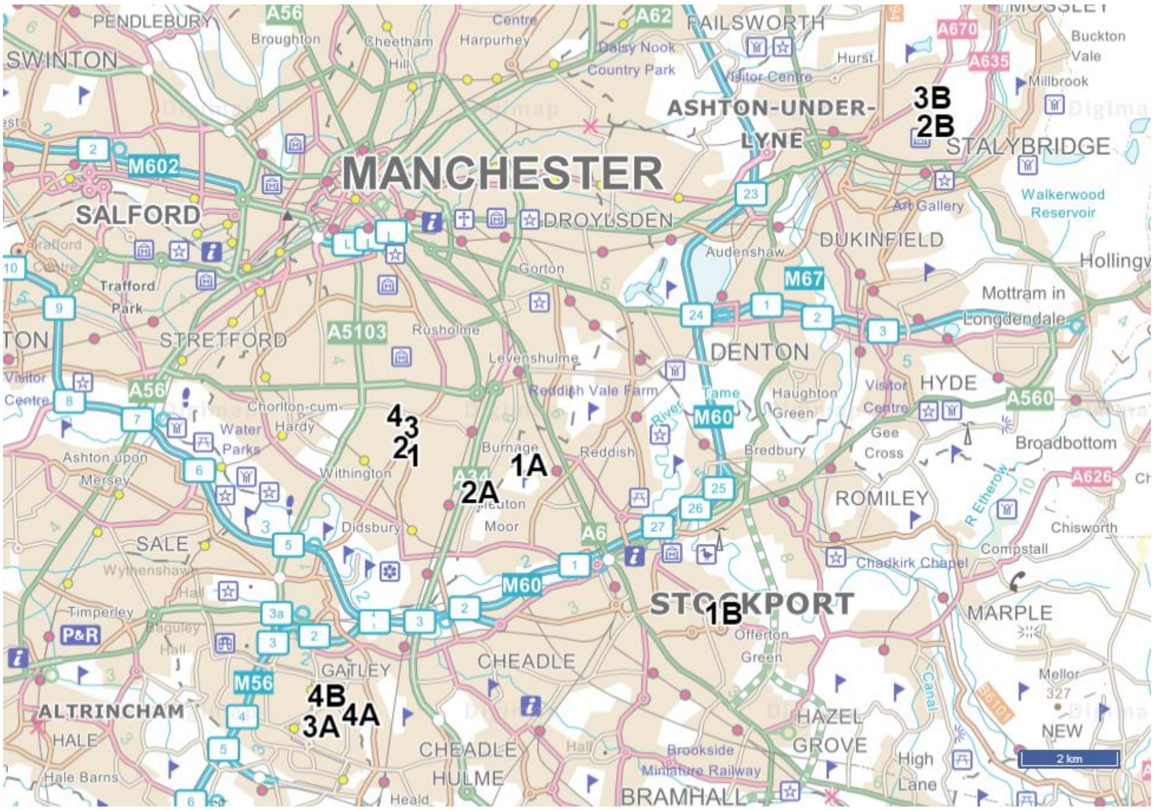


Figure 3.5. Map showing the location of all intervention and comparison sites in Greater Manchester. *The numbers/ letters refer to intervention and comparison sites e.g. '1' is Intervention site 1, and '1A' and '1B' are the corresponding comparison sites: Comparison site 1A and Comparison site 1B, respectively. See Table 3.1 for details on each intervention and comparison site. © Crown Copyright and Database Right (2018). Ordnance Survey (Digimap Licence).*

Figure 3.6. Intervention site 1 (top) and Comparison site 1A (bottom) at baseline.

Photographs taken by Jack Benton.

3.4.6 Logic model

Hypothesised causal pathways are outlined in the logic model (Figure 3.7), which is based on the framework suggested by Panter et al. [48]. It is proposed that improvements in the aesthetic quality of green space will increase overt appreciation in the intervention sites. The interventions are hypothesised to influence only one known variable associated with physical activity: aesthetic quality of the route. More aesthetically pleasing streetscapes, including the presence of attractive and well-maintained trees and greenery, are valued by older adults in facilitating physical activity [38]. The artificial tree decorations (e.g. tree socks) will be present all year round, ensuring the presence of the intervention in colder seasons.

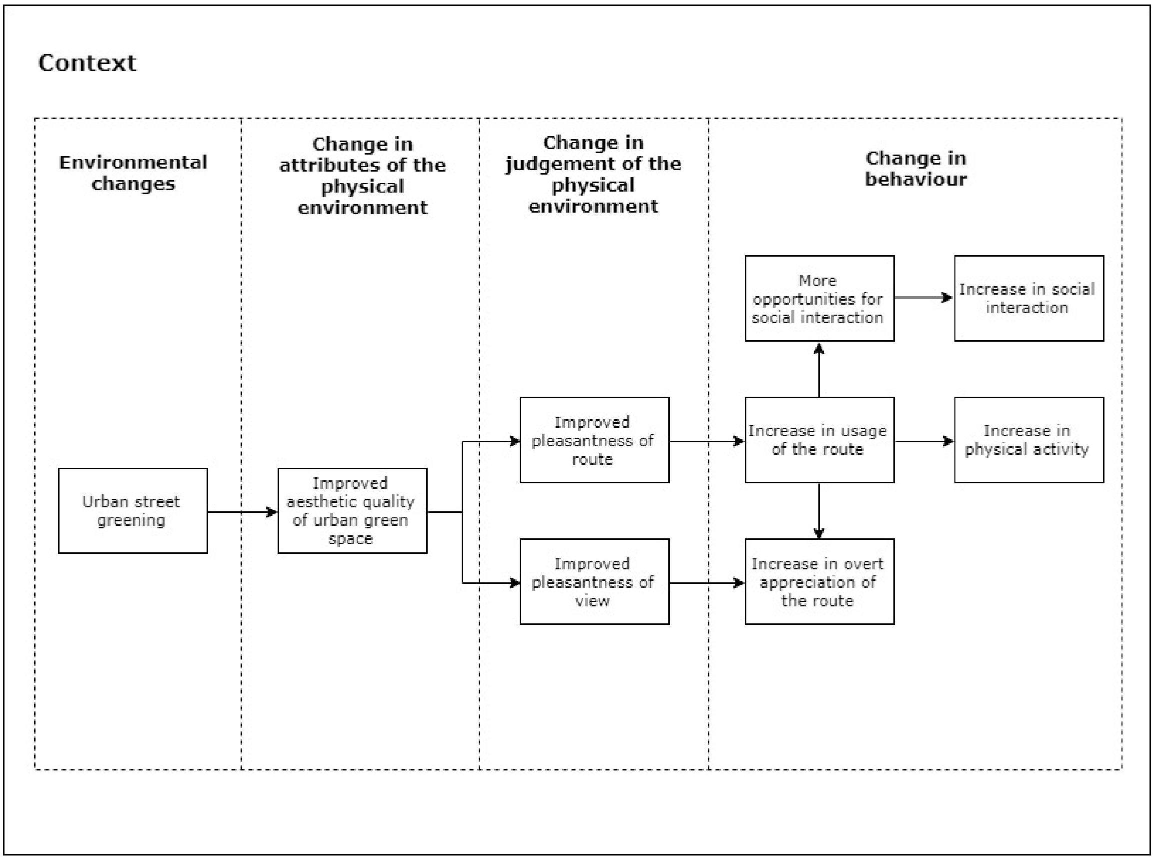


Figure 3.7. Urban street greening logic model.

3.4.7 Confounders

3.4.7.1 Weather

Observations will be carried out regardless of weather conditions, unless weather conditions become so extreme that they compromise the observer's safety. To control for the confounding influence of weather, the observer will record the duration of any precipitation that occurs during each observation period. These data will inform a sensitivity analysis.

3.4.7.2 Other known confounders

Data collection at all time points will be carried out during UK school term dates to control for the change of activity around school holidays. Data collection is planned so that it does not overlap with the daylight saving clock change on 25th March 2018. A liaison from Southway Housing Trust will be contacted before data collection at each time point to enquire about any unrelated significant events that could influence the outcomes, such as other unrelated planned changes to the built environment. Media outlets (e.g. Twitter) will also be monitored to check for any significant events near to sites during data collection periods at each time point. Where possible, data collection will be arranged so that it does not co-occur with any unrelated significant events.

3.4.8 Analysis plan

3.4.8.1 Inter-rater reliability

High inter-rater reliability for MOHAWk has previously been established between pairs of observers across three studies for assessing people's behaviours and their characteristics (ICCs > 0.8) [34]. Inter-rater reliability will be calculated between each pair of observers at each time point to assess agreement on demographic and activity categories. Inter-rater reliability will be analysed using single rater two-way random effects ICC and percentage agreement.

3.4.8.2 Main analysis

The analysis will estimate the effect of the intervention on the number of older adults using the site at each of the two time points, and the types of behaviours they engage in per observation period, compared to the two corresponding comparison sites. The analysis will control for known confounders, including weather, day and time. The primary analysis will analyse data for older adults, and a secondary analysis will analyse data for all adults using the same methods.

The primary outcome will be a count per observation period of Take Notice behaviour at 12 months. Take Notice behaviour is the primary outcome because the interventions are expected to improve the aesthetic quality of green space by providing visual information of value to users, thus causing more overt appreciation in the intervention sites. The secondary outcome will be the overall count of older adults per observation period. Additional exploratory analyses will assess a count per observation period separately for each physical activity level (Sedentary, Walking, Vigorous) and Connect behaviour.

For each outcome we will follow three steps. Firstly, using a dataset that only includes the baseline data, and is blinded to group allocation, we will build a regression model to examine the relationship between the baseline count outcome and the covariates (weather, day, time). The overall count of older adults per observation period will be used as an additional covariate when analysing each of the behaviours (i.e. Sedentary, Walking, Vigorous, Take Notice, Connect). We will consider this count outcome as either a Poisson distribution, a zero-inflated Poisson or a normal distribution. We will consider all these approaches, and we will choose the most suitable approach by seeing which model has the best fit, assessed using Akaike's information criterion (AIC) and measures of over-dispersion. Secondly, once we have developed a suitable model with the baseline data, we will combine the pre- and post-intervention data and apply the model from step one.

Thirdly, we will add into the model Group (intervention or comparison) and Period (pre or post). The treatment effect will be the coefficient for the interaction of Group and Period. This is a form of Difference in Differences analysis [49,50].

3.4.8.3 Sensitivity analysis

A sensitivity analysis will be conducted to assess for any potential bias in the analysis of the primary and secondary outcome.

Observation periods will be removed for the sensitivity analysis if there is any precipitation that lasts for more than 50% of the observation period i.e. an overall accumulated duration of 30 minutes or more (recorded by the observer). This is in line with recommendations from MOHAWk [34].

Analysis will be undertaken using SPSS version 23 or later.

3.4.9 Power calculation

Given the lack of previous studies of the causal effects of urban street greening, it is difficult to estimate the plausible size of the effect that interventions will have on the outcomes. However, the interventions only target one known variable that can influence physical activity and wellbeing within the context of a broader complex 'system' [51]; therefore, the effects on the measured outcomes, particularly physical activity, are likely to be small.

To assess the power of the study, we conducted a power calculation of the primary outcome measure: counts per observation period (hour) of Take Notice behaviour in older adults. We used the approach suggested by Donner and Klar [52, p.66] for calculating the sample size for a matched pair design: calculate the number of clusters required for a completely randomised cluster design, and then multiply that by one minus the correlation between the mean outcomes in the two groups. This suits our context because it allows us to account for multiple comparison groups for each intervention group. We used the means and standard deviations (SD) from the two sites in the feasibility study. We assumed that one site was the 'comparison' group (i.e. the intervention site at baseline) and the other site was the 'intervention' group (this site had two benches, more greenery and was thus more aesthetically appealing based on ratings using the NGST). For the SDs for each site, we used the 80% upper one-sided confidence limit of the SD to account for the possibility that the SD from the feasibility data may be an underestimate.

A power calculation showed that if we match four intervention sites and eight comparison sites (12 in total) and have eight observation periods per cluster, we will have 99% power ($p = .05$, two-tailed test) to detect a difference between 0.5 ($SD = .77$) counts per observation period of Take Notice behaviour in older adults in the control group and 1.6 ($SD = 1.72$) in the intervention group. This assumes the ICC is 0.02, $p = .05$, two-tailed test: this low ICC value has been used in previous studies that have evaluated homogeneous parks (B. Han, personal communication). This also assumes that the 'comparison' group and 'intervention' groups from the feasibility study accurately represent the comparison and intervention sites in the main study at follow-up. However, due to the lack of data from previous studies on changes in older adult's Take Notice behaviours in urban street greening interventions, data from the feasibility data was the most suitable available data to inform the power calculation.

3.5 Discussion

This natural experimental study permits a rare and valuable opportunity to evaluate the causal effects of 'real life' changes in four small urban green spaces, within a deprived neighbourhood in an understudied population (older adults) and setting (UK). The findings will be useful for policy- and decision-makers in GM, as well as other urban areas in the UK and elsewhere in Europe.

This study will provide important methodological contributions by addressing seven out of eight key methodological weaknesses identified in a recent review to reduce bias in natural experimental studies [24] based on MRC guidance [20]. Bias due to confounding is a particularly pervasive problem in natural experimental studies [20]. We reduced the risk of bias due to confounding by developing a rigorous approach to comparison site matching and using appropriate statistical analyses to control for important known potential confounders (e.g. weather). Strengths of our novel approach to comparison site matching include the use of several objective and subjective variables at different levels of the environment (neighbourhood and street level) and multiple comparison sites to increase the likelihood of finding balanced comparison groups. Other methodological improvements in the present study include a published study protocol with a priori analyses specified, pioneering a newly developed tool to objectively measure physical activity and other wellbeing-related behaviours, clear reporting of interventions in line with the TIDieR checklist and a power calculation.

There is a lack of evidence and understanding on what specific kinds of changes in urban green space produce which outcomes in different contexts [48] i.e. what works, for whom and in what circumstances? [53]. The present study will address this issue by formally measuring the specific environmental change (i.e. aesthetic quality) using a UK validated tool for measuring the quality of green space (NGST). Using an objective outcome measure that is directly measured within the environmental context of interest will enable us to more confidently attribute changes in outcomes to the environmental change. This study will therefore provide an accurate insight into the effects of urban street greening on older adults' physical activity and wellbeing in a deprived urban neighbourhood in the UK i.e. helping to answer what works, for whom and in what circumstances.

Understanding the pathways underlying the potential link between changes in urban green space (exposure) and physical activity and wellbeing (outcome) is important from a theoretical point of view, but also in terms of translating evidence into intervention or policy change [54]. However, there has been limited consideration and measurement of how changes in urban green space may work to change behaviour, particularly physical activity, in existing studies to date [15]. We developed a logic model explaining how the interventions are expected to influence older adults' physical activity and wellbeing. We will seek to conduct a qualitative process analysis testing this logic model in a separate sub-study.

This study will use a parallel-groups design with a binary distinction between exposed intervention and unexposed comparison groups. A common difficulty when using this type of design is finding equally matched comparison sites. However, this design is suitable for the present study because intervention sites are located on residential streets; a type of land-use frequently found across the majority of neighbourhoods, thus providing ample unexposed potential matches. An alternative type of comparison site that can be used in natural experimental studies involves graded measures of exposure [55], such as distance from the intervention, as recommended by MRC guidance [20]. However, the interventions in this study are small; it is therefore unlikely that there will be any meaningful variation in exposure outside of the intervention site and thus graded measures of exposure would be less suitable.

3.6 Conclusion

This study permits a rare opportunity to carry out a natural experimental study in the 'real world' as part of a multi-sectoral interdisciplinary collaboration. This study will also demonstrate the feasibility of incorporating rigorous methodology into the challenging field of natural experimental studies. As a result, this study will produce unique robust evidence on the causal effect of changes to urban green space in the UK on physical activity and wellbeing, in a growing but understudied ageing population in the context of environmental interventions.

3.7 References

1. Taylor AH, Cable NT, Faulkner G, Hillsdon M, Narici M, Van Der Bij AK. Physical activity and older adults: a review of health benefits and the effectiveness of interventions. J Sports Sci. 2004;22(8):703-25.
2. UK Department of Health. Start active, stay active. A report on physical activity for health from our four home countries' chief medical officer. UK Department of Health London. 2011. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216370/dh_128210.pdf
3. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. The Lancet. 2012;380(9838):247-57.
4. NHS Information Centre for Health and Social Care. Health survey for England – 2012: Physical activity and fitness. Leeds: NHS. 2014. Available from: https://digital.nhs.uk/catalogue/PUB13218
5. Chatterji S, Byles J, Cutler D, Seeman T, Verdes E. Health, functioning, and disability in older adults—present status and future implications. The Lancet. 2015;385(9967):563-75.
6. Murray JM, Brennan SF, French DP, Patterson CC, Kee F, Hunter RF. Effectiveness of physical activity interventions in achieving behaviour change maintenance in young and middle aged adults: A systematic review and meta-analysis. Soc Sci Med. 2017;192:125-33.
7. Hobbs N, Godfrey A, Lara J, Errington L, Meyer TD, Rochester L, White M, Mathers JC, Sniehotta FF. Are behavioral interventions effective in increasing physical activity at 12 to 36 months in adults aged 55 to 70 years? A systematic review and meta-analysis. BMC Med. 2013;11(1):75.

8. National Institute for Health and Clinical Excellence (NICE). Behaviour change: general approaches. NICE. 2007. Available from: https://www.nice.org.uk/guidance/ph6/resources/behaviour-change-general-approaches-pdf-55457515717
9. Wahl HW, Iwarsson S, Oswald F. Aging well and the environment: Toward an integrative model and research agenda for the future. The Gerontologist. 2012;154:306-16.
10. Jackson RJ, Dannenberg AL, Frumkin H. Health and the built environment: 10 years after. Am J Public Health. 2013;103(9):1542-4.
11. Barnett DW, Barnett A, Nathan A, Van Cauwenberg J, Cerin E. Built environmental correlates of older adults' total physical activity and walking: a systematic review and meta-analysis. Int J Behav Nutr Phys Act. 2017;14(1):103.
12. Lachowycz K, Jones AP. Greenspace and obesity: a systematic review of the evidence. Obes Rev. 2011;12(5):183-9.
13. Schipperijn J, Bentsen P, Troelsen J, Toftager M, Stigsdotter UK. Associations between physical activity and characteristics of urban green space. Urban Forestry & Urban Greening. 2013;12(1):109-16.
14. Giles-Corti B, King AC. Creating active environments across the life course: "thinking outside the square". Br J Sports Med. 2009;43(2):109-13.
15. Hunter RF, Christian H, Veitch J, Astell-Burt T, Hipp JA, Schipperijn J. The impact of interventions to promote physical activity in urban green space: a systematic review and recommendations for future research. Soc Sci Med. 2015;124:246-56.
16. Van Cauwenberg J, De Bourdeaudhuij I, De Meester F, Van Dyck D, Salmon J, Clarys P, Deforche B. Relationship between the physical environment and physical activity in older adults: a systematic review. Health & Place. 2011;17(2):458-69.

17. Lachowycz K, Jones AP. Towards a better understanding of the relationship between greenspace and health: development of a theoretical framework. Landsc Urban Plan. 2013;118:62-9.

18. Lee AC, Maheswaran R. The health benefits of urban green spaces: a review of the evidence. J Public Health. 2010;33(2):212-22.

19. World Health Organization. Urban green space interventions and health: A review of impacts and effectiveness. WHO. 2017. Available from: http://www.euro.who.int/__data/assets/pdf_file/0010/337690/FULL-REPORT-for-LLP.pdf?ua=1

20. Craig P, Cooper C, Gunnell D, Hajow S, Lawson K, Macintyre S, Ogilvie D, Petticrew M, Reeves B, Sutton M, Thompson S. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. J Epidemiol Community Health. 2012;66(12):1182-6.

21. Cerin E, Nathan A, Van Cauwenberg J, Barnett DW, Barnett A. The neighbourhood physical environment and active travel in older adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act. 2017;14(1):15.

22. Handy S. Does the Built Environment Influence Physical Activity: Examining the Evidence – TRB Special Report 282. Transportation Research Board. 2005. Available from: http://www.onlinepubs.trb.org/onlinepubs/sr/sr282.pdf

23. Tucker P, Gilliland J. The effect of season and weather on physical activity: a systematic review. Pub Health. 2007;121(12):909–22.

24. Benton JS, Anderson J, Hunter RF, French DP. The effect of changing the built environment on physical activity: a quantitative review of the risk of bias in natural experiments. Int J Behav Nutr Phys Act. 2016;13(1):107.

25. MacMillan F, George ES. Feng X, Merom D, Bennie A, Cook A., … Astell-Burt T. Do Natural Experiments of Changes in Neighborhood Built Environment Impact

Physical Activity and Diet? A Systematic Review. Int J Environ Res Public Health. 2018;15(2):217.

26. Roberts H, McEachan R, Margary T, Conner M, Kellar I. Identifying Effective Behavior Change Techniques in Built Environment Interventions to Increase Use of Green Space: A Systematic Review. Environ Behav. 2016; 1-28.
27. Anderson J, Ruggeri K, Steemers K, Huppert F. Lively Social Space, Well-Being Activity, and Urban Design Findings From a Low-Cost Community-Led Public Space Intervention. Environ Behav. 2016.
28. Bowler DE, Buyung-Ali LM, Knight TM, Pullin AS. A systematic review of evidence for the added benefits to health of exposure to natural environments. BMC Public Health. 2010;10(1):456.
29. Morris N. Health, well-being and open space: literature review. Edinburgh: Edinburgh College of Art and Heriot-Watt University. 2003.
30. van den Berg M, Wendel-Vos W, van Poppel M, Kemper H, van Mechelen W, Maas J. Health benefits of green spaces in the living environment: A systematic review of epidemiological studies. Urban Forestry & Urban Greening. 2015;14(4):806-16.
31. Clark C, Myron R, Stansfeld S, Candy B. A systematic review of the evidence on the effect of the built and physical environment on mental health. J Public Ment Health. 2007;6(2):14-27.
32. Anderson, J., Ruggeri, K., Steemers, K and Huppert, F.A. The need for a well-being science of cities: mixed findings from a mixed methods study in multiple prize-winning neighbourhoods. In preparation.
33. New Economics Foundation. Five ways to wellbeing: The evidence. London, England: New Economics Foundation. 2008. Available from: http://www.neweconomics.org/projects/entry/five-ways-to-well-being
34. Benton JS, Anderson J, Cotterill S, Hunter RF, Pulis M, French DP. Method for Observing pHysical Activity and Wellbeing (MOHAWk): development of an

observation tool to measure physical activity and wellbeing in urban spaces. Under development.

35. Copeland JL, Esliger DW. Accelerometer assessment of physical activity in active, healthy older adults. J Aging Phys Act. 2009;17(1):17-30.
36. Davis MG, Fox KR, Hillsdon M, Sharp DJ, Coulson JC, Thompson JL. Objectively measured physical activity in a diverse sample of older urban UK adults. Med Sci Sports Exerc. 2011;43(4):647-54.
37. Grant PM, Granat MH, Thow MK, Maclaren WM. Analyzing free-living physical activity of older adults in different environments using body-worn activity monitors. J Aging Phys Act. 2010;18(2):171-84.
38. Moran M, Van Cauwenberg J, Hercky-Linnewiel R, Cerin E, Deforche B, Plaut P. Understanding the relationships between the physical environment and physical activity in older adults: a systematic review of qualitative studies. Int J Behav Nutr Phys Act. 2014;11(1):79.
39. McKenzie TL, Cohen DA, Sehgal A, Williamson S, Golinelli D. System for Observing Play and Recreation in Communities (SOPARC): reliability and feasibility measures. J Phys Act Health. 2006;3 Suppl 1:208-22.
40. National Research Council. Subjective well-being: Measuring happiness, suffering, and other dimensions of experience. Washington, DC: The National Academies Press; 2013.
41. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, Altman DG, Barbour V, Macdonald H, Johnston M, Lamb SE. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ. 2014;348:1756-1833.
42. TEP. Towards a green infrastructure framework for Greater Manchester: Full Report. Manchester: Association of Greater Manchester Authorities. 2008. Available from:

http://www.greeninfrastructurenw.co.uk/resources/1547.058_Final_Report_September_2008.pdf

43. Phillipson C, White S, Aftab F, Hammond M. Old Moat: Age-Friendly Neighbourhood Report. Southway Housing Trust. 2013. Available from: http://www.southwayhousing.co.uk/getattachment/Southway-in-Your-Community/Age-friendly-Neighbourhoods/Old-Moat-Age-friendly-Neighbourhood-Report.pdf.aspx
44. Manchester City Council. Old Moat Ward Electoral Registration Summary. 2014. Available from: https://www.manchester.gov.uk/info/500330/the_electoral_register
45. Manchester City Council. F1 Indices of Deprivation 2015. 2015. Available from: http://www.manchester.gov.uk/downloads/download/414/research_and_intelligence_population_publications_deprivation
46. Department for Communities and Local Government. English Indices of Deprivation 2015 [computer file]. 2015. Downloaded from: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015. Licensed under: https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3. Accessed 30 Oct 2017.
47. Gidlow CJ, Ellis NJ, Bostock S. Development of the neighbourhood green space tool (NGST). Landscape and Urban Planning. 2012;106(4):347-58.
48. Panter J, Guell C, Prins R, Ogilvie D. Physical activity and the environment: conceptual review and framework for intervention research. Int J Behav Nutr Phys Act. 2017;14(1):156.
49. Blundell R, Costa Dias M. Evaluation methods for non-experimental data. Fisc Stud. 2000;21(4):427-68.
50. Craig P, Katikireddi SV, Leyland A, Popham F. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. Annu Rev Public Health. 2017;38:39-56.

51. Hawe P, Shiell A, Riley T. Theorising interventions as events in systems. Am J Community Psychol. 2009;43(3-4):267-76.

52. Donner A, Klar N. Design and Analysis of Cluster Randomization Trials in Health. London: Arnold Publishers; 2000.

53. Pawson R, Tilley N: Realistic evaluation. London: Sage; 1997.

54. Wells NM, Ashdown SP, Davies EH, Cowett FD, Yang Y. Environment, design, and obesity: Opportunities for interdisciplinary collaborative research. Environ Behav. 2007;39(1):6-33.

55. Humphreys DK, Panter J, Sahlqvist S, Goodman A, Ogilvie D. Changing the environment to improve population health: a framework for considering exposure in natural experimental studies. J Epidemiol Community Health. 2016;70(9):941-6.

Chapter 4. Impact of a low-cost urban green space intervention on wellbeing behaviours in older adults: a natural experimental study

This article has been submitted for publication in Wellbeing, Space & Society and is currently under peer review:

- Benton JS, Cotterill S, Anderson J, Macintyre VG, Gittins M, Dennis M, Lindley SJ, French DP. Impact of a low-cost urban green space intervention on wellbeing behaviours in older adults: a natural experimental study. Wellbeing, Space & Society. Under review.

4.1 Abstract

This study assessed the impact of low-cost physical changes to four urban amenity green spaces on older adults' wellbeing behaviours (people taking notice of the environment, physical activity levels, social interactions), relative to eight matched comparison sites in Greater Manchester, United Kingdom. Systematic observations were conducted at baseline, six- and twelve months, with a nested qualitative study. There was no evidence that the intervention increased observable wellbeing behaviours or use of these spaces. This study demonstrates the feasibility of robust natural experimental methods and highlights the need for more nuanced policy and practice recommendations for improving urban green spaces.

4.2 Background

Worldwide, the proportion of older adults in the population is growing rapidly [1]. Older adults (≥ 60 years old) currently represent approximately 12% of the global population, but this is expected to rise to 20% by 2050 [2]. This increased life expectancy raises the importance of promoting healthy ageing to extend quality of life [3].

Improving or creating urban green spaces has substantial potential to improve health and wellbeing in all age groups [4]. Urban green spaces are urban open spaces wholly or partly covered by vegetation and/ or water, including dedicated recreational spaces such as urban parks, but also other types of green infrastructure within the wider urban fabric such as street trees. Growing urbanisation increases the importance of urban green spaces in supporting health and wellbeing [5].

For all age groups, including older adults, the use of urban green spaces may promote wellbeing via various behavioural pathways. For example, exposure to urban green spaces has been associated with people engaging more with nature (e.g. watching wildlife, smelling wildflowers) [6], which can have psychological and wellbeing benefits due to higher nature connectedness [7] and improved restoration [8]; these benefits have been shown to be greater in older age groups [9,10]. Use of urban green spaces has been associated with increased social interactions [11], which is particularly important for older adults as social isolation is a major health problem in older adults [12]. Urban green spaces can also provide settings for people to engage in physical activity, which is pertinent to older adults who are the least physically active age group [13].

Despite these numerous plausible behavioural pathways by which urban green spaces might influence wellbeing, the majority of research to date has focused on physical activity. Urban green space availability, accessibility and usage have been associated with increased physical activity in all age groups [14], including older adults [15]; however, most of this evidence is cross-sectional. Randomised controlled trials (RCTs) of urban green space interventions are often not feasible because researchers do not have control of the environment. Therefore, making use of natural experiments has become a priority to investigate the causal effects of urban green spaces on physical activity [14]. Natural experiments are real world events or interventions that are not under the control of researchers. Researchers can design studies around a natural experiment; known as a 'natural experimental study' [16].

Recent systematic reviews of natural experimental studies suggest that urban green space interventions can have positive effects on physical activity [17–22]. Such ostensibly effective interventions range from low-cost interventions like street greening (e.g. tree planting) to more substantial interventions like new greenways and major park renovations.

However, natural experimental studies in this area are scarce and often have a high risk of bias; that is, a high risk of underestimating or overestimating the true intervention effect due to flaws in the study design, conduct, analyses and/ or reporting [23]. Internal validity is the extent to which a study is free from bias [24]. A recent review of systematic reviews assessed the risk of bias in natural experimental studies of built environment interventions on physical activity [25]. Using three existing exemplar systematic reviews, they identified only 12 controlled natural experimental studies (15 physical activity outcomes), including nine studies of urban green space interventions. They found that all studies had outcomes that had critical (n = 12) or serious (n = 3) risk of bias, which shows that even the strongest studies conducted to date have substantial risk of bias; a conclusion in line with other recent systematic reviews [17,19,20,26].

Eight recommendations have been provided to improve the internal validity of future natural experimental studies in this field [25], influenced by the Medical Research

Council (MRC) guidance for using natural experiments [16]. The present study implemented these recommendations:

1. Publish study protocols with a priori analyses specified;
2. Conduct sample size calculations;
3. Better matching of comparison sites;
4. Use multiple comparison sites;
5. Use adequate outcome measurements;
6. Measure intervention exposure to the intervention at the individual level;
7. Control for confounders in statistical analyses;
8. Better reporting of samples and interventions.

Aside from risk of bias, there are other limitations with existing natural experimental studies of urban green space interventions in this area. There is a shortage of studies among older adults; three recent systematic reviews of studies of the physical environment on physical activity among older adults [15,27,28] found just one natural experimental study [29]. There is also a dearth of studies in the UK and the rest of Europe; most studies are in the United States (US) [14]. There are many contextual differences between the UK and US that influence physical activity which makes it difficult to generalise findings from the US to the UK, such as climate, land use, population density, and physical activity patterns [30]. Finally, similar to physical activity research, there is a scarcity of robust natural experimental studies of urban green space interventions on other behaviours important for wellbeing [31].

Systematic observation (i.e. direct observations of behaviour using pre-determined criteria) is a promising objective method of unobtrusively assessing a range of behaviours that are important for wellbeing. On behalf of the UK Government's Foresight project, New Economics Foundation (NEF) conducted a review of the wellbeing literature and identified five behaviours for which there was good evidence that engaging in these behaviours improves wellbeing ('Five Ways to Wellbeing') [32]: Be Active (engage in physical activity); Take Notice (be aware of the environment); Connect (socially interact with others); Keep Learning (acquire knowledge or skill in something new); Give (contribute to the community). Despite empirical support for the Five Ways to Wellbeing, there is a lack of intervention research that has examined how to effectively promote these behaviours [33]. The present study uses a newly validated observation tool that quantitatively assesses three of the Five Ways behaviours (Be Active, Take Notice, Connect) - MOHAWk: Method for Observing pHysical Activity and Wellbeing [34].

4.2.1 Research aim and objectives

The present study took advantage of an opportunity for a prospective natural experimental study. A local housing association planned to make low-cost changes to four small urban amenity green spaces in Manchester, UK (hereafter referred to as an 'intervention'). As suggested in several UK policy and practice guidance recommendations [35–39], the housing association believed that increasing the provision of green infrastructure (e.g. planting trees) would be an effective intervention to increase the number of people using the area and improve wellbeing in the local community (Southway Housing Trust Environment Manager, personal communication). Hence, the aims of the intervention provider are reflected in the objectives of this study. This study focused primarily on older adults because the housing association has a strong focus on older adults, and there is a high proportion of older adults living in their properties. The effects of environmental interventions are likely to vary for different age groups (e.g. older adults tend to spend more time in their local living environment compared to younger adults [40]); hence data were also collected on younger adults.

The aims of this study were to develop methodological innovations to address key weaknesses causing high risk of bias in previous natural experimental studies in this area, and to implement these new methods to assess the impact of the intervention on wellbeing behaviours in older adults using systematic observation. Specific objectives were to examine whether the intervention increased the following at 12 months relative to baseline, relative to matched comparison sites where no such changes occurred:

1. Older adults' Take Notice behaviour (primary outcome). The primary outcome is Take Notice behaviour due to the largest effects on this behaviour being anticipated from changes in the aesthetic quality of green space at the intervention sites;
2. The total number of older adults using the sites (secondary outcome);
3. Older adults' physical activity levels (Sedentary, Walking, Vigorous activity) and Connect behaviour (exploratory outcomes).
4. The total number of adults using the sites and wellbeing behaviours in adults.

As recommended by MRC guidance for process evaluation [41], it is important to use qualitative methods as part of a mixed-methods evaluation to understand how people experience the intervention. Given that this was a low-cost intervention without a specific awareness or marketing program, it was particularly important to assess awareness of the intervention. Therefore, a nested qualitative study was conducted to explore local older adults' views and experiences of these small urban green spaces at the intervention sites; these qualitative findings have been published and reported in detail elsewhere [42]. For the purposes of the present study, the qualitative data enabled us to address an additional objective:

5. How many people were aware of the intervention?

4.3 Methods

The methods and analyses for this study have been described in the study protocol [43].

4.3.1 Feasibility study

A feasibility study was carried out in Manchester in July 2017 to inform the natural experimental study [34]. Observations were conducted between 8am-6pm in 50-minute observation periods (e.g. 8–8.50 am, 9–9.50 am etc.) by one observer for seven consecutive days in Intervention Site 1 (Table 4.1). The data were used to inform a power calculation and the frequency and timing of observations in the natural experimental study.

4.3.2 Study design

This was a prospective controlled natural experimental study, with four intervention sites and eight matched comparison sites.

4.3.3 Procedure

4.3.3.1 Comparison site matching

We developed a five-step process of matching (see Figure 4.1). We identified comparison sites that were matched to each intervention site using ten key objective and subjective variables. Variables for matching were identified from one qualitative [44] and two quantitative [15,28] systematic reviews that recently investigated built environment correlates of older adults' physical activity. Two comparison sites were matched to each intervention site to increase statistical power. Each step of this matching process is described in more detail in Appendix G.

4.3.3.2 Data collection time points

Baseline data collection occurred in September 2017 before the intervention was implemented in November 2017. Follow-up data collection occurred in February/ March 2018 (6 months post-baseline) to capture initial short-term effects of the intervention, and

September 2018 (12 months post-baseline) at the same time of year as baseline to control for seasonal variation.

4.3.3.3 Observation schedule and procedure

The newly validated MOHAWk observation tool was used to assess three wellbeing behaviours (people taking notice of the environment, physical activity levels, social interactions) [34]. Informed by the feasibility study, observations for each site were conducted over two weekdays, four times a day at each time point; providing a total of eight observation periods for each site at each time point. Observing on two days, four times a day can provide a reliable estimation of activity [34]. All observation periods were one hour: 10-11am (morning), 12-1pm (lunchtime), 3-4pm (afternoon), and 5-6pm (evening). Days of observation for each intervention site and its two matched comparison sites were the same at all time points e.g. observations for Intervention site 1 and Comparison sites 1A and 1B were conducted on a Monday and Wednesday at all time points. Observations for each site were spread over two weeks, rather than a single week, to provide a more robust assessment of activity over a longer period; observations were therefore counterbalanced to control for week, day of week, and time of day. The same observation schedule was used at all time points.

Four observers were used across the study, with two observers independently collecting data at each time point. Observers were trained using the MOHAWk instruction manual [34] and by practising observations in the study sites. There was one target area per site, and the same target areas were used at all time points (Appendix H).

Agreement between each pair of observers (i.e. inter-rater reliability) was analysed using two-way mixed, single measure, consistency intraclass correlation coefficients (ICC) at each time point. Inter-rater reliability for counts of people, their characteristics and behaviours was ‘good’ or ‘excellent’, with only one ‘moderate’ ICC [45] (range of ICCs in brackets): total number of people (0.96-0.99), age group (0.76-1), gender (0.90-1), ethnicity (0.66-1), Take Notice (1), Sedentary (0.91-1), Walking (0.91-1), Vigorous (0.83-1), and Connect (0.96-1).

4.3.4 Study population

Data were collected on older adults and adults (age groups estimated using observable criteria – see 'Outcome measure' section). Analyses were conducted separately for older adults and adults.

4.3.5 Intervention

This study was set in Greater Manchester: a large metropolitan county in North West England containing ten metropolitan boroughs, including the city of Manchester. The intervention sites were located in Old Moat (Figure 4.2): a ward (i.e. a subsection of a city that is represented by a councillor) with a population of 14,657 (1,174 older adults ≥ 65 years old) located in South Manchester. Old Moat is characterised by a semi-detached 'garden suburb' layout [46]. Manchester is the fifth most deprived local authority in England [47], and Old Moat is the 22nd most deprived ward in Manchester out of 32 wards [48], based on the Index of Multiple Deprivation (IMD) Score [49].

The intervention was implemented on four small public amenity green spaces (0.04 - 0.14 hectares) adjacent to residential streets. These are public spaces that primarily contribute to the appearance of the area rather than being used as informal recreation spaces. The intervention included tree (mostly deciduous fruit) and bulb planting, tree decorations including tree socks (knitted yarn around the trunks of trees) and strings of small electric lights, and an information board (total cost ≈ £6,000) (Figure 4.3). They were designed and funded by Southway Housing Trust: a housing association in Manchester. Southway Housing Trust stated that the aims of the intervention were to increase the number of people actively using the intervention sites and enhance wellbeing in the local community (Southway Housing Trust Environment Manager, personal communication). The intervention was implemented in November 2017 by two arborists, staff members at Southway Housing Trust, local community members from Old Moat and children from a local school.

Key characteristics of all intervention and comparison sites can be found in Table 4.1. An example of one of the intervention sites and one of its comparison sites at baseline are shown in Figure 4.4.

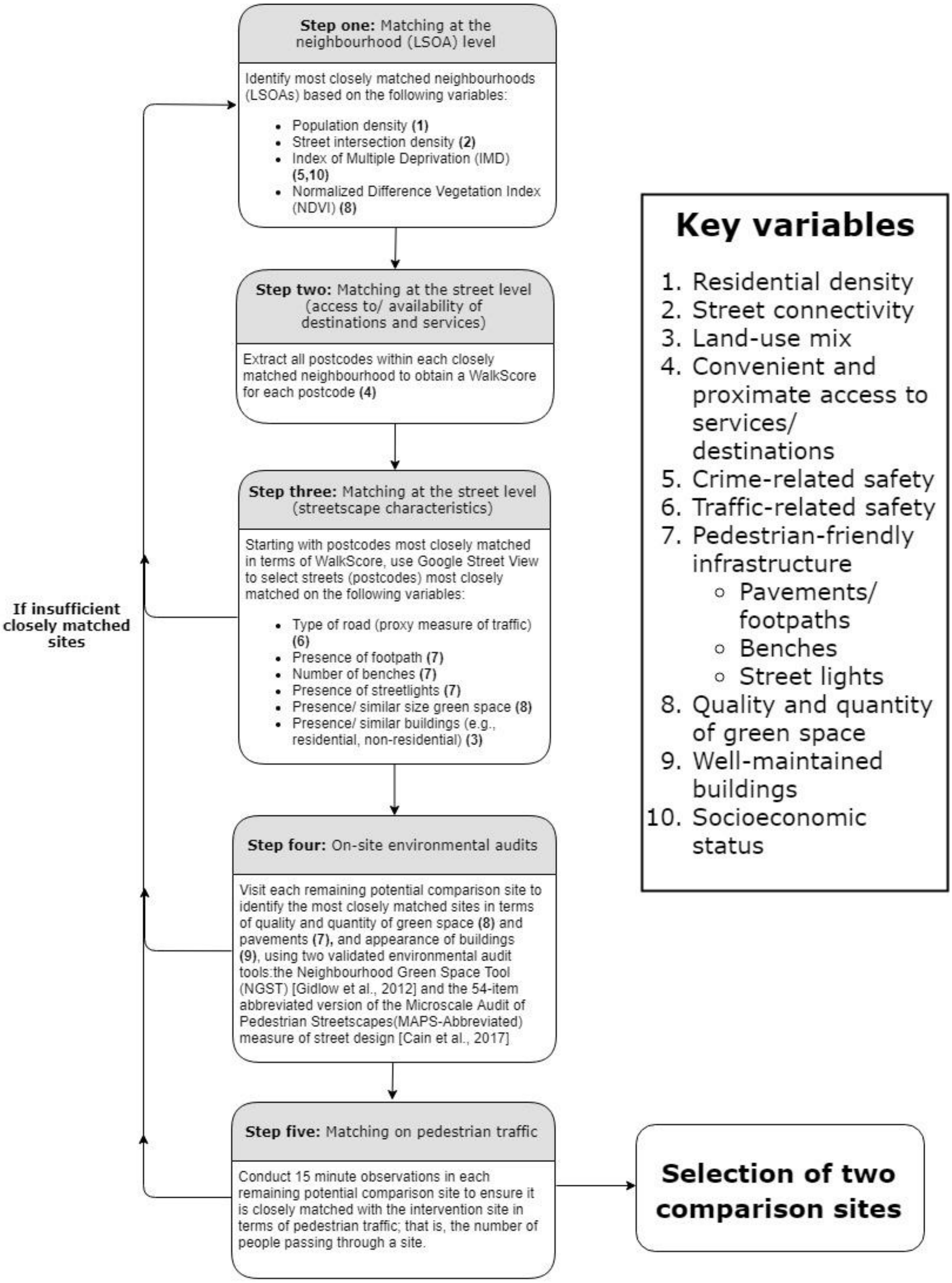


Figure 4.1. Overview of the five-step comparison site matching process. *Numbers in brackets refer to the key variables used for matching. References for the environmental audit tools: Cain et al.* [50] *and Gidlow et al.* [51]. *Image originally published elsewhere* [43].

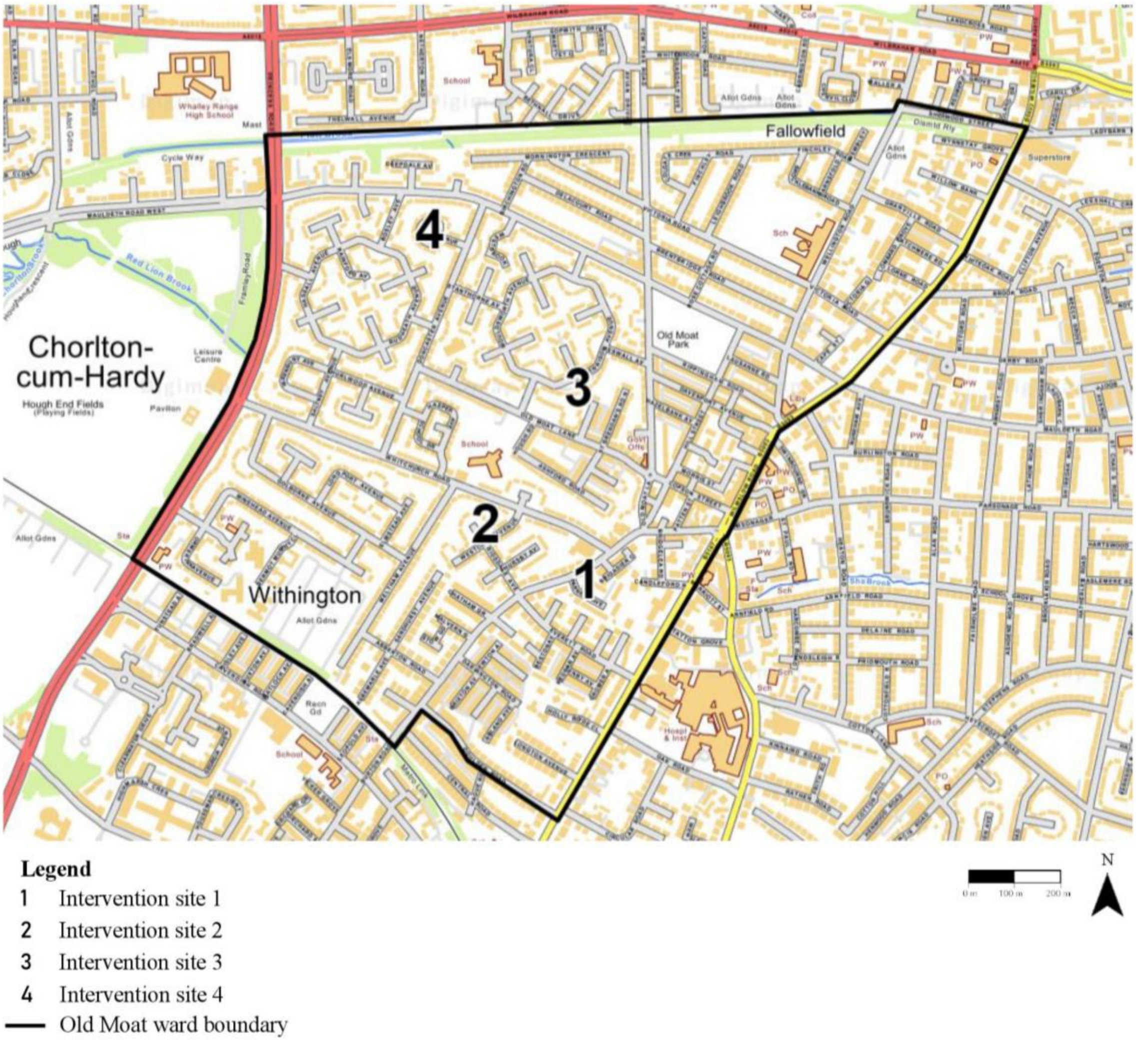


Figure 4.2. Map showing the ward boundary of Old Moat and location of all intervention sites. © *Crown Copyright and Database Right (2018). Ordnance Survey (Digimap Licence). Image originally published elsewhere* [43].

Figure 4.3. Intervention components. *Tree planting (left), tree socks (middle), information board (top right), and string lights (bottom right). Photographs were taken by Jack Benton and Vanessa Macintyre (September 2018).*

Figure 4.4. Intervention site 1 (top) and Comparison site 1A (bottom) at baseline. *Photographs were taken by Jack Benton (September 2017). Originally published elsewhere* [43].

Table 4.1. Key characteristics of all sites and LSOAs.

Site	Postcode	Intervention components	Green space size (hectares)	Walk Score[a]	LSOA [b] characteristics				
					LSOA name	Population density [c]	Intersection density [d]	IMD [e]	NDVI[f]
Intervention site 1	**M20 3GB**	**2 planted trees; bulb planting**	**0.04**	**94**	**Manchester 038C**	**79.50**	**15.37**	**23.08**	**0.38**
Comparison site 1A	M19 1EN	-	0.02	71	Manchester 034B	82.58	16.13	29.87	0.47
Comparison site 1B	SK2 6DS	-	0.08	63	Stockport 019D	69.92	16.18	18.02	0.40
Intervention site 2	**M20 1FU**	**12 planted trees; bulb planting; string lights**	**0.11**	**90**	**Manchester 038A**	**69.77**	**15.54**	**36.84**	**0.37**
Comparison site 2A	M20 6FE	-	0.10	83	Manchester 040A	69.85	14.27	38.24	0.44
Comparison site 2B	OL6 8HH	-	0.06	54	Tameside 004C	73.19	16.68	51.14	0.46
Intervention site 3	**M20 1GF**	**3 planted trees; string lights; tree socks**	**0.09**	**92**	**Manchester 035A**	**80.88**	**15.52**	**47.92**	**0.46**
Comparison site 3A	M22 9PS	-	0.05	80	Manchester 050D	78.94	17.36	54.98	0.40
Comparison site 3B	OL6 8HW	-	0.08	63	Tameside 004C	73.19	16.68	51.14	0.46
Intervention site 4	**M20 1AQ**	**8 planted trees; bulb planting; string lights; tree socks; information board**	**0.14**	**48**	**Manchester 035A**	**80.88**	**15.52**	**47.92**	**0.46**
Comparison site 4A [g]	M22 9SZ	-	0.07	82	Manchester 050D	78.94	17.36	54.98	0.40
Comparison site 4B	M22 9PU	-	0.05	75	Manchester 050D	78.94	17.36	54.98	0.40

[a] WalkScore uses a Google search algorithm to calculate a weighted score (1-100) based on the number and accessibility of amenities (such as shops and parks) within a 1-mile radius of a user-entered postcode, whereby closer amenities with the most accessible walking routes are weighted more strongly; used as a measure of 'access to/ availability of destinations and services'. Higher scores indicate more 'walkable' areas.
[b] Lower Layer Super Output Area (LSOA): census reporting units containing between 1000 and 3000 individuals
[c] Population density: number of persons per hectare; used as a proxy measure of residential density
[d] Intersection density: the number of 3-way junctions standardised by LSOA area; used as a measure of street connectivity
[e] Index of Multiple Deprivation score (IMD) [49]: an area deprivation score that combines several indicators of deprivation including income, employment, health and crime. Higher scores indicate more deprived areas.
[f] Normalised Difference Vegetation Index (NDVI): a validated normalised scale of healthy vegetation cover; used as a measure for the presence of greenery at the neighbourhood-level. Higher scores indicate areas with more healthy vegetation cover.
[g] This site was removed from the study at baseline due to ethical concerns regarding observer safety.

4.3.6 Outcome measure

Outcomes were measured using MOHAWk [34]: a systematic observation tool for assessing people taking notice of the environment (Take Notice), social interactions (Connect) in urban spaces and three levels of physical activity (Sedentary, Walking, Vigorous). There is evidence of high inter-rater reliability when using MOHAWk, and there is evidence of validity for recording Take Notice behaviours using MOHAWk [34]. To date, over 650 hours of MOHAWk data have been collected in 34 unique sites across Greater Manchester and Belfast. MOHAWk is also being used in two other natural experimental studies [52,53]. All MOHAWk materials have recently been published elsewhere [34]. MOHAWk is freely available for use.

Observers used MOHAWk to record the characteristics and behaviours of each person entering a pre-determined boundary ('target area') during hour-long observation periods. Observers used the standardised MOHAWk observation form to record the following data for each person that entered the target area during each observation period: age group (Adult or Older Adult), gender (Female or Male), ethnicity (White or Non-white), taking notice of the environment (Take Notice or No Take Notice), social interaction (Connect or No Connect), and physical activity level (Sedentary, Walking, Vigorous). Observers estimated age groups based on gait, clothing and other physical attributes.

4.3.6.1 Take Notice and Connect behaviours

Take Notice behaviours occur when individuals stop or slow down and appear as if they are making a conscious decision to appreciate their surroundings. Examples of Take Notice behaviours include taking a photograph, extended viewing of the scenery, engaging with wildlife (e.g. touching or smelling flowers), or a pronounced head swivel to look at a specific object, view or person. Take Notice behaviours are akin to the growing positive psychology-based research, which shows that noticing the good things about one's surroundings can improve wellbeing [54,55].

Connect behaviours occur when individuals are engaging or interacting with a person or the people around them in some way. Examples of Connect behaviours include talking, holding hands, hugging or participation in a group activity.

4.3.7 Known confounders

Observations were carried out regardless of weather conditions. To control for potential bias associated with precipitation, observers recorded the duration of any precipitation that occurred during each observation period, as recommended in MOHAWk procedures. Based on the feasibility study, the outcomes are likely to be affected by day of week and time of day.

4.3.8 Logic model

Hypothesised causal pathways were outlined in a logic model [43], based on the framework suggested by Panter et al. [56]. It was proposed that the intervention would improve the aesthetic quality of green space, which would improve pleasantness of view and therefore increase overt appreciation in the intervention sites (i.e. Take Notice behaviour). The improved pleasantness of routes in the intervention sites was predicted to increase the use of these routes, subsequently causing increases in physical activity (i.e. Walking and Vigorous activity) and social interaction (i.e. Connect behaviour).

4.3.9 Sample size

We used the approach suggested by Donner and Klar ([57], p.66) for calculating the sample size for a matched pair design. Due to lack of data from previous studies on Take Notice behaviours, the sample size calculation used data from the feasibility study. Matching four intervention sites and eight comparison sites, with eight observation periods per site, we would have 99% power (p = .05, two-tailed test) to detect a difference between 0.5 (SD =

.77) counts per observation period of Take Notice behaviour in older adults (primary outcome) in the comparison group and 1.6 (SD = 1.72) in the intervention group. We assumed an ICC of 0.02, $p = .05$, two-tailed test: this low ICC has been used in previous studies that have evaluated homogeneous parks (B. Han, personal communication).

4.3.10 Analysis

The unit of analysis for all analyses was at the level of the observation period i.e. counts per observation period per site.

4.3.10.1 Wellbeing behaviours

Counts of the primary outcome measure (older adults' taking notice) and some of the other wellbeing behaviour counts for older adults and adults were very small, making the planned regression analysis inappropriate. We therefore calculated the median and interquartile range (IQR) at baseline and 12 months follow-up for each of the wellbeing behaviours (Take Notice, Sedentary, Walking, Vigorous, Connect) for the intervention and comparison groups, separately for older adults and adults. A Mann-Whitney *U* test was used to determine if there were significant differences in the change in counts of each behaviour at baseline and 12 months between intervention and comparison groups.

4.3.10.2 Total number of people

We estimated the effect of the intervention on the total number of older adults using the site per observation period at 12 months (secondary outcome), compared to comparison sites, controlling for day, time of day and precipitation.

We followed three steps to build a suitable regression model. First, using baseline data only (without group allocation), we built a regression model to examine the relationship between the baseline count outcome and three covariates (day, time of day, precipitation). We used a negative binomial regression model, which is an extension of the Poisson regression model that can adjust for overdispersion [58]. Second, we combined the pre- and post-intervention data and applied the same negative binomial model from step one. Third, we added Intervention Group (intervention or comparison) and Time Point (baseline or follow-up) into the model. The intervention effect was the coefficient for the interaction of Intervention Group and Time Point. This is a form of difference in differences analysis [59].

We obtained model estimates using *menbreg* (i.e. multilevel mixed-effects negative binomial regression) in Stata version 14.1.20. We repeated this analysis separately for the total number of older adults at 12 months, and adults at 6 and 12 months.

4.3.10.3 Sensitivity analysis

To assess for any potential bias due to precipitation, observation periods were removed if precipitation occurred for 50% or more of the observation period (recorded by the observer), as recommended in MOHAWk procedures. We did not conduct sensitivity analyses for the regression models because precipitation was included as a covariate in these analyses.

4.3.11 Nested qualitative study

A nested qualitative study was conducted to explore older adults' views and experiences of small urban green spaces in Old Moat (study published elsewhere [42]). In brief, older adults living in or near to Old Moat were interviewed using walk-along interviews and photo elicitation methods. The participants walked through several small urban green spaces in Old Moat, including the intervention sites.

For the present study, the qualitative data allowed exploration of participants' awareness of the intervention using manifest content analysis [60] of answers to the question: 'There have been a number of recent changes to Old Moat – which of these changes have you noticed?'. The lead researcher independently read and analysed the texts to count the number of participants that reported noticing any of the changes at the intervention sites.

4.4 Results

Comparison site 4A was removed from the study at baseline due to safety concerns for the observer. This incident was reported to the institutional Research Ethics Committee; no further action was required. Eleven sites were included in the analyses: four intervention sites and seven comparison sites. Table 4.2 shows baseline characteristics of the observation periods, sample and outcome measures in the intervention and comparison sites. Appendix I provides a breakdown of the total number of people observed at each site at each time point.

Table 4.2. Baseline description of observation periods, sample and outcome measures.

	Comparison sites (*k* = 7)	**Intervention sites (*k* = 4)**
Observation periods		
Total number of observation periods	56	32
High precipitation – n (% of total)	12 (21.5%)	5 (15.5%)
Sample [a]		
Female – n (% of total)	1161 (52.6%)	691 (47.4%)
Older adults – n (% of total)	307 (13.9%)	120 (8.2%)
White – n (% of total)	1871 (84.8%)	1134 (77.8%)
Wellbeing behaviours		
Take Notice behaviours in older adults – n (median, IQR) [b]	3 (0, 0)	1 (0, 0)
Sedentary behaviours in older adults – n (median, IQR)	18 (0, 1)	5 (0, 0)
Walking behaviours in older adults – n (median, IQR)	287 (3, 5.8)	113 (2, 6)
Vigorous behaviours in older adults – n (median, IQR)	13 (0, 0)	8 (0, 0)
Connect behaviours in older adults – n (median, IQR)	77 (1, 2)	22 (0, 1)
Total number of people		
Older adults – n (median, IQR) [c]	307 (4, 6)	120 (2, 6.5)
10-11am – n (median, IQR)	77 (5.5, 5.25)	41 (2.5, 7.25)

	12-1pm – n (median, IQR)	87 (5.5, 11.5)	23 (0.5, 6.75)
	3-4pm – n (median, IQR)	93 (2.5, 10)	32 (2, 7.75)
	5-6pm – n (median, IQR)	50 (3, 4.25)	24 (3, 5.5)
Adults – n (median, IQR)		1900 (33.9, 19.8)	1338 (22.5, 42.5)
	10-11am – n (median, IQR)	286 (18.5, 10.5)	212 (17.5, 42.25)
	12-1pm – n (median, IQR)	316 (17, 25.25)	283 (16, 69.25)
	3-4pm – n (median, IQR)	831 (28.5, 96)	351 (34, 60.25)
	5-6pm – n (median, IQR)	467 (21, 36.25)	492 (25.5, 124)

[a] Sample includes adults and older adults; [b] Primary outcome; [c] Secondary outcome; All values to one decimal place

4.4.1 Main analysis

4.4.1.1 Primary outcome - Older adults' Take Notice behaviour

At 12 months post-baseline, there was no significant difference in the change of Older Adults' Take Notice behaviour between the intervention group (increased 0.06 per hour) and comparison group (increased 0.05 per hour) (p = .93) (Table 4.3).

4.4.1.2 Secondary outcome – total number of Older Adults

At 12 months post-baseline, there was no significant difference between the intervention and comparison group in the total number of Older Adults using the sites, after controlling for day, time of day and precipitation (95% CI = 0.94 – 1.85, p = 0.11) (Table 4.4).

4.4.1.3 Exploratory outcomes (Older Adults)

As shown in Table 4.3, there were no significant differences in the change of Older Adults' Sedentary (p = .77), Walking (p = .13), Vigorous (p = .97) and Connect (p = .86) behaviour between the intervention group and comparison group at 12 months post-baseline.

4.4.1.4 Exploratory outcomes (Adults)

For the total number of Adults using the sites, there was no significant difference between the intervention and comparison group, after controlling for day, time of day and precipitation (95% CI = 0.72 – 1.00, p = 0.051) (Table 4.4).

As shown in Table 4.3, counts of Adults Walking decreased in the intervention group (1.43 per hour) but increased in the comparison group (4.46 per hour) at 12 months post-baseline; this was the only behaviour that had a significant difference (p = 0.01). However, there were no significant differences in the change of Adults' Take Notice (p = .57), Sedentary (p = .18), Vigorous (p = .21) and Connect (p = .50) behaviour between the intervention group and comparison group at 12 months post-baseline (Table 4.3).

Table 4.3. Median counts of all five wellbeing behaviours at baseline and 12 months.

Outcome		Intervention observation periods (n = 64)					Comparison observation periods (n = 112)					Effect (difference between the change in the two groups)	p-value [a]
		Baseline		12 months		Change in median	Baseline		12 months		Change in median		
		Median (IQR)	Total	Median (IQR)	Total		Median (IQR)	Total	Median (IQR)	Total			
Older adults	Take Notice	0 (0)	1	0 (0)	3	0	0 (0)	5	0 (0)	8	0	0	.93
	Sedentary	0 (0)	5	0 (1)	12	0	0 (1)	18	0 (1)	41	0	0	.77
	Walking	2 (6)	113	2 (4.75)	115	0	3 (5.75)	287	2.5 (5)	223	-0.5	0.5	.13
	Vigorous	0 (0)	8	0 (0)	5	0	0 (0)	13	0 (0)	9	0	0	.97
	Connect	0 (1)	22	0 (1)	22	0	1 (2)	77	1 (2)	76	0	0	.86
Adults	Take Notice	0 (0)	2	0 (0)	11	0	0 (0)	23	0 (0)	25	0	0	.57
	Sedentary	0 (1.75)	24	2 (4.5)	104	2	1 (2)	76	3.5 (5.75)	268	2.5	-0.5	.18
	Walking	21 (35.25)	1146	17.5 (32)	1100	-3.5	18 (17.5)	1677	20.5 (21.75)	1927	2.5	-6	.01*
	Vigorous	2.5 (7.75)	253	3 (11.5)	245	0.5	2.5 (3.75)	265	3 (5)	307	0.5	0	.21
	Connect	7 (16.25)	334	6.5 (11.5)	359	-0.5	7 (8.75)	666	8.5 (8.75)	715	1.5	-2	.50

[a] Mann-Whitney *U* test was carried out to determine if there were significant differences in the change in counts of behaviours at baseline and 12 months between comparison and intervention groups;
* Statistically significant at $p < 0.05$ (z-test, two-tailed)

Table 4.4. Multilevel mixed-effects negative binomial regression results for the total number of older adults and adults. *All models adjusted for day, time of day and precipitation.*

Time point	Age group	Total count (intervention, comparison)	IRR [a]	95% CI	Robust standard error	*p*-value	Random effect variance (site)
Baseline	Older adults	427 (120, 307)	-	-	-	-	-
	Adults	3238 (1338, 1900)	-	-	-	-	-
6 months post-baseline	Older adults	402 (119, 283)	1.23	0.76 – 1.98	0.30	0.39	0.21
	Adults	3467 (1468, 1999)	0.98	0.82 – 1.18	0.09	0.86	0.09
12 months post-baseline	Older adults	365 (122, 243)	1.32	0.94 – 1.85	0.23	0.11	0.16
	Adults	3563 (1374, 2189)	0.85	0.72 – 1.00	0.72	0.051	0.10

[a] Incidence rate ratio;

* Statistically significant at $p < 0.05$ (z-test, two-tailed)

4.4.2 Sensitivity analysis

Thirty-eight observation periods were removed due to precipitation: 14% (13/96) of observation periods from the intervention group and 15% (25/168) from the comparison group.

At 12 months post-baseline, there remained no significant difference in the change of Older Adults' Take Notice behaviour between the intervention group (increased 0.09 per hour) and comparison group (increased 0.05 per hour). Appendix J contains the results of the sensitivity analyses.

4.4.3 Nested qualitative study – content analysis

There were fifteen participants (five males and 10 females). Twelve participants (80%) lived in Old Moat, two (13%) used to live in Old Moat but still spend time there, and one (7%) lived in an area adjacent to Old Moat. Only four participants (27%) reported noticing any of the changes at the intervention sites. Of these four participants, four noticed the new trees and two noticed the tree socks.

4.5 Discussion

We found no evidence that low-cost physical changes to urban amenity green spaces had an impact on wellbeing behaviours in older adults or adults. There was also no evidence that these changes had an impact on the number of either older adults or other adults using these spaces. The nested qualitative study suggested that the intervention was not substantial enough to be noticed, especially compared to other recent neighbourhood changes in the area.

4.5.1 How this study compares to the current literature

These null findings are in line with the only existing UK natural experimental study specifically among older adults in this area, which found that making residential streets more attractive (e.g. installing tree planters) and safer (e.g. reducing traffic) had no impact on older adults' physical activity at 3 to 6 months post-intervention [29]. Their study, together with findings from the present study, suggest that making low-cost changes to residential streets may not be enough to influence behaviour in older adults, at least within a relatively short time frame after the intervention (< 10 months).

This evidence is important in relation to current national policy and practice recommendations for planning and designing urban green spaces in the UK, which imply that interventions of any scale will produce changes in wellbeing outcomes. For example, 'Spatial Planning for Health' [35] is an evidence resource commissioned by Public Health England for urban planners and public health professionals, which recommends that neighbourhood tree planting can be effective in increasing physical activity. The intervention provider's belief in the present study that planting trees would encourage outdoor activity is in line with this resource, even though the intervention sites were not primarily used as dedicated recreational spaces.

These null findings highlight the need for more nuanced recommendations that specify key contextual factors which may influence the effectiveness of urban green space interventions; such as the dose of intervention, type of green space (e.g. amenity, park, community garden), the scale of intervention (e.g. street, neighbourhood, city), and local population characteristics (e.g. socioeconomics, age, population density). For example, in the present study, the intervention neighbourhood was already relatively green; this may have reduced the salience, and thus effectiveness, of the intervention. It might be the case that a similar intervention implemented in a different context (e.g. more derelict urban areas) may lead to bigger effects on behaviour. Whilst the impact of context is currently poorly understood in this research area, a recent review has started to theorise the role of context in the success of physical environment interventions [61].

The intervention in the present study was implemented on a smaller scale and was less costly compared to most previous intervention research in this area; there is a lack of research on how implementing smaller green space features (such as street trees and roadside vegetation) impact health and wellbeing outcomes [62,63]. Small-scale interventions can offer a low-cost and quick approach to improving the local environment, particularly in densely populated urban areas like Manchester, where the potential for exposure is high but opportunities for large-scale interventions are limited. While there were no measurable intervention effects on behaviour in the present study, there are examples of small-scale environmental interventions which have been effective [33,64]. For example, low-cost improvements to a public space in Manchester city centre (UK) led to increases in the number of users, and increases in the number of people taking notice of the environment and social interactions, following the introduction of benches that were particularly important among older people [33]. It is therefore likely that small changes that address key barriers for older adults use of green spaces (e.g. lack of benches, poor quality pavements) may be more effective in increasing older adults' enjoyment of the outdoors and confidence to walk within their local environments [65].

4.5.2 Strengths and limitations

We addressed seven key methodological weaknesses causing high risk of bias in previous natural experimental studies in this area (see Table 4.5). The most important improvements addressed one of the major sources of bias in natural experimental studies: confounding. To reduce confounding, we developed a new five-step process to identify multiple matched comparison sites using ten objective and subjective environmental correlates of physical activity (at neighbourhood and street-level) (Figure 4.1). As shown in Table 4.1, this systematic process identified comparison sites that were well-matched on several important variables. We also used a difference in differences analysis to control for known confounders (day, time of day, precipitation), as recommended by MRC guidance [16].

The methods and analyses used to reduce confounding in the present study are vast improvements on many previous natural experimental studies in this area, which have often used poorly matched comparison groups and poor control of confounding variables [25].

Table 4.5. How the present study implemented recommendations to address methodological weaknesses in previous natural experimental studies.

Recommendation	**Improvements in the present study**
1. Better matching of comparison sites	Developed a five-step process to identify matched comparison sites, using ten objective and subjective variables consistently associated with physical activity.
2. Use multiple comparison sites	Two matched comparison sites were identified for each intervention site.
3. Controlling for confounders in statistical analyses	A difference in differences analysis controlled for important known confounders, as recommended by MRC guidance for natural experiments [16].
4. Publish study protocols with a priori analyses specified	A study protocol was published [43], which was submitted for publication before follow-up data were collected. Any important deviations from the protocol were reported.
5. Use adequate outcome measurements	A reliable and valid observation tool was used: MOHAWk [34]. Observations were counterbalanced over more than one week at each time-point, and we conducted multiple follow-ups. A nested qualitative study provided a mixed-methods approach.
6. Better reporting of samples and interventions	The sample was reported in line with the Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) statement checklist [66]. The intervention was reported in line with the Template for Intervention Description and Replication for Population Health and Policy interventions (TIDieR-PHP) checklist [67].
7. Conduct sample size calculations	A power calculation was conducted for the primary outcome measure, using an approach suggested by Donner and Klar [57] for a matched pair design.
8. Measure intervention exposure at the individual level	Measuring intervention exposure is less important when using observations (compared with self-report) because observations are conducted directly in the intervention site.

We also made other methodological improvements. We published a study protocol with a priori analyses specified and reported any deviations, to improve transparency and reduce the risk of selective reporting. We conducted a power calculation to ensure there was an appropriate sample size to detect an effect. These are substantial improvements on previous natural experimental studies: none of the included studies in a recent review published a study protocol or attempted sample size calculations [25]. We also provided clear and transparent reporting using existing checklists for reporting samples [66] and interventions [67]. Triangulation between the observations and interview data provided a mixed-methods approach, which enabled us to explore why the intervention was not effective in influencing the measured outcomes.

A potential limitation is insufficient statistical power. This study was the first application of a newly validated observation tool (MOHAWk) in a natural experimental study. Therefore, due to limited data, it was difficult to estimate the size of the effect that the intervention would have on the outcomes for the power calculation. There were also a high proportion of zero counts for many of the wellbeing behaviours, which prevented us from fitting a suitable regression model. Data from this study will contribute to the scarce evidence base to inform sample size calculations in future natural experimental studies.

The intervention may have influenced wellbeing via pathways not measured in the present study, such as an increased sense of pride for local residents. Qualitative data from the nested qualitative study (reported in detail elsewhere [42]) suggested that the intervention did not affect subjective wellbeing in older adults, most of whom lived in Old Moat. The intervention may also have had a bigger influence on some of the measured wellbeing behaviours at weekends and at different times of the year, particularly during warmer and drier months in spring and summer. The scheduling of data collection was not sensitive to wider phenological timings (e.g. the timing of flowering of spring bulbs), which may have influenced results found. Further, the intervention may provide additional benefits beyond the 12 month time frame of the present study; for example, when mature, street trees can provide shelter from wind and direct sunlight, and improve stormwater management [68]. It is important to acknowledge other potential co-benefits that may be important for decision-makers.

4.5.3 Implications for policy and practice

This study provides robust practice-based evidence to inform stakeholders who are responsible for the planning, design and management of urban green spaces. More robust evidence of this sort is needed to better inform future policy and practice, particularly as many UK policy and industry guidelines advocate improving urban green spaces as an effective approach to support health and wellbeing, without considering factors like scale and local demographics [35–39]. Where locational and demographically specific guidelines do exist, they often rely on expert consensus due to the absence of robust evidence [69]. Therefore, producing better evidence is vital to ensure limited funding available for investing in urban green spaces in the UK is utilised most effectively [70].

It is crucial that null findings from robust studies like this are published. Publication bias towards only publishing novel or significant results is a common issue across the health and social sciences [71]. However, knowing which interventions are unlikely to work for desired outcomes is equally as important for policy and practice as knowing which interventions work, to produce more nuanced recommendations that prevent false expectations. For example, recent UK National Institute for Health and Care Excellence (NICE) guidelines on the physical environment and physical activity [38] lacks guidance on which interventions are likely to be ineffective, and potential ineffectiveness of interventions in different physical and social contexts. This study therefore contributes to the sparse evidence base about what does not work for whom and why, which will ultimately help stakeholders redirect efforts into actions that are more likely to be effective in improving health and wellbeing in the population.

As indicated in the nested qualitative study, older adults in the community preferred to visit larger green spaces and perceived the intervention sites as belonging to other people who lived nearby [42]. This highlights the importance of involving the local community in the planning and design process of urban green space interventions, which has been found to provide an effective way to ensure the intervention is accepted and used by the local community, especially in vulnerable groups such as older adults [72].

4.5.4 Implications for research

Evaluating natural experiments is essential to produce causal evidence for preventive public health policy, yet conducting this type of research is often fraught with practical, political and cultural obstacles [73]. Using rigorous yet feasible mixed-methods, this study provides an exemplar to strengthen causal inferences in future natural experimental studies of environmental interventions. This study also contributes evidence required for the development of theoretical models for understanding contextual influences. Similar methods are being applied in two separate natural experimental studies of different types of urban green space interventions in Greater Manchester, including changes along an urban canal [52] and a new park [53]. More natural experimental studies like this are needed to provide a better understanding of which small green space changes can best enable improvements in health and wellbeing outcomes, especially interventions that are scalable to bring about transformative change.

4.6 Conclusions

This controlled natural experimental study provides rare robust evidence on the impact of an urban green space intervention on a range of wellbeing behaviours, in an understudied population (older adults) and setting (UK). There was no evidence that low-cost physical changes to urban amenity green spaces lead to an increase in wellbeing behaviours or usage of green spaces in older adults or adults. This study used methods with substantially lower risk of bias than previous research and therefore demonstrates how to address key methodological weaknesses in previous natural experimental studies in this field. Results suggest the need for more nuanced policy and practice recommendations for planning and designing urban green spaces. More robust natural experimental studies like this are now needed to better understand the effectiveness of a range of urban green space interventions, especially small-scale changes that are readily replicable and scalable for transformative change in wider urban areas.

4.7 References

1. World Health Organization. The Global strategy and action plan on ageing and health. World Health Organization; 2017. Available from: http://who.int/ageing/global-strategy/en/

2. United Nations. World Population Ageing 2013. New York; 2013. Available from: https://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2013.pdf

3. Chatterji S, Byles J, Cutler D, Seeman T, Verdes E. Health, functioning, and disability in older adults—present status and future implications. Lancet. 2015;385(9967):563–75.

4. Lee ACK, Maheswaran R. The health benefits of urban green spaces: a review of the evidence. J Public Health. 2010;33(2):212–22.

5. United Nations. World Urbanization Prospects: The 2014 Revision, Highlights. New York; 2015. Available from: https://esa.un.org/unpd/wup/publications/files/wup2014-report.pdf

6. Richardson M, Passmore H, Barbett L, Lumber R, Thomas R, Hunt A. The green care code: How nature connectedness and simple activities help explain pro-nature conservation behaviours. People Nat. 2020.

7. Pritchard A, Richardson M, Sheffield D, McEwan K. The Relationship Between Nature Connectedness and Eudaimonic Well-Being: A Meta-analysis. J Happiness Stud. 2020;21(3):1145–67.

8. Hartig T, Mitchell R, De Vries S, Frumkin H. Nature and Health. Annu Rev Public Heal. 2014;35:207–28.

9. Cervinka R, Röderer K, Hefler E. Are nature lovers happy? on various indicators of well-being and connectedness with nature. J Health Psychol. 2012;17(3):379–88.

10. Wyles KJ, White MP, Hattam C, Pahl S, King H, Austen M. Are Some Natural Environments More Psychologically Beneficial Than Others? The Importance of Type and Quality on Connectedness to Nature and Psychological Restoration. Environ Behav. 2019;51(2):111–43.

11. World Health Organization. Urban Green Space Interventions and Health: A review of impacts and effectiveness. Copenhagen; 2017. Available from: http://www.euro.who.int/__data/assets/pdf_file/0010/337690/FULL-REPORT-for-LLP.pdf?ua=1

12. Nicholson NR. A Review of Social Isolation: An Important but Underassessed Condition in Older Adults. 2012;33(2–3):137–52.

13. Sport England. Active Lives Adult Survey - May 16/17 Report. London; 2017. Available from: https://www.sportengland.org/media/12458/active-lives-adult-may-16-17-report.pdf

14. World Health Organization. Urban Green Space Interventions and Health: A review of evidence. Copenhagen; 2017. Available from: http://www.euro.who.int/__data/assets/pdf_file/0005/321971/Urban-green-spaces-and-health-review-evidence.pdf?ua=1

15. Barnett DW, Barnett A, Nathan A, Van Cauwenberg J, Cerin E. Built environmental correlates of older adults’ total physical activity and walking: a systematic review and meta-analysis. Int J Behav Nutr Phys Act. 2017;14(1):103.

16. Craig P, Cooper C, Gunnell D, Haw S, Lawson K, Macintyre S, et al. Using

natural experiments to evaluate population health interventions: new Medical Research Council guidance. J Epidemiol Community Health. 2012;66(12):1182–6.

17. Roberts H, McEachan R, Margary T, Conner M, Kellar I. Identifying Effective Behavior Change Techniques in Built Environment Interventions to Increase Use of Green Space: A Systematic Review. Environ Behav. 2018;50(1):28–55.

18. Mayne SL, Auchincloss AH, Michael YL. Impact of policy and built environment changes on obesity-related outcomes: A systematic review of naturally occurring experiments. Obes Rev. 2015;16(5):362–75.

19. Hunter RF, Christian H, Veitch J, Astell-Burt T, Hipp JA, Schipperijn J. The impact of interventions to promote physical activity in urban green space: A systematic review and recommendations for future research. Soc Sci Med. 2015;124:246–56.

20. Kärmeniemi M, Lankila T, Ikäheimo T, Koivumaa-Honkanen H, Korpelainen R. The Built Environment as a Determinant of Physical Activity: A Systematic Review of Longitudinal Studies and Natural Experiments. Ann Behav Med. 2018;52(3):239–51.

21. Hunter RF, Cleary A, Brauback M. Environmental, Health and Equity Effects of Urban Green Space Interventions. In: Marselle MR, Stadler J, Korn H, Irvine KN, Bonn A, editors. Biodiversity and Health in the Face of Climate Change. 2019. p. 381–409.

22. Kondo M, Fluehr J, McKeon T, Branas C. Urban Green Space and Its Impact on Human Health. Int J Environ Res Public Health. 2018;15(3):445.

23. Higgins JP, Altman DG. Chapter 8. Assessing Risk of Bias in Included Studies. In: Cochrane Handbook for Systematic Reviews of Interventions. Chichester, UK:

John Wiley & Sons, Ltd; 2008. p. 187–241. Available from: http://doi.wiley.com/10.1002/9780470712184.ch8

24. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343(7829).

25. Benton JS, Anderson J, Hunter RF, French DP. The effect of changing the built environment on physical activity: a quantitative review of the risk of bias in natural experiments. Int J Behav Nutr Phys Act. 2016;13(1):107.

26. MacMillan F, George ES, Feng X, Merom D, Bennie A, Cook A, et al. Do Natural Experiments of Changes in Neighborhood Built Environment Impact Physical Activity and Diet? A Systematic Review. Int J Environ Res Public Health. 2018;15(2):217.

27. Van Cauwenberg J, Nathan A, Barnett A, Barnett DW, Cerin E. Relationships Between Neighbourhood Physical Environmental Attributes and Older Adults' Leisure-Time Physical Activity: A Systematic Review and Meta-Analysis. Sport Med. 2018;48:1635–60.

28. Cerin E, Nathan A, van Cauwenberg J, Barnett DW, Barnett A. The neighbourhood physical environment and active travel in older adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act. 2017;14(1):15.

29. Ward Thompson C, Curl A, Aspinall P, Alves S, Zuin A. Do changes to the local street environment alter behaviour and quality of life of older adults? The "DIY Streets" intervention. Br J Sports Med. 2014;48(13):1059–65.

30. Sallis JF, Bull F, Burdett R, Frank LD, Griffiths P, Giles-Corti B, et al. Use of science to guide city planning policy and practice: how to achieve healthy and

sustainable future cities. Lancet. 2016;388(10062):2936–47.

31. Houlden V, Weich S, Porto de Albuquerque J, Jarvis S, Rees K. The relationship between greenspace and the mental wellbeing of adults: A systematic review. PLoS One. 2018;13(9):e0203000.

32. New Economics Foundation. Five ways to wellbeing: The evidence.. London; 2008. Available from: http://b.3cdn.net/nefoundation/8984c5089d5c2285ee_t4m6bhqq5.pdf

33. Anderson J, Ruggeri K, Steemers K, Huppert F. Lively Social Space, Well-Being Activity, and Urban Design: Findings From a Low-Cost Community-Led Public Space Intervention. Environ Behav. 2017;49(6):685–716.

34. Benton JS, Anderson J, Pulis M, Cotterill S, Hunter RF, French DP. Method for Observing pHysical Activity and Wellbeing (MOHAWk): validation of an observation tool to assess physical activity and other wellbeing behaviours in urban spaces. Cities Health. 2020.

35. Public Health England. Spatial Planning for Health An evidence resource for planning and designing healthier places. 2017. Available from: https://www.gov.uk/government/publications/spatial-planning-for-health-evidence-review

36. Cavill N. Building health: creating and enhancing places for healthy active lives: what needs to be done? London; 2007.

37. Sport England. Active Design: Planning for health and wellbeing through sport and physical activity. London; 2015.

38. National Institute for Health and Clinical Excellence (NICE). Physical activity and the environment. 2018. Available from: www.nice.org.uk/guidance/ng90

39. Department for Environment Food and Rural Affairs. A Green Future: Our 25 Year Plan to Improve the Environment. London; 2018. Available from: https://www.gov.uk/government/publications/25-year-environment-plan

40. Wahl H-W, Iwarsson S, Oswald F. Aging Well and the Environment: Toward an Integrative Model and Research Agenda for the Future. Gerontologist. 2012;52(3):306–16.

41. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ. 2015;350.

42. Macintyre VG, Cotterill S, Anderson J, Phillipson C, Benton JS, French DP. "I Would Never Come Here Because I've Got My Own Garden": Older Adults' Perceptions of Small Urban Green Spaces. Int J Environ Res Public Health. 2019;16(11):1994.

43. Benton JS, Anderson J, Cotterill S, Dennis M, Lindley SJ, French DP. Evaluating the impact of improvements in urban green space on older adults' physical activity and wellbeing: protocol for a natural experimental study. BMC Public Health. 2018;18(1):923.

44. Moran M, Van Cauwenberg J, Hercky-Linnewiel R, Cerin E, Deforche B, Plaut P. Understanding the relationships between the physical environment and physical activity in older adults: a systematic review of qualitative studies. Int J Behav Nutr Phys Act. 2014;11(1):79.

45. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155–63.

46. Phillipson C, White S, Aftab F, Hammond M. Old moat: Age-friendly

Neighbourhood report. 2013. Available from: https://www.southwayhousing.co.uk/media/1016/old-moat-age-friendly-neighbourhood-report.pdf

47. Manchester City Council. Old Moat Ward Electoral Registration Summary. 2014. Available from: www.manchester.gov.uk/download/downloads/id/22217/old_moat

48. Manchester City Council. F1 Indices of Deprivation 2015. 2015;

49. Department for Communities and Local Government. English Indices of Deprivation 2015 [computer file]. Downloaded from: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015. Licensed under: https://www.nationalarchives.gov.uk/doc/open-government-licence/ version/3. 2015.

50. Cain KL, Gavand KA, Conway TL, Geremia CM, Millstein RA, Frank LD, et al. Developing and validating an abbreviated version of the Microscale Audit for Pedestrian Streetscapes (MAPS-Abbreviated). J Transp Heal. 2017;5:84–96.

51. Gidlow CJ, Ellis NJ, Bostock S. Development of the Neighbourhood Green Space Tool (NGST). Landsc Urban Plan. 2012;106(4):347–58.

52. Benton J, Anderson J, Cotterill S, Dennis M, French DP. The impact of new walking infrastructure and changes to green space along an urban canal on physical activity and wellbeing: protocol for a natural experimental study. Open Sci Framew. 2018; https://osf.io/zcm7v/

53. Anderson J, French DP, Benton JS, Dennis M, Rothwell J. The effects of the Grow Green urban greening improvements on physical health and wellbeing behaviours in Manchester UK: A natural experimental study. Open Sci Framew.

2020; https://osf.io/zqgcn

54. Seligman MEP, Steen TA, Peterson C. Positive Psychology Progress: Empirical Validation of Interventions. Am Psychol. 2005;60(5):410–21.

55. McEwan K, Richardson M, Sheffield D, Ferguson FJ, Brindley P. A Smartphone App for Improving Mental Health through Connecting with Urban Nature. Int J Environ Res Public Health. 2019;16(18):3373.

56. Panter J, Guell C, Prins R, Ogilvie D. Physical activity and the environment: conceptual review and framework for intervention research. Int J Behav Nutr Phys Act. 2017;14(1):156.

57. Donner A, Klar N. Design and Analysis of Cluster Randomization Trials in Health Research. John Wiley & Sons; 2000. 178 p.

58. Cameron AC, Trivedi PK. Regression analysis of count data. Cambridge University Press; 1998.

59. Craig P, Vittal Katikireddi S, Leyland A, Popham F. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. Annu Rev Public Heal. 2017;38:39–56.

60. Downe-Wamboldt B. Content analysis: Method, applications, and issues. Health Care Women Int. 1992;13(3):313–21.

61. Panter J, Guell C, Humphreys D, Ogilvie D. Can changing the physical environment promote walking and cycling? A systematic review of what works and how. Health Place. 2019;58:102161.

62. Lovell R, Depledge M, Maxwell S. Health and the natural environment: A review of evidence, policy, practice and opportunities for the future. 2018. Available from: http://randd.defra.gov.uk

63. Moore THM, Kesten JM, López-López JA, Ijaz S, McAleenan A, Richards A, et al. The effects of changes to the built environment on the mental health and well-being of adults: Systematic review. Health Place. 2018;53:237–57.

64. Prins RG, Kamphuis CBM, Van Lenthe FJ. The effects of small-scale physical and social environmental interventions on walking behaviour among Dutch older adults living in deprived neighbourhoods: results from the quasi-experimental NEW.ROADS study. Int J Behav Nutr Phys Act. 2019;16:133.

65. McGowan LJ, Powell R, French DP. How acceptable is reducing sedentary behavior to older adults? Perceptions and experiences across diverse socioeconomic areas. J Aging Phys Act. 2019;27(5):642–52.

66. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health. 2004;94(3):361–6.

67. Campbell M, Katikireddi SV, Hoffmann T, Armstrong R, Waters E, Craig P. TIDieR-PHP: a reporting guideline for population health and policy interventions. BMJ. 2018;361.

68. Silvera Seamans G. Mainstreaming the environmental benefits of street trees. Urban For Urban Green. 2013;12(1):2–11.

69. Douglas O, Lennon M, Scott M. Green space benefits for health and well-being: A life-course approach for urban planning, design and management. Cities. 2017;66:53–62

70. Dickinson J, Bennett E, Marson J. Challenges facing green space: is statute the answer? J Place Manag Dev. 2019;12(1):121–38.

71. Franco A, Malhotra N, Simonovits G. Publication bias in the social sciences:

Unlocking the file drawer. Science. 2014;345(6203):1502–5.

72. World Health Organization. Urban Green Space Interventions and Health: A review of impacts and effectiveness. Copenhagen; 2017.

73. Ogilvie D, Adams J, Bauman A, Gregg EW, Panter J, Siegel KR, et al. Using natural experimental studies to guide public health action: turning the evidence-based medicine paradigm on its head. J Epidemiol Community Heal. 2019;74(2):203–8.

SECTION 4. Natural experimental study 2

Section overview

This section relates to the second natural experimental study of new walking infrastructure and green space improvements along an urban canal on canal usage and wellbeing behaviours in adults and older adults in Greater Manchester, UK. Chapter 5 is the protocol paper describing the planned methods and analyses, which is reported using the headings from the template form for pre-registering studies on the Open Science Framework. Chapter 6 reports on the findings of this study.

Chapter 5. The impact of new walking infrastructure and changes to green space along an urban canal on physical activity and wellbeing: protocol for a natural experimental study

This article has been published in the Open Science Framework:

- Benton J, Anderson J, Cotterill S, Dennis M, French DP. The impact of new walking infrastructure and changes to green space along an urban canal on physical activity and wellbeing: protocol for a natural experimental study. Open Sci Framew. 2018; https://osf.io/zcm7v/

5.1 Aims and objectives

5.1.1 Research questions

The overall aim of this natural experimental study is to investigate the effect of new walking infrastructure and changes to green space along an urban canal (hereafter referred to as an 'urban canal intervention') on adults' and older adults' physical activity and wellbeing.

The specific objectives are as follows:

1. To examine whether an urban canal intervention increases the number of adults and older adults using the intervention site in comparison to matched comparison sites where no such changes occur (primary outcome);
2. To examine whether an urban canal intervention increases adults and older adults physical activity levels (walking and vigorous activity) in comparison to matched comparison sites where no such changes occur (secondary outcomes);
3. To examine whether an urban canal intervention increases adults and older adults Sedentary, Connect and Take Notice behaviour in comparison to matched comparison sites where no such changes occur (exploratory outcomes);
4. To assess canal users awareness of and views regarding the intervention; self-reported changes in use of the canal and physical activity following the intervention; and whether self-reported reasons for using the canal are in line with hypotheses proposed in a logic model (exploratory outcomes).

5.1.2 Hypotheses

1. The urban canal intervention will significantly increase the overall count of adults and older adults per observation period using the intervention site at each follow-up in comparison to matched comparison sites where no such changes occur (primary outcome);

2. The urban canal intervention will significantly increase the overall count of adults and older adults per observation period engaging in Walking and Vigorous behaviour at each follow-up in comparison to matched comparison sites where no such changes occur (secondary outcomes);
3. The urban canal intervention will significantly increase the overall count of adults and older adults engaging in Sedentary, Connect and Take Notice behaviour at each follow-up in comparison to matched comparison sites where no such changes occur (exploratory outcomes).

5.2 Sampling plan

5.2.1 Existing data

Registration prior to analysis of the data: as of the date of submission of this research plan for preregistration, baseline data has been collected.

5.2.2 Data collection procedures

5.2.2.1 Comparison site matching

Due to the absence of randomisation in natural experiments, comparison (or control) groups ought to be matched on all important variables that influence the outcome to strengthen internal validity and improve accuracy of the estimated intervention effect [1]. However, previous studies in this area have often used poorly matched comparison groups; in particular, there is an absence of any matching based on objective features of the environment that correlate with physical activity e.g. population density, street connectivity [2]. To address this issue, the present study selected comparison sites that were matched to the intervention site, using several key objective and subjective environmental variables. Two comparison sites were matched to the intervention site to increase the likelihood of finding comparable comparison sites.

There is an absence of evidence on how characteristics of the built environment may influence wellbeing; accordingly, all variables that comparison sites were matched on were from a systematic review of reviews of physical activity environmental correlates (using four systematic reviews, one meta-analysis and one systematic review of reviews that focused on adults) [3]; and three recent systematic reviews that have investigated built environmental correlates of older adults' physical activity: one qualitative [4] and two quantitative systematic reviews [5,6]. Figure 5.1 displays the variables that were reported as consistent correlates of physical activity in at least two of the systematic reviews, meta-analysis or systematic review of reviews investigating adults' physical activity; and in at least two of the systematic reviews investigating older adults' physical activity. These correlates represent the key variables that were used for comparison site matching.

There are no agreed-upon standards for how researchers ought to identify matched comparison sites when studying the effects of the built environment on physical activity [2]. Therefore, a five-step process of matching was developed. An overview of this process is displayed in Figure 5.1, and each step is described in more detail in Appendix K.

This matching process has been used to identify multiple matched comparison sites in a previous natural experimental study [7]. However, only one comparison site was found using this method of matching in the present study. This is because the intervention site in the present study is located along an urban canal; a type of land-use not frequently found across the majority of neighbourhoods, thus providing a limited number of unexposed potential matches. To identify a second matched comparison site, the same matching process was used but potential neighbourhoods for step one (see Figure 5.1) were purposefully identified from the same canal route as the intervention site, within the boundaries of Greater Manchester.

The second comparison site is approximately 8.8 km walking distance from the intervention site. This second comparison site is not directly linked to the intervention site via the northern canal footpath, thus reducing the risk of contamination.

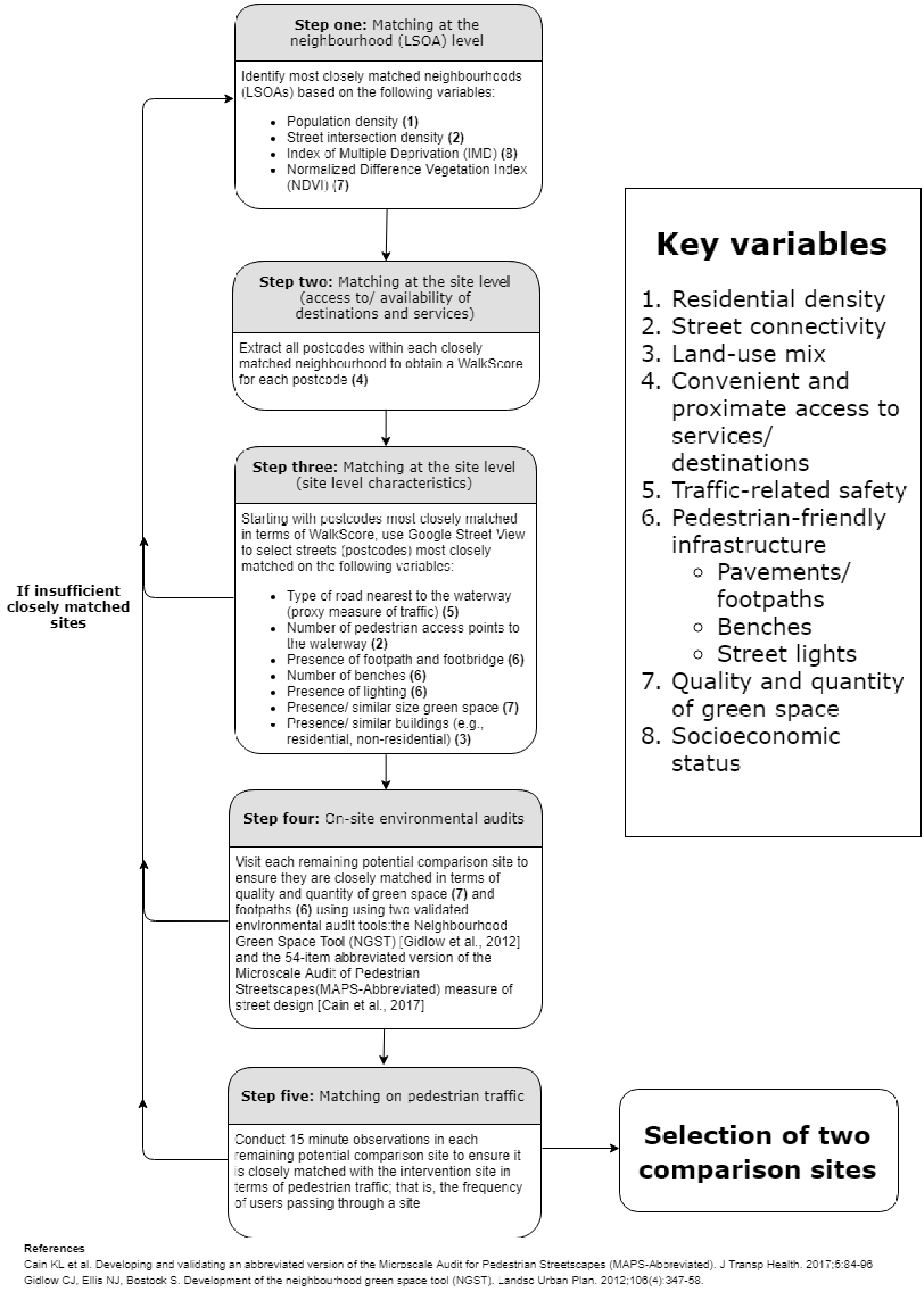


Figure 5.1. Overview of the five steps constituting the comparison site matching and selection process. *Numbers in brackets refer to the key variables used for comparison site matching.*

Key

X = Observer location

= Target area boundary

= Target area boundary for exploratory analysis

Figure 5.2. Target area boundaries and primary location for the observer at the intervention site. *Maps drawn using: www.digimap.edina.ac.uk. © Crown Copyright and Database Right (2018). Ordnance Survey (Digimap Licence).*

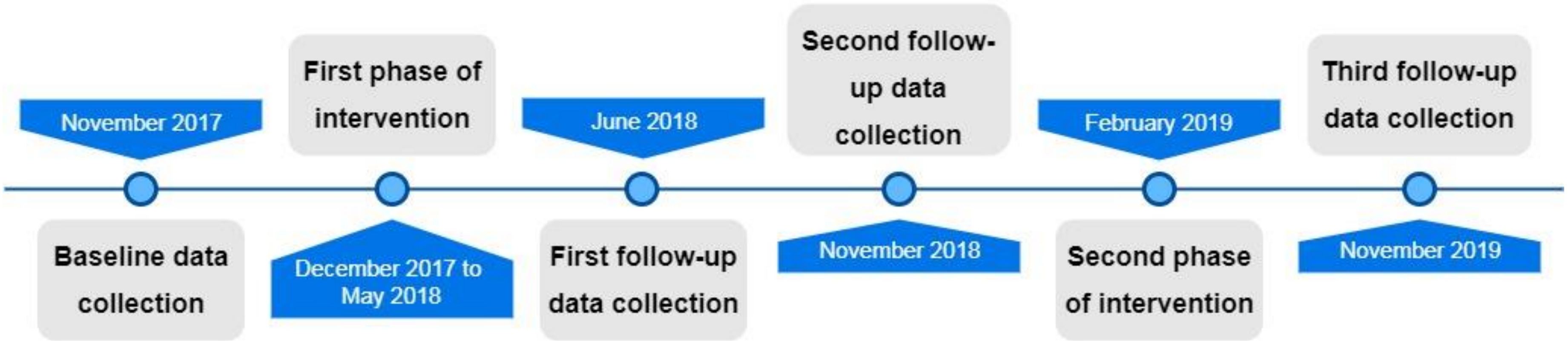


Figure 5.3. Intervention and data collection timeline.

	November 2017 (Baseline)													
	Monday		**Tuesday**		**Wednesday**		**Thursday**		**Friday**		**Saturday**		**Sunday**	
	6th		**7th**		**8th**		**9th**		**11th**		**11th**		**12th**	
8am											1	-		
10am											1	-		
1pm											1A	-		
3pm											1A	-		
	13th		**14th**		**15th**		**16th**		**17th**		**18th**		**19th**	
8am			1A	1	1B	-								
10am			1A	1	1B	-								
1pm			1B	1	1	-								
3pm			1B	1	1	-								
	20th		**21st**		**22nd**		**23rd**		**24th**		**25th**		**26th**	
8am											1A	1B		
10am											1A	1B		
1pm											1	1B		
3pm											1	1B		
	27th		**28th**		**29th**		**30th**		**1st December**		**2nd December**		**3rd December**	
8am			1B	-	1	1A								
10am			1B	-	1	1A								
1pm			1A	-	1B	1A								
3pm			1A	-	1B	1A								

Figure 5.4. Observation schedule for two observers during baseline data collection for the intervention and two comparison sites. *The two columns within each day correspond to the two observers i.e. one column for each observer. The numbers/ letters refer to intervention and comparison sites i.e. '1' is the Intervention site, and '1A' and '1B' are the corresponding comparison sites: Comparison site 1A and Comparison site 1B, respectively. See Table 5.1 for details on each intervention and comparison site.*

5.2.3 Outcomes

Data will be collected using two outcome measures: (1) direct observations of behaviour (systematic observation); and (2) intercept surveys.

5.2.3.1 Systematic observation

Data will be collected on all individuals (infants, children, teens, adults, older adults) entering the target area in the intervention site and two comparison sites during observations. However, this study is focused on adults and older adults; therefore analyses will only consider data from adults and older adults. Individuals being observed are not ‘participants’ in the usual sense: they should not be aware that they are being observed for the purposes of a study. For this reason, details of recruitment efforts are not applicable for this outcome measure.

The same procedure will be used as set out in the MOHAWk (Method for Observing pHysical Activity and Wellbeing) observation manual (Appendix A). All observers will be trained in using MOHAWk and will be required to demonstrate high agreement with the trainer before making observations in the present study. Prior to observations, observers will visit each site to agree on the boundaries of the target area in which all participants will be recorded. Target areas are of similar size between corresponding intervention and comparison sites, and the same target areas will be used at all time points (see Appendix L).

The target area at the intervention site includes anyone who uses the northern footpath on the Bridgewater Canal; this footpath is the main component of the intervention. However, observers will also record anyone who passes through Moss House Lane (this road is one of the intervention components) but does not enter the target area (see Figure 5.2). Only participants who entered the target area during observation periods, and therefore used the northern footpath, will be included in the primary and secondary analyses. An exploratory analysis will compare counts of participants who used the northern footpath (i.e. entered the target area) with participants who passed through Moss House Lane but did not enter the target area (see Analysis plan). This exploratory analysis will therefore examine the impact of the changes along the northern footpath on the number of adults and older adults who used Moss House Lane to access or exit the unchanged southern side of the canal.

Baseline data collection occurred in November 2017 before the first phase of the intervention was implemented between December 2017 and May 2018. The second phase of the intervention is planned for implementation in February 2019. Follow-up data collection will be repeated soon after the main phase of improvements has been completed: in June 2018 (7 months post-baseline) and November 2018 (12 months post-baseline). A further round of follow up data will be collected after the final phase of the intervention has been completed, in November 2019 (24 months post-baseline). Figure 5.3 displays a timeline for each phase of the intervention and data collection.

Observations will be conducted over three days (Tuesday, Wednesday, Saturday), four times a day at each time point, resulting in a total of twelve observation periods for each site at each time point. Observations will be conducted at four set hour-long observation periods per day: early morning (8-9am), late morning (10-11am), early afternoon (1-2pm), and late afternoon (3-4pm).

Observations for the intervention site and two comparison sites will be spread over four weeks. This will provide a more robust assessment of activity over a longer period rather than observing activity during a single week. Observations will be counterbalanced to control for week, day of week and time of day. The observation schedule used for baseline data collection is displayed in Figure 5.4. This schedule will be replicated at all follow-ups.

5.2.3.2 Intercept surveys

Survey data will be collected from the intervention site only. Interview-administered surveys will be conducted with all eligible consenting canal users aged 18 years or older entering the target area in the intervention site during data collection periods. For groups of two people or more, the group member with the birthday closest to the day of the interview will be asked to respond to the survey. People engaged in activities that would clearly be interrupted by a request to participate (e.g. people on the phone, people directly interacting with young children) will not be approached. Participants must be able to understand and speak English competently and not display any obvious learning difficulties as they may struggle to understand what is asked of them and provide informed consent. Informed consent will be obtained verbally face-to-face by the researcher. Surveys will be completely anonymous and will not require any identifiable information. A detailed participant information sheet will be given to all participants (see Appendix M). There will be no payment for participation.

No baseline survey data collection occurred. Follow-up survey data collection will be collected in June 2018 (7 months) and November 2018 (12 months). Surveys will be conducted on the same days and times as the observations but in the following weeks after observations are completed at each time point. Surveys will be conducted at four set hour-long periods per day: early morning (8-9am), late morning (10-11am), early afternoon (1-2pm), and late afternoon (3-4pm).

To capture users along the northern footpath on both sides of Moss House Lane, the interviewer will be positioned for half the observation period on the west access point, and the other half at the east access point. The positioning of the interviewer will be counterbalanced to control for week, day of week and time of day.

5.2.4 Sample size

5.2.4.1 Systematic observation

Observations will be conducted over three days (Tuesday, Wednesday, Saturday), four times a day at each time point, resulting in a total of twelve hours of observations for each site at each time point.

5.2.4.2 Intercept surveys

Intercept surveys will be conducted over three days (Tuesday, Wednesday, Saturday), four times a day at each time point, resulting in a total of twelve hours of data collection at the intervention site (June and November 2018). The desired sample size for the intercept survey is 48 participants across both time points.

5.2.5 Sample size rationale

5.2.5.1 Systematic observation

Conducting observations over three days, four times a day can produce consistency approaching that obtained by observing 10 hours per day [7]. Further, the timing of observation periods provide sufficient time for breaks and possible travel to other sites in between observation periods, whilst ensuring that no observations are conducted outside of daylight hours to reduce the risk of possible danger to the observers.

To assess the power of the study, we conducted a power calculation of the primary outcome measure: overall counts per observation period (hour) of adults. We followed the approach suggested by Donner and Klar [8, p.66] for calculating the sample size for a matched pair design: first, calculate the number of sites that would be required for a completely randomised cluster design, and then multiply by one minus the correlation between the number of people in the intervention and control sites. This suits our context because it allows us to account for multiple comparison groups for each intervention group.

One intervention site, matched to two control sites (three in total) with 12 observation periods per cluster will provide 80% power ($p = .05$, two-tailed test) to detect a difference between 20 (SD = 27) counts per observation period in the control group and 80 (SD = 27) in the intervention group. This assumes an ICC of 0.02 (this low ICC value has been used in previous studies that have evaluated homogeneous parks (B. Han, personal communication)). The SD (27) and the correlation between sites (0.37) is taken from a previous feasibility study of two small spaces (not canals) [7]. However, due to the lack of data from previous studies on changes in numbers of people using canals, data from the feasibility data was the most suitable available data to inform the power calculation.

5.2.5.2 Intercept surveys

One researcher will be conducting the intercept surveys. Therefore, the desired sample size (n = 48) was calculated based on a feasible sample size that one researcher could attain across the two time points. Assuming one person can realistically survey no more than four people in one hour, the maximum number of people that one person can survey across the two points is 96 participants; this is because there are a total of 24 hours of survey data collection. However, there were 50 observed adults and older adults who used the unimproved canal path at baseline (i.e. entered the target area), which is an average of approximately four people per hour. If we assume that 50% of these people can be recruited for the intercept surveys, then this gives us a predicted sample size of 48 participants; assuming a survey completion rate of two participants per hour across the two time points.

5.2.6 Stopping Rule

5.2.6.1 Systematic observation

Observations will be carried out during the pre-specified observation periods; hence there is no stopping rule for the observations.

5.2.6.2 Intercept surveys

If the desired sample size (n = 48) has not been reached after twenty-four hours of data collection across two time points, a further three days of intercept surveys will be conducted using the same sampling schedule i.e. four times a day across three days

(Tuesday, Wednesday, Saturday). This would result in a further twelve hours of data collection.

5.3 Variables

5.3.1 Manipulated variables

N/A (this is an observational study)

5.3.2 Measured variables

5.3.2.1 Systematic observation

Systematic observation of activity in each of the intervention and comparison sites will be carried out using MOHAWk [1]: an observation tool that measures three levels of physical activity intensity (Sedentary, Walking and Vigorous) and two other behavioural indicators of wellbeing (Connect: connecting with other people; and Take Notice: taking notice of the environment). This tool applies interval time sampling techniques using continuous observation of activities and characteristics of all individuals entering predefined target areas during hour-long observation periods. MOHAWk requires observers to record the following characteristics for all observed participants: age, gender, ethnicity and whether they require assistance to move. Observers also document weather conditions (precipitation) and incivilities in the target area (e.g. general litter, graffiti). MOHAWk is a newly developed tool, for which there is preliminary evidence of validity [9].

MOHAWk was chosen because existing validated tools for systematic observation of physical activity, such as SOPARC (System for Observing Play and Recreation in Communities) [10], use momentary scans of activity and were developed for outdoor environments that attract consistently high numbers of users or large groups (e.g. large regional parks). The planned study will evaluate narrow linear green space along a canal that has lower numbers of users and less consistent usage; thus momentary scans would be unable to reliably capture people's activity within or passing through that space.

MOHAWk also measures behavioural indicators of wellbeing; wellbeing has predominantly been measured used self-report, which is more susceptible to recall bias and poor response rates [11].

5.3.2.2 Intercept surveys

An intercept survey was developed for the purposes of the present study (see Appendix N). The intercept survey was developed to assess: (1) reasons for using the canal; (2) awareness and views regarding the intervention; and (3) changes in use of the canal and changes in physical activity. The age group (adult or older adult), gender (male or female), ethnicity (white or non-white) and primary activity (e.g. walking, cycling) of each participant will be estimated using the same procedures specified in MOHAWk.

5.4 Design plan

5.4.1 Study type

Observational study.

5.4.2 Blinding

No blinding is involved in this study.

5.4.3 Study design

This is a prospective controlled before and after natural experimental study (repeated cross-sectional).

5.4.4 Interventions

The following descriptions are in line with the Template for Intervention Description and Replication (TIDieR) checklist [12].

This study is set in Greater Manchester (GM); a large metropolitan county in North West England with a population of around 2.8 million and containing ten metropolitan boroughs.

This study will evaluate an urban canal intervention designed and implemented by Peel Land and Property Group (hereafter referred to as 'Peel'): a private company which owns and manages 13 million square feet of property and 33,000 acres of land and water across the United Kingdom (UK); this portfolio is valued at £2.3 billion. Peel state that the aims of the intervention are:

> '…to enhance existing recreational provision which is used by the local community and the key connections between these recreational spaces. The improvements will enable and encourage circular walks to be undertaken by the local community in Boothstown. The improvements will help in creating pleasant routes for recreation and exercise to promote health and wellbeing, while also increasing biodiversity.' [13]

The intervention site is located in Boothstown and Ellensbrook: a suburban ward with a population of 9,532, located in the city of Salford [14]. Although Salford is ranked as the 22nd most deprived local authority in England [15], Boothstown and Ellensbrook is ranked as the 2nd least deprived ward in Salford out of 20 wards [15] (based on the Index of Multiple Deprivation (IMD) Score [16]). Also, there are no areas in Boothstown and Ellenbrook that are within the 30% most deprived in England [14].

The intervention is located around the Bridgewater Canal and Bridgewater Nature Park (see Figure 5.5). The Bridgewater Canal is a 65 km canal in North West England which partly runs through Salford. The Bridgewater Canal is owned and operated by The Bridgewater Canal Company Limited, part of The Peel Group, in conjunction with the Bridgewater Canal Trust. Bridgewater Nature Park is a semi-natural area of green space covering approximately 37 acres located in Boothstown and Ellensbrook next to the Bridgewater Canal.

There are six components of the intervention, which are split into two phases.

5.4.4.1 First phase

The first phase of the intervention contains four components, which are planned to be implemented between December 2017 and May 2018. The largest component of the intervention is on the northern footpath on the Bridgewater Canal, which includes resurfacing of the path (for approximately 1 km), management of vegetation, installation of benches, new directional signage and new informal play equipment (implementation December 2017 – May 2018).

Another component focuses on Moss House Lane; this is a route that provides access to the northern footpath on the Bridgewater Canal. Changes along Moss House Lane include the upgrade of surfacing and management of vegetation (implementation April – May 2018).

The final two components of the intervention focus on existing green spaces: Bridgewater Nature Park and Vicars Hall Village Green. At Bridgewater Nature Park, changes include resurfacing of footpaths, replacement and installation of benches, clearance and management of existing vegetation, and new interpretation boards and directional signage (implementation January – March 2018). At Vicars Hall Village Green, changes include replacement seating areas, tree and bulb planting, improved access points and resurfacing of the existing footpath (implemented January – February 2018). The Neighbourhood Green Space Tool (NGST) [17], a UK validated environmental audit tool for measuring the quality of green space, will be used to measure the environmental changes at Vicars Hall Village Green and Bridgewater Nature Park.

5.4.4.2 Second phase

The second phase of the intervention focuses on changes along Vicars Hall Lane. One component of the intervention is a new footpath along the side of the existing track (implementation in February 2019 to coincide with a new housing development due to be completed between 2020 and 2021). Another component of the intervention along Vicars

Hall Lane will provide new footpath links directly to the new housing development site (implementation February 2019).

The total estimated cost for the implementation of all six components of the intervention and maintenance for the next 20 years is £920,000 (plus VAT). All components of the intervention will be implemented by the Peel team who have extensive experience in the development of land across a broad range of sectors and in a multitude of locations (C. Culshaw, personal communication).

A description of the key characteristics and locations of the intervention and comparison sites can be found in Table 5.1 and Figure 5.6, respectively.

Figure 5.5. Map showing the location of all components of the intervention. *Taken from http://www.peellandp.co.uk/boothstown-recreational-improvements*

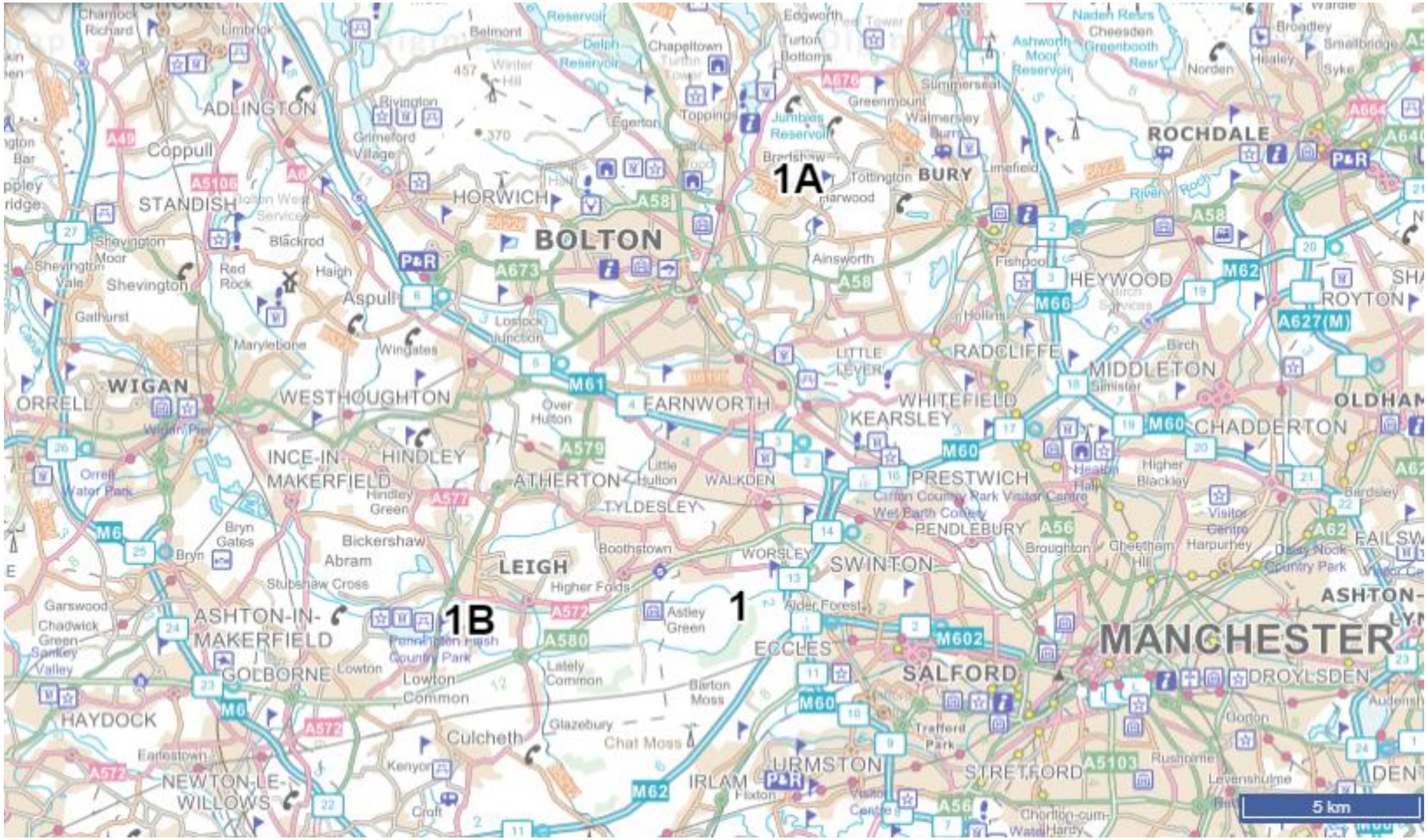


Figure 5.6. Map showing the location of all intervention and comparison sites in Greater Manchester. *The numbers/ letters refer to intervention and comparison sites i.e. '1' is Intervention site 1, and '1A' and '1B' are the corresponding comparison sites: Comparison site 1A and Comparison site 1B, respectively. See Table 5.1 for details on the intervention and comparison sites. Maps drawn using: www.digimap.edina.ac.uk. © Crown Copyright and Database Right (2018). Ordnance Survey (Digimap Licence).*

Table 5.1. Key characteristics of the intervention and comparison sites at baseline.

Intervention and comparison sites	**Location (postcode)**	**LSOA [a]**	**Site level features**	**Population density (persons Ha^{-1})[b]**	**Intersection density (per $1000m^{-2}$)[c]**	**IMD[d]**	**NDVI[e]**	**Walk Score[f]**
Intervention site	M28 1JD	Salford 013G	Canal waterway with two canalside paths (one unsurfaced and one surfaced); two access routes; a bridge; no benches; no lighting; links to Bridgewater Nature Park and Vicars Hill Village Green	22.67	9.18	4.28	0.55	35
Comparison site 1A	BL2 3EQ	Bolton 002D	Brook with two brookside paths (both unsurfaced); two access routes; a footbridge; no benches; no lighting; links to Jumbles Country Park	23.91	7.58	8.37	0.54	48
Comparison site 1B	WN7 4QP	Wigan 033C	Canal waterway with two canalside paths (one unsurfaced and one surfaced); two access routes; a footbridge; no benches; no lighting; links to Pennington Flash Country Park	37.93	9.68	27.05	0.52	50

[a] Lower Layer Super Output Area (LSOA): census reporting units containing between 1000 and 3000 individuals
[b] Population density: number of persons per hectare; used as a proxy measure of residential density
[c] Intersection density: the number of 3-way junctions standardised by LSOA area; used as a measure of street connectivity
[d] Index of Multiple Deprivation score (IMD) [16]: an area deprivation score that combines several indicators of deprivation including income, employment, health and crime. Higher scores indicate more deprived areas.
[e] Normalised Difference Vegetation Index (NDVI): a validated normalised scale of healthy vegetation cover; used as a measure for presence of greenery at the neighbourhood-level. Higher scores indicate areas with more healthy vegetation cover.
[f] WalkScore uses a Google search algorithm to calculate a weighted score (1-100) based on the number and accessibility of amenities (such as shops and parks) within a 1-mile radius of a user-entered postcode, whereby closer amenities with the most accessible walking routes are weighted more strongly; used as a measure of 'access to/ availability of destinations and services'. Higher scores indicate more 'walkable' areas.

5.4.5 Logic model

Hypothesised causal pathways are outlined in the logic model (Figure 5.7), which is based on the framework suggested by Panter et al. [18]. It is proposed that the changes will impact a number of key variables that will encourage more people to use the canal; particularly the convenience and pleasantness of the route (i.e. the northern footpath on the Bridgewater Canal) and accessibility and pleasantness of key destinations (i.e. Bridgewater Nature Park and Vicars Hall Village Green). This, in turn, will lead to increases in physical activity, social interaction and overt appreciation of the canal routes and destinations.

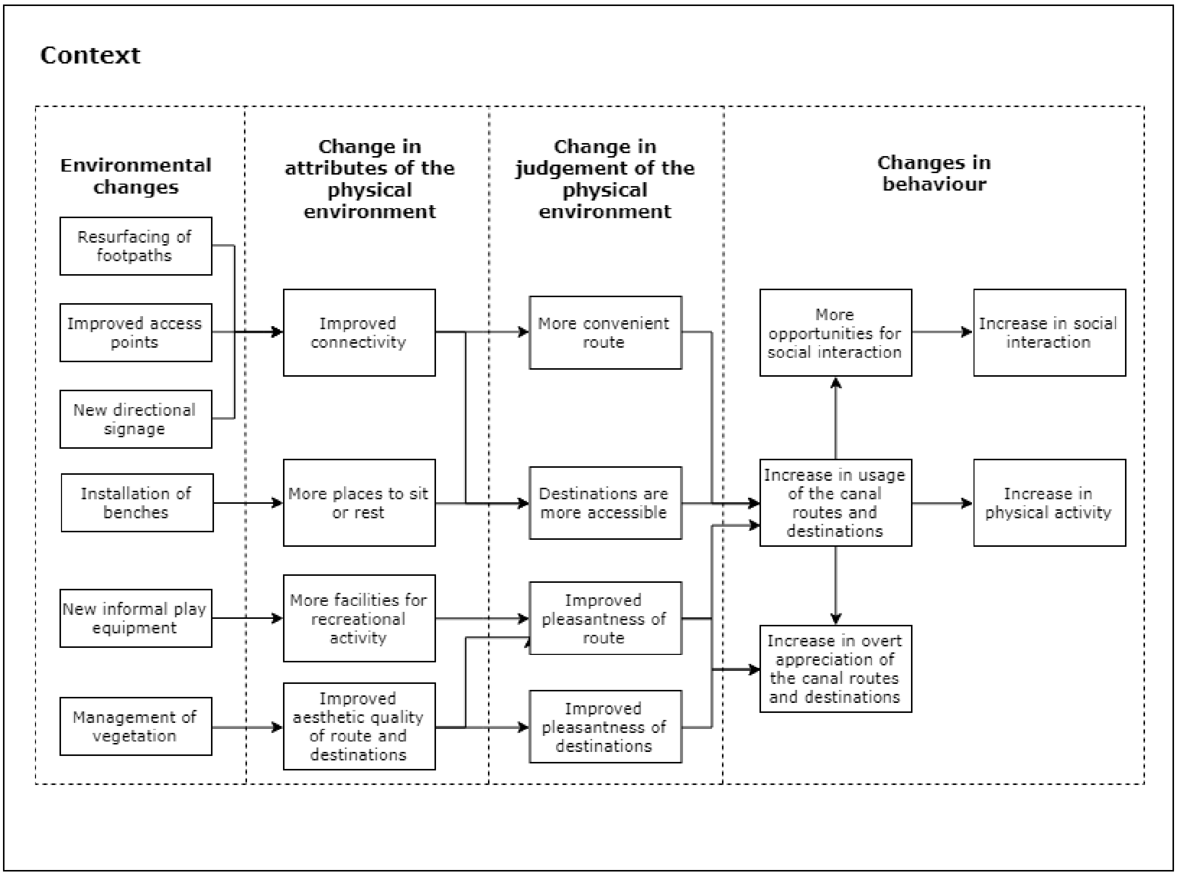


Figure 5.7. Urban canal intervention logic model

5.4.6 Confounders

5.4.6.1 Weather

Observations will be carried out regardless of weather conditions, unless weather conditions become so extreme that they compromise the observer's safety. To control for the confounding influence of weather, the observer will record the duration of any precipitation that occurs during each observation period. These data will inform a sensitivity analysis.

5.4.6.2 Other known confounders

Data collection at all time points will be carried out during UK school term dates to control for the change of activity around school holidays. A liaison from Peel will be contacted before data collection at each time point to enquire about any unrelated significant events that could influence the outcomes, such as other unrelated planned changes to the built environment. Media outlets (e.g. Twitter) will also be monitored to check for any significant events near to sites during data collection periods at each time point. Where possible, data collection will be arranged so that it does not co-occur with any unrelated significant events.

5.4.7 Randomisation

N/A

5.5 Analysis plan

5.5.1 Statistical models

5.5.1.1 Inter-rater reliability

High inter-rater reliability for MOHAWk has previously been established between pairs of observers across four studies for assessing people's behaviours and their characteristics (intraclass correlation coefficients (ICC) > 0.8) [7,9]. Inter-rater reliability will be calculated between each pair of observers at each time point to assess agreement on demographic and activity categories. Inter-rater reliability will be analysed using single rater two-way random effects ICC.

5.5.1.2 Systematic observation

The analysis will estimate the effect of the intervention on the number of adults and older adults using the intervention site at each of the three time points, and the types of behaviours they engage in per observation period, compared to the two corresponding comparison sites. The analysis will control for known confounders, including weather, day and time. The primary analysis will analyse data for all adults, and a secondary analysis will analyse data for all older adults using the same methods.

The primary outcome will be an overall count of adults per observation period at 12 months, which is six months after the first phase of the intervention is completed. A secondary outcome will be an overall count of adults per observation period at 24 months, which is nine months after the second phase of the intervention is completed. Further secondary outcomes will be a count per observation period for Walking and Vigorous behaviour. Exploratory outcomes will be a count per observation period for Sedentary, Connect and Take Notice behaviour.

For each outcome we will follow three steps. Firstly, using a dataset that only includes the baseline data, and is blinded to group allocation, we will build a regression model to examine the relationship between the baseline count outcome and the covariates (weather, day, time). The overall count of adults per observation period will be used as an additional covariate when analysing each of the behaviours (i.e. Sedentary, Walking, Vigorous, Take Notice, Connect). We will consider this count outcome as either a Poisson distribution, a zero-inflated Poisson or a normal distribution. We will consider all these approaches, and we will choose the most suitable approach by seeing which model has the best fit, assessed using Akaike's information criterion (AIC) and measures of over-dispersion. Secondly, once we have developed a suitable model with the baseline data, we will combine the pre- and post-intervention data together and apply the model from step one. These first two steps will develop a model of the effect of known covariates on behaviour, blinded to group allocation: this will enable us to ensure we control for appropriate confounders in the main analysis.

The final step will be to test the effect of the experimental group on the outcome, controlling for known confounders: we will add into the model Group (intervention or comparison) and Period (pre or post). The treatment effect will be the coefficient for the interaction of Group and Period. This is a form of Difference in Differences analysis [19,20].

5.5.1.3 Intercept surveys

Descriptive statistics (e.g. means, frequencies, percentages) will be used to summarise characteristics of the sample and the survey responses. No inferential statistics will be carried out for the survey data.

5.5.2 Inference criteria

5.5.2.1 Systematic observation

We will consider the count outcomes as either a Poisson distribution, a zero-inflated Poisson or a normal distribution. We will consider all these approaches, and we will choose the most suitable approach by seeing which model has the best fit, assessed using Akaike's information criterion (AIC) and measures of over-dispersion.

5.5.2.2 Intercept surveys

N/A

5.5.3 Data exclusion

5.5.3.1 Systematic observation

A sensitivity analysis will be conducted to assess for any potential bias in the analysis of the primary and secondary outcomes. Observation periods will be removed for the sensitivity analysis if there is any precipitation that lasts for more than 50% of the observation period i.e. an overall accumulated duration of 30 minutes or more (recorded by the observer). This is in line with recommendations from MOHAWk [1].

5.5.3.2 Intercept surveys

N/A

5.5.4 Missing data

5.5.4.1 Systematic observation

Procedures in the MOHAWk were developed to avoid incomplete or missing data. For instance, any missed observations (e.g. due to illness) will be rescheduled. Therefore, there are no analysis plans for missing data.

5.5.4.2 Intercept surveys

Analyses for the survey data are descriptive and thus do not require statistical methods to deal with missing data. Further, item non-response is unlikely because the survey is brief (no more than 10 minutes), and there are no questions based on sensitive issues or topics. Therefore, there are no analysis plans for missing data.

5.5.5 Exploratory analysis

An exploratory analysis will be conducted to explore the difference in the total number of adults and older adults using the northern footpath at the intervention site (i.e. entering the target area), compared to the total number of adults and older adults who did not enter the target area but accessed or exited the southern side of the canal via Moss House Lane. Firstly, a line graph will be used to plot the total number of adults and older adults (y-axis) at each time point (x-axis), with one line showing participants who entered the target area and one line showing participants who were counted but did not enter the target area. Secondly, if the data is normally distributed, a two-sample t-test will be used to compare the mean change score (mean change from baseline to 12 month follow-up) between participants who entered the target area and participants who were counted but did not enter the target area. If the assumption of normality is not met, a Mann-Whitney U test will be used instead.

5.6 Protocol amendments

Subsequent to publication of this protocol, we submitted protocol amendments on 21st November 2019; this section reports the amendments.

5.6.1 Timing of amendments

At the time of the amendments, we have collected baseline, 7-month and 12-month follow-up observation data. We have collected 7-month and 12-month intercept survey data. The final 24-month follow-up observation data collection began on 9th November 2019, and is due to finish on 27th November 2019. Final intercept survey data collection begins on 3rd December 2019 and will finish on 7th December 2019. We have inputted all data collected so far; however, we have only analysed baseline data. The following amendments have been made.

5.6.2 Amendments to the analysis plan

We initially proposed that we would consider the count outcomes as either a Poisson distribution, a zero-inflated Poisson or a normal distribution. From analysing the baseline data (as originally planned), and based on our experience in a recently completed similar natural experimental study that used the same outcome measures [7], it is unlikely that any of these distributions will appropriately fit the data. Therefore, we will firstly consider the count outcomes as negative binomial, which is preferred for modelling outcomes that are overdispersed i.e. where the variance is greater than the mean. Negative binomial is an extension of the Poisson regression model and can adjust for overdispersion. If a negative binomial model is inappropriate for any of the outcomes, then we will attempt to use an ordinal logistic regression model. If this approach is not suitable for the raw count data, we will attempt to group the outcome data e.g. 0, 1-2, 2-

3, ≥4. Finally, if an ordinal logistic regression is not suitable for any of the outcomes, we will consider fitting a binary logistic regression by classifying the count outcome as 0 and ≥1. Each of the models will be rejected if the outcome does not follow the appropriate distribution or if the model fails to fit in Stata.

We initially proposed that we would build an appropriate regression model using the baseline data only, blinded to group allocation. However, we will build an appropriate regression model using baseline data and follow-up data (including the covariates prespecified in the protocol i.e. weather, day of week, time of day), still blinded to group allocation. We need to include follow-up data when building an appropriate regression model to ensure the model fits the whole dataset. All remaining steps will be conducted as originally specified i.e. once we have developed a suitable model with the baseline data and follow-up data, we will then add into the model Group (intervention or comparison).

5.6.2 Amendments to the intercept survey

After conducting intercept surveys at 7-month follow-up, we decided to reduce the number of questions in the surveys to improve poor response rates and incomplete surveys. We have kept questions that will be most important in understanding reasons for any possible behaviour change measured using observations and will therefore improve causal inference (e.g. questions around changes in behaviour). We have also reworded some of the questions to improve clarity. We have attached the amended intercept survey that was used at 12-month follow-up, and will be used again at 24-month follow-up.

5.7 References

1. Craig P, Cooper C, Gunnell D, Hajow S, Lawson K, Macintyre S, Ogilvie D, Petticrew M, Reeves B, Sutton M, Thompson S. Using natural experiments to evaluate population health interventions: new Medical Research Council guidance. J Epidemiol Community Health. 2012;66(12):1182-6.
2. Benton JS, Anderson J, Hunter RF, French DP. The effect of changing the built environment on physical activity: a quantitative review of the risk of bias in natural experiments. Int J Behav Nutr Phys Act. 2016;13(1):107.
3. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJ, Martin BW, Lancet Physical Activity Series Working Group. Correlates of physical activity: why are some people physically active and others not?. The Lancet. 2012;380(9838):258-71.
4. Moran M, Van Cauwenberg J, Hercky-Linnewiel R, Cerin E, Deforche B, Plaut P. Understanding the relationships between the physical environment and physical activity in older adults: a systematic review of qualitative studies. Int J Behav Nutr Phys Act. 2014;11(1):79.
5. Barnett DW, Barnett A, Nathan A, Van Cauwenberg J, Cerin E. Built environmental correlates of older adults' total physical activity and walking: a systematic review and meta-analysis. Int J Behav Nutr Phys Act. 2017;14(1):103.
6. Cerin E, Nathan A, Van Cauwenberg J, Barnett DW, Barnett A. The neighbourhood physical environment and active travel in older adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act. 2017;14(1):15.
7. Benton JS, Anderson J, Cotterill S, Dennis M, Lindley SJ, French DP. Evaluating the impact of improvements in urban green space on older adults' physical activity and wellbeing: protocol for a natural experimental study. BMC Public Health. In press.
8. Donner A, Klar N. Design and Analysis of Cluster Randomization Trials in Health. London: Arnold Publishers; 2000.

9. Benton JS, Anderson J, Cotterill S, Hunter RF, Pulis M, French DP. Method for Observing pHysical Activity and Wellbeing (MOHAWk): development of an observation tool to measure physical activity and wellbeing in urban spaces. Under development.
10. McKenzie TL, Cohen DA, Sehgal A, Williamson S, Golinelli D. System for Observing Play and Recreation in Communities (SOPARC): reliability and feasibility measures. J Phys Act Health. 2006;3 Suppl 1:208-22.
11. National Research Council. Subjective well-being: Measuring happiness, suffering, and other dimensions of experience. Washington, DC: The National Academies Press; 2013.
12. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, Altman DG, Barbour V, Macdonald H, Johnston M, Lamb SE. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. BMJ. 2014;348:1756-1833.
13. Peel Land and Property Group. Improvements to green spaces and recreational routes in Boothstown Salford. Available from: http://peellandp.co.uk/boothstown-recreational-improvements/?rq=boothstown
14. Salford City Council. Boothstown and Ellenbrook Ward profile - March 2016. 2016. Available from: https://www.salford.gov.uk/people-communities-and-local-information/my-local-community/ward-profiles/
15. Salford City Council. Indices of Deprivation 2015 – the City of Salford's position. 2015. Available from: https://www.salford.gov.uk/media/388062/index_of_multiple_deprivation_report_2015.pdf
16. Department for Communities and Local Government. English Indices of Deprivation 2015 [computer file]. 2015. Downloaded from: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015.

Licensed under: https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3. Accessed 27 April 2017.

17. Gidlow CJ, Ellis NJ, Bostock S. Development of the neighbourhood green space tool (NGST). Landsc Urban Plan. 2012;106(4):347-58.
18. Panter J, Guell C, Prins R, Ogilvie D. Physical activity and the environment: conceptual review and framework for intervention research. Int J Behav Nutr Phys Act. 2017;14(1):156.
19. Blundell R, Costa Dias M. Evaluation methods for non-experimental data. Fisc Stud. 2000;21(4):427-68.
20. Craig P, Katikireddi SV, Leyland A, Popham F. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. Annu Rev Public Health. 2017;38:39-56.

Chapter 6. A natural experimental study of improvements along an urban canal: impact on canal usage, physical activity and other wellbeing behaviours

This article has been submitted for publication in the International Journal of Behavioral Nutrition and Physical Activity (IJBNPA) and is currently under peer review:

- Benton JS, Cotterill S, Anderson J, Macintyre VG, Gittins M, Dennis M, French DP. A natural experimental study of improvements along an urban canal: impact on canal usage, physical activity and other wellbeing behaviours. International Journal of Behavioral Nutrition and Physical Activity. Under review.

6.1 Abstract

Background: There are few robust natural experimental studies of improving urban green spaces on physical activity and wellbeing. The aim of this controlled natural experimental study was to examine the impact of green space improvements along an urban canal on canal usage, physical activity and two other wellbeing behaviours (social interactions and taking notice of the environment) among adults in Greater Manchester, UK. The intervention included resurfaced footpaths, removal of encroaching vegetation, improved entrances, new benches and signage.

Methods: Two comparison sites were matched to the intervention site using a systematic five-step process, based on eight correlates of physical activity at the neighbourhood (e.g. population density) and site (e.g. lighting) levels. Outcomes were assessed using systematic observations at baseline, and 7, 12 and 24 months post-baseline. The primary outcome was the change in the number of people using the canal path from baseline to 12 months. Other outcomes were changes in physical activity levels (Sedentary, Walking, Vigorous), Connect and Take Notice behaviours. Data were analysed using multilevel mixed-effects negative binomial regression models, comparing outcomes in the intervention group with the matched comparison group, controlling for day, time of day and precipitation. A process evaluation assessed potential displacement of activity from a separate existing canal path using intercept surveys and observations.

Results: The total number of people observed using the canal path at the intervention site increased more than the comparison group at 12 months post-baseline (IRR 2.10, 95% CI 1.79 – 2.48); there were similar observed increases at 7 and 24 months post-baseline. There was some evidence that the intervention brought about increases in walking and vigorous physical activity, social interactions, and people taking notice of the environment. The process evaluation suggested that there was some displacement of activity, but the intervention also encouraged existing users to use the canal more often.

Conclusions: Urban canals are promising settings for interventions to encourage green space usage and potentially increase physical activity and other wellbeing behaviours. Interventions that improve access to green corridors along canals and provide separate routes for different types of physical activities may be particularly effective and warrant further research.

Study protocol: Study protocol published in Open Science Framework in July 2018 before the first follow-up data collection finished (https://osf.io/zcm7v). Date of registration: 28 June 2018.

6.2 Background

There is mounting evidence that the use of urban green spaces can provide a range of mental and physical health benefits [1–3]. Urban green spaces are open spaces in urban areas wholly or partly covered by vegetation, ranging from 'man-made' spaces (e.g. parks) to more 'natural' spaces (e.g. woodlands), and can include landscapes around blue spaces (e.g. canals) [4]. Due to rapid global urbanisation, more people are living in dense built-up urban areas with limited access to high quality green space [5]. Therefore, improving or creating new urban green spaces is a potentially promising population-level intervention to promote health and wellbeing [6].

Numerous behavioural pathways have been proposed to explain how urban green spaces may promote health and wellbeing [7]. One of these behavioural pathways is physical activity, since many different physical activities can be performed in green spaces, including walking, running, cycling and sports; all of which have positive effects on physical and mental health [8,9]. Physical activity undertaken in green spaces ('green exercise') may be more beneficial than physical activity done in other settings [10]. In addition to physical activity, urban green spaces can facilitate social interactions and cultivate social cohesion [11]; social relationships are critical for health and wellbeing [12]. Green spaces also offer opportunities for people to get outdoors and take notice of nature [13,14], which can have important benefits for wellbeing [14,15].

While there are multiple plausible behavioural pathways by which urban green spaces may influence health and wellbeing, the majority of research to date has focused on physical activity [16]. Growing cross-sectional evidence has found positive associations between accessibility, availability and quality of urban green spaces and physical activity [6]. However, cross-sectional evidence cannot determine causality and tells us little about how to design new or improve existing urban green spaces to increase physical activity. Due to researchers' lack of control over environmental changes, making use of natural experiments is the optimal approach to understand the causal effects of environmental interventions [17]. Natural experiments are real world interventions that are not under the control of researchers. Researchers can design studies around a natural experiment to compare exposed and unexposed groups (or groups with different levels of exposure) to assess intervention effectiveness; known as natural experimental studies [18].

There is a dearth of natural experimental studies on the effect of urban green space interventions on physical activity and these studies have weak designs. A recent review of systematic reviews assessed the risk of bias in natural experimental studies of built environment interventions on physical activity [19]. From three existing exemplar systematic reviews, the researchers identified only 12 natural experimental studies (15 physical activity outcomes) with comparison groups, which included nine studies of urban green space interventions. All outcomes in these studies were assessed as having critical (n = 12) or serious (n = 3) risk of bias, thus showing that natural experimental studies conducted to date have questionable internal validity; a conclusion in line with other recent systematic reviews [16,20–22].

Eight recommendations have been made to improve the methodological rigour of future natural experimental studies of environmental interventions [19], influenced by Medical Research Council (MRC) guidance for using natural experiments [18]. The present study implemented these recommendations: (1) publish study protocols with a priori analyses specified; (2) conduct sample size calculations; (3) better matching of comparison sites; (4) use multiple comparison sites; (5) use adequate outcome measurements; (6) measure intervention exposure at the individual level; (7) control for key confounders in statistical analyses; and (8) better reporting of samples and interventions.

Another limitation of the evidence base to date is that few natural experimental studies of urban green space interventions have been conducted in the United Kingdom

(UK) and elsewhere in Europe; most have been in the United States (US) [17]. There are important differences between the UK and US that influence physical activity and make it hard to generalise findings, such as climate, residential density and population characteristics [23]. Also, most US studies are of park-based or large-scale redevelopment interventions [16]. As a result, the impact of small-scale replicable urban green space interventions, such as improving canals, greening residential streets or creating informal amenity green spaces, remains largely unknown.

Systematic observation (i.e. direct observations of behaviour using predetermined criteria) offers an objective method of unobtrusively assessing a range of behaviours that are important for wellbeing. The New Economics Foundation (NEF) conducted a review of the wellbeing literature on behalf of the UK Government's Foresight project and identified five evidence-based behaviours that improve wellbeing (hereafter referred to as 'wellbeing behaviours'), known as the 'Five Ways to Wellbeing' [24]: Be Active (engage in physical activity); Take Notice (awareness of one's internal and external environment); Connect (socially interact with others); Keep Learning (acquire knowledge or skill in something new); and Give (pursue altruistic activities). The present study uses Method for Observing pHysical Activity and Wellbeing (MOHAWk) [25]: a recently validated observation tool which assesses three of these wellbeing behaviours (Be Active, Take Notice, Connect) in urban spaces.

6.2.1 Research aim and objectives

The present study took advantage of an opportunity for a prospective natural experimental study. A private developer planned to implement new walking infrastructure and green space improvements along an urban canal in the UK (hereafter referred to as 'canal improvements'). Canals are inland waterways that may produce important health and wellbeing benefits [26]; for example, areas surrounding canals can provide opportunities for a range of recreational activities, such as walking, jogging and cycling. To date, there are no natural experimental studies of the impact of interventions along urban canals on human health and wellbeing [27]. Growing investment into the regeneration of urban canals in the UK [28,29] and elsewhere in Europe [30] increases the importance of producing robust practice-based evidence to inform decision makers how urban canals can effectively be used to improve population health and wellbeing.

The aim of this two-year natural experimental study was to evaluate the impact of the canal improvements on canal usage, physical activity and two other wellbeing behaviours (social interactions and taking notice of the environment) in adults. In line with the published study protocol [31], pre-planned objectives were to examine whether the canal improvements increased the following at 7, 12 and 24 months post-baseline, compared to two matched comparison sites where no changes occurred:

1. Total number of people (primary outcome at 12 months);
2. Walking and Vigorous physical activity behaviours (secondary outcomes);
3. Sedentary, Connect and Take Notice behaviours (exploratory outcomes).

An additional objective was to explore potential displacement of activity from a separate existing canal path, through process evaluation using intercept surveys and systematic observation.

6.3 Methods

The study protocol is published elsewhere [31].

6.3.1 Study design

This was a prospective controlled natural experimental study, with an intervention group (one intervention site) and a matched comparison group (two matched comparison sites). The study was set in Greater Manchester, UK: a large metropolitan authority containing ten boroughs.

6.3.2 Intervention

The intervention was implemented in the city of Salford, which is the 22nd most deprived local authority in England [32]. However, the intervention was located in an affluent area within Salford called Boothstown and Ellensbrook (population = 9,532), which is the 2nd least deprived ward in Salford [33].

The intervention was implemented along the north side of an urban canal (Bridgewater Canal) and nearby connecting footpaths. There is an existing well-surfaced towpath on the south side of the canal, finished with a tarmac/ gravel surface, which has no vegetation other than a narrow grass strip between the path and the canal water. The intervention was designed, funded and implemented by Peel Land and Property Group (hereafter referred to as 'Peel'): a private company which owns and manages 13 million square feet of property and 33,000 acres of land and water across the UK. According to Peel, the aim of the intervention was:

> '… to enhance existing recreational provision which is used by the local community and the key connections between these recreational spaces. The improvements will enable and encourage circular walks to be undertaken by the local community in Boothstown. The improvements will help in creating pleasant routes for recreation and exercise to promote health and wellbeing, while also increasing biodiversity.' [34]

There were two phases of intervention implementation. The main phase (Dec 2017 to May 2018) included new footpaths; resurfacing of existing footpaths using golden gravel; enhancement of an existing nature park and village green (e.g. improved entrance points); removal of encroaching vegetation; new benches, signage and informal play equipment. The second phase (Feb 2019) included new footpaths which link to a new nearby residential housing development due to be completed in 2021, although new residents started to occupy houses from early 2019. The total estimated cost of the intervention and 20-year ongoing maintenance is £920,000 (≈ $1,242,000 US dollars), excluding Value Added Tax. Figure 6.1 shows the intervention site pre and post-intervention. Figure 6.2 shows the intervention path on the north side of the canal and the existing unchanged path on the south side.

6.3.3 Comparison group

We used a recently developed five-step process to identify two matched comparison sites in Greater Manchester (Figure 6.3) [35]. The eight variables used for matching were based on nine systematic reviews of physical activity environmental correlates in adults [36–41] and older adults [42–44]. In brief, the first step involves identifying the most closely matched neighbourhoods to the index intervention neighbourhood, using spatial data at the Lower Layer Super Output Area (LSOA) level (population density, street connectivity, deprivation, neighbourhood greenness). The next four steps involve searching for the most closely matched comparison sites within the potential matched neighbourhoods identified in step one, using variables at the site level (e.g. footpath, benches, lighting). Steps two and three are conducted using Google Street View to narrow down potential matched comparison sites. Steps four and five involve in-person site audits. Appendix K provides further details of this matching process.

The two matched comparison sites were pooled together into one comparison group. The main reason for pooling the comparison sites was to increase statistical power. Also, including multiple comparison sites in the analysis provides increased confidence that any variation in confounding variables across comparison sites is offset, therefore reducing the risk that the intervention effect is confounded by site-specific variables in a single comparison site.

A description of the key characteristics of the intervention and two comparison sites can be found in Table 6.1. Figure 6.4 shows the comparison sites at baseline.

Table 6.1. Key characteristics of all study sites and LSOAs at baseline.

Site	Description of sites at baseline	Ward (postcode)	Walk Score [a]	LSOA [b] characteristics			
				Population density [c]	Intersection density [d]	IMD [e]	NDVI [f]
Intervention site	Canal with paths on both sides (one unsurfaced and one surfaced); two access routes; a bridge; no benches; no lighting; links to Bridgewater Nature Park and Vicars Hill Village Green	Boothstown and Ellenbrook (M28 1JD)	35	22.67	9.18	4.28	0.55
Comparison site 1A	Brook with paths on both sides (both unsurfaced); two access routes; a footbridge; no benches; no lighting; links to Jumbles Country Park	Bromley Cross (BL2 3EQ)	48	23.91	7.58	8.37	0.54
Comparison site 1B	Canal with two paths on both sides (one unsurfaced and one surfaced); two access routes; a footbridge; no benches; no lighting; links to Pennington Flash Country Park	Leigh West (WN7 4QP)	50	37.93	9.68	27.05	0.52

[a] WalkScore uses a Google search algorithm to calculate a weighted score (1-100) based on the number and accessibility of amenities (such as shops and parks) within a 1-mile radius of a user-entered postcode, whereby closer amenities with the most accessible walking routes are weighted more strongly, used as a measure of 'access to/ availability of destinations and services'. Higher scores indicate more 'walkable' areas;
[b] Lower Layer Super Output Area (LSOA): census reporting units containing between 1000 and 3000 people;
[c] Population density: number of persons per hectare; used as a proxy measure of residential density. Higher values indicate areas with higher population density;
[d] Intersection density: the number of 3-way junctions per $1000m^{-2}$ standardised by LSOA area; used as a measure of street connectivity. Higher values indicate areas with higher street connectivity;
[e] Index of Multiple Deprivation score (IMD) [78]: an area deprivation score that combines multiple indicators of deprivation including income, employment, health and crime. Higher values indicate more deprived areas;
[f] Normalised Difference Vegetation Index (NDVI): a validated normalised scale of healthy vegetation cover; used as a measure for presence of greenery at the neighbourhood-level. Higher values indicate areas with more healthy vegetation cover.

Pre intervention (Sept 2017) **Post intervention: 7 months post-baseline (June 2018)** **Post intervention: 12 months post-baseline (Nov 2018)**

Figure 6.1. Intervention site pre and post-intervention. *Photographs taken by Peel Land and Property Group and Jack Benton.*

Figure 6.2. Intervention path (left side) and unchanged path (right side) post-intervention. *Photograph taken by Jack Benton in June 2018.*

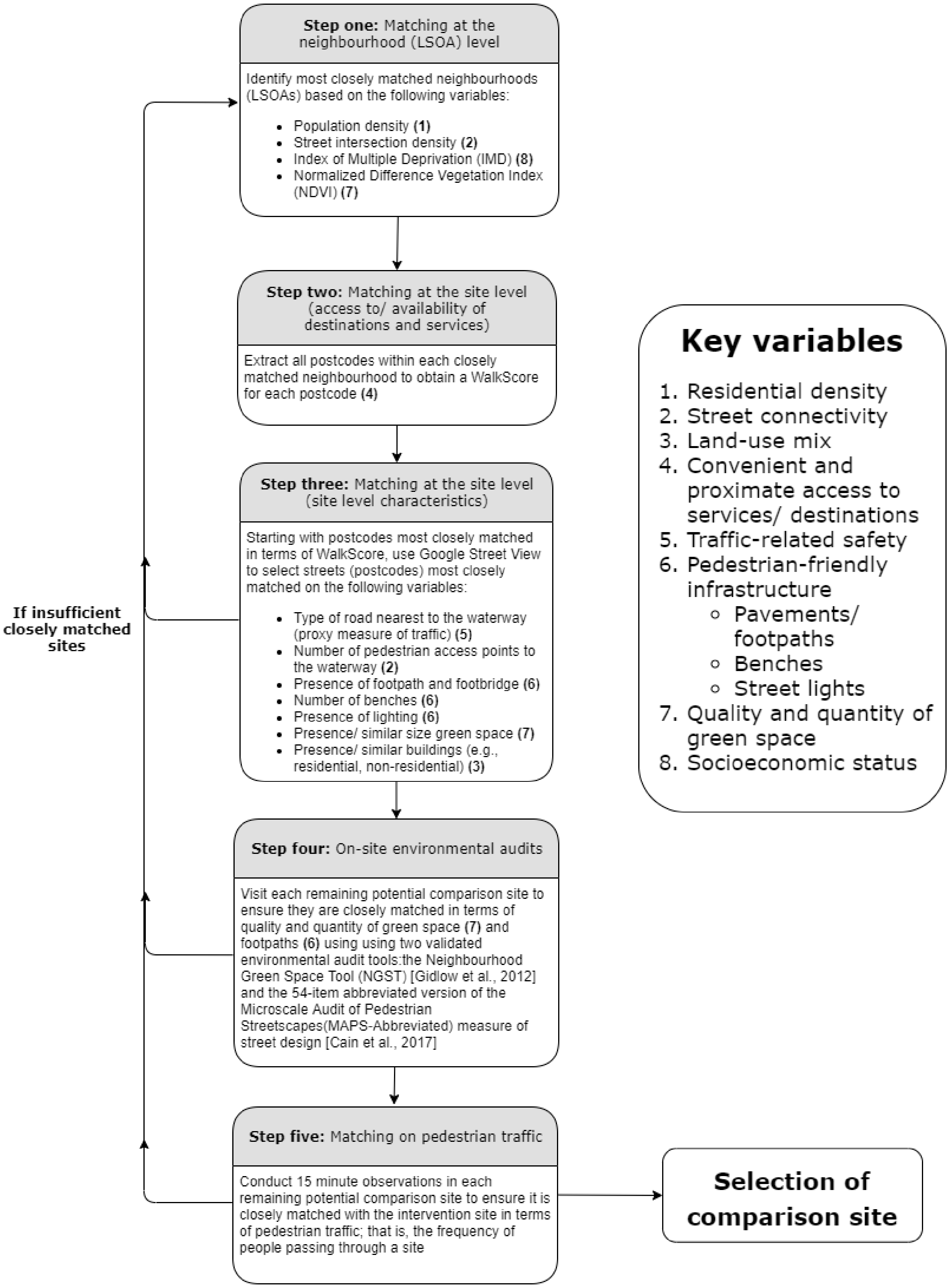


Figure 6.3. Overview of the five-step matching process used to identify comparison sites. *Numbers in brackets refer to the key variables used for matching. References for the environmental audit tools: Cain et al.* [45] *and Gidlow et al.* [46]*. Based on a similar graphic originally published in Benton et al.* [35].

Comparison site 1A

Comparison site 1B

Figure 6.4. Comparison sites at baseline. *Photographs taken by Jack Benton in Nov 2017.*

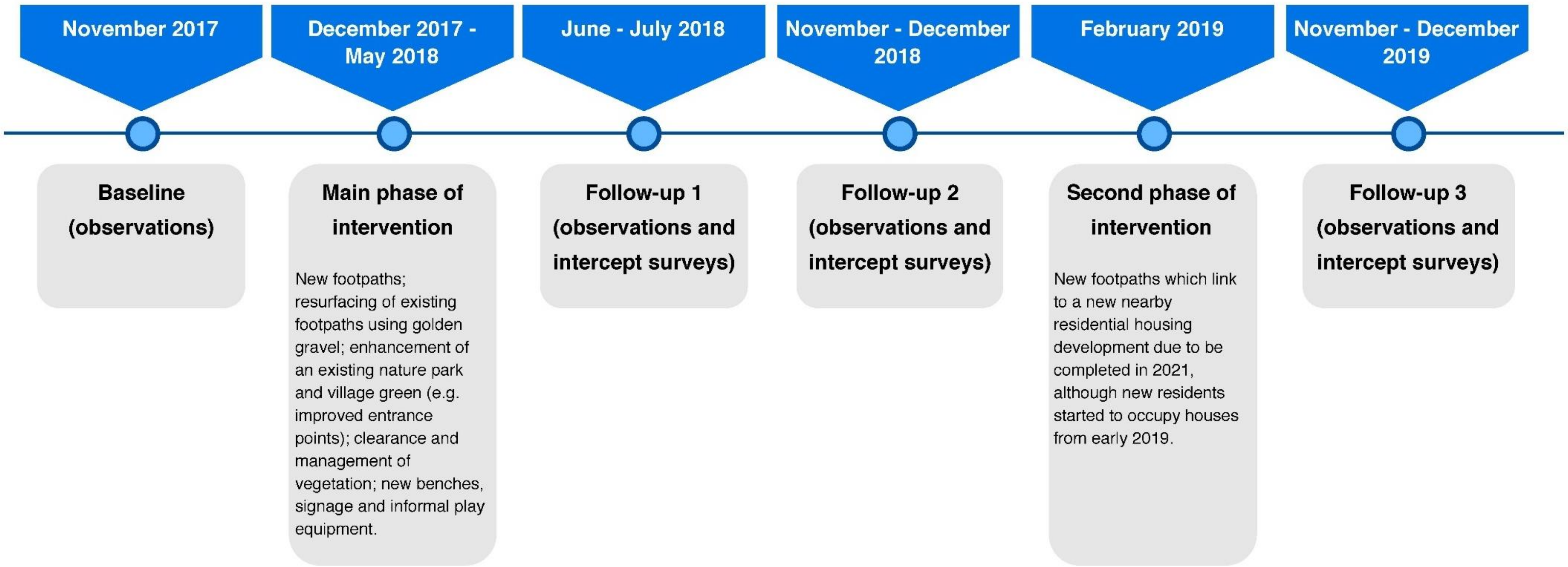


Figure 6.5. Timeline of data collection and intervention implementation.

6.3.4 Study sample

Data were collected on adults observed using the intervention and comparison sites.

6.3.5 Outcomes

Systematic observations were conducted at baseline (Nov 2017) and three separate follow-ups: 7 months (June 2018); 12 months (Nov 2018); and 24 months (Nov 2019) post-baseline. Observations were conducted during four hour-long observation periods (8-9am, 10-11am, 1-2pm, 3-4pm) on three days (Tuesday, Wednesday, Saturday) at each site i.e. twelve hours of observations for each site at each time point. However, at 24 months, two observation periods at the intervention site were cancelled because of safety concerns for the observer, which involved the observer being approached in a hostile manner by a member of the public. In line with the procedure specified in our approved ethics application, the researcher left the area immediately and we ceased all remaining observations at that site. This incident was reported to the institutional Research Ethics Committee; no further action was required. Figure 6.5 provides an overview of data collection and intervention timings.

Observations were conducted using MOHAWk (Method for Observing pHysical Activity and Wellbeing): a systematic observation tool for assessing three levels of physical activity (Sedentary, Walking, Vigorous) and two other wellbeing behaviours (Take Notice: taking notice of the environment; Connect: social interactions) in urban spaces [25]. To date, there are over 650 hours of MOHAWk data across 36 unique sites in Greater Manchester, Belfast and Valencia. There is now evidence of high inter-rater reliability between pairs of observers when using MOHAWk and criterion-related validity [25]. MOHAWk has been used in two separate natural experimental studies of urban green space interventions [47,48].

Two observers were used at each time point, with each observer independently collecting data at one site. As there were three study sites but only two observers, observations were counterbalanced to control for week, day of the week, and time of day. The same observation schedule was used at all time points; this observation schedule provides a reliable estimation of activity [25,49]. Observations were carried out regardless of weather conditions; observation periods were coded as high precipitation if the observer recorded an accumulated duration of any precipitation lasting for 50% or more of the observation period (i.e. 30 minutes or more), as recommended by MOHAWk procedures.

The target area at the intervention site was the path identified for improvements ('intervention path'). Observers recorded anyone who entered onto the intervention path, regardless of which direction they were coming from. Target areas at the two comparison sites were defined to resemble the target area at the intervention site, by capturing people who entered the unsurfaced canal side path regardless of which direction they were coming from.

At each time point, JB (first observer) trained the second observer using the MOHAWk instruction manual and by practicing observations in the study sites. At the end of training, inter-rater reliability was assessed between the pair of observers for two hours using 5-minute observation periods (i.e. 24 data points). Inter-rater reliability for counts of people, their characteristics and behaviours was ‘good’ or ‘excellent’, analysed using two-way mixed, single measure, consistency intraclass correlation coefficients (ICCs) (< 0.5 = poor; 0.5 – 0.75 = moderate; 0.76 – 0.9 = good; and > 0.9 = excellent [50]).

6.3.6 Logic model

Hypothesised causal pathways were outlined in the logic model (see study protocol [31]), based on a framework suggested by Panter et al. [51]. The model predicted that the intervention would improve accessibility, convenience and pleasantness of the canal route and key destinations, which would in turn encourage more people to use the canal. This, in turn, would lead to observable increases in physical activity, social interactions, and people taking notice of the environment.

6.3.7 Sample size

We used the method suggested by Donner and Klar ([52], p.66) for calculating the sample size for a matched pair design: calculate the number of clusters (sites) required for a completely randomised cluster design, and then multiply that by one minus the correlation between the mean outcomes in the two groups. This was conducted for the primary outcome measure: the total number of people per hour (observation period). Due to the absence of studies of urban canal interventions, the sample size calculation was informed by MOHAWk data from a feasibility study of two UK residential streets [35]. Based on these data, we assumed a mean number of people per hour in the comparison group of 20, a standard deviation in both groups of 27, and correlation between intervention and comparison sites of 0.37. We assumed an intraclass correlation coefficient (ICC) of 0.02 (this relatively low ICC value has been used in previous studies that have evaluated homogeneous parks [53]). We calculated that matching one intervention site (intervention group) and two comparison sites (comparison group), with twelve observation periods per site, provided 80% power ($p = .05$, two-sided test) to detect an increase in the mean number of people from 20 counts per hour in the comparison group to 80 counts per hour in the intervention group. The calculation was conducted using the *clsampsi* command in Stata version 14.1.20.

6.3.8 Analyses

The unit of analysis was at the level of the observation period i.e. counts per observation period per site. Analyses were conducted separately for data at 7, 12 and 24 month follow-ups, in order to produce separate effect estimates for each follow-up time point. We originally planned to analyse adults and older adults separately but combined them to increase statistical power. Analyses were conducted using Stata version 14.1.20.

6.3.8.1 Primary outcome – total number of people using the canal

We estimated the effect of the intervention on the total number of people counted per observation period, compared to the comparison group, controlling for three covariates: day (Tuesday, Wednesday, Saturday), time of day (8-9am, 10-11am, 1-2pm, 3-4pm) and precipitation (Yes or No). Precipitation was a binary variable, which was coded as 'Yes' if the observer recorded an accumulated duration of any precipitation lasting for 50% or more of the observation period (i.e. 30 minutes or more).

Multilevel mixed-effects negative binomial regression models were used because they can account for overdispersion [54]. We specified a random intercept at the site level to take into account any unobserved differences between sites. We used robust standard errors to obtain unbiased standard errors. Results were reported as incidence rate ratios (IRRs), which represent the relative change in the total number of people at the intervention group compared to the comparison group.

We followed three steps to build a suitable negative binomial regression model. First, using baseline data only (without group allocation), we built a regression model to examine the relationship between the baseline data and covariates. Second, we combined the baseline and follow-up data and applied the same regression model from step one. Third, we added the Intervention Group (intervention or comparison) and Time Point (baseline or follow-up) into the model. The intervention effect was the interaction of the Intervention Group and Time Point. This is a difference-in-differences approach i.e. analysis of the change in outcomes between intervention and comparison groups pre- and post-intervention, using a regression with terms for group, period, and the group-by-period interaction [55]. Model estimates were obtained using the *menbreg* Stata command.

6.3.8.2 Secondary and exploratory outcomes – physical activity and other wellbeing behaviours

For some of the physical activity levels (Sedentary, Walking, Vigorous) and other wellbeing behaviours (Connect, Take Notice), the number of observations was too low to undertake statistical modelling as originally planned. Instead, we calculated the change in counts of each behaviour from baseline to follow-up for each of the intervention and comparison groups. Mann-Whitney U tests were then used to assess whether any differences in the change in counts between intervention and comparison groups were statistically significant, at a priori alpha level of 0.05. There were no adjustments for multiple comparisons.

6.3.8.3 Sensitivity analyses

To assess for any potential bias due to precipitation when analysing the physical activity and other wellbeing behaviour outcomes, observation periods were removed if precipitation occurred for 50% or more of the observation period (recorded by the observer).

6.3.9 Process evaluation

6.3.9.1 Systematic observation

At the intervention site, observers separately recorded people who accessed the existing unchanged path on the other side of the canal (Figure 6.2). Specifically, these were people who were following the same route that leads to the intervention path, but did not use the intervention path and instead travelled over the bridge to use the unchanged path on the other side of the canal. This allowed us to assess the extent to which the intervention affected use of the unchanged path.

6.3.9.2 Intercept surveys

Short face-to-face intercept surveys were conducted to assess change in use of the intervention path and potential displacement of activity from the unchanged canal path at the intervention site. Surveys were conducted at each of the follow-ups (June/ July 2018; Nov/ Dec 2018; Nov/ Dec 2019) using the same schedule as observations but on different days. Surveys were attempted with all English-speaking adult canal users (aged 18+ checked in the introduction). Informed consent was obtained verbally and surveys were completely anonymous. The estimated gender, age group and ethnicity of each participant were recorded using MOHAWk coding procedures. Appendix O contains the intercept survey questions.

6.4 Results

Table 6.2 provides information on baseline observation periods, sample and outcomes. The intervention and comparison group samples were similar at baseline in terms of gender and ethnicity. Although the intervention site had higher numbers of people at baseline, proportions of wellbeing behaviours were similar between intervention and comparison sites.

Table 6.2. Baseline information on observation periods, sample and outcomes

	Comparison group (2 sites)	**Intervention group (1 site)**
Observation periods		
Total number of observation periods	24	12
8am – n (% of total)	6 (25%)	3 (25%)
10am – n (% of total)	6 (25%)	3 (25%)
1pm – n (% of total)	6 (25%)	3 (25%)
3pm – n (% of total)	6 (25%)	3 (25%)
Tuesday – n (% of total)	8 (33.33%)	4 (33.33%)
Wednesday – n (% of total)	8 (33.33%)	4 (33.33%)

Saturday – n (% of total)	8 (33.33%)	4 (33.33%)
High precipitation – n (% of total)	5 (20.83%)	0 (0%)
Sample		
Female – n (% of total)	31 (47.69%)	30 (58.82%)
White – n (% of total)	65 (100%)	50 (98.04%)
Primary outcome		
Overall count of people – n (median, IQR)	65 (2, 1.75)	51 (4, 2.5)
Secondary outcomes		
Walking behaviours – n (median, IQR)	64 (2, 1.75)	48 (4, 1.75)
Vigorous behaviours – n (median, IQR)	1 (0, 0)	3 (0, 0.75)
Exploratory outcomes		
Connect behaviours – n (median, IQR)	24 (0.5, 2)	17 (1.5, 2)
Take Notice behaviours – n (median, IQR)	5 (0, 0)	3 (0, 0.75)
Sedentary behaviours – n (median, IQR)	4 (0, 0)	2 (0, 0)

All values to two decimal places; *IQR* interquartile range

6.4.1 Total number of people using the canal

Table 6.3 shows the results of the mixed-effects negative binomial regression models. Compared to the comparison group, the total number of people observed using the intervention canal path increased at all three follow-ups, controlling for day, time of day and precipitation: 7 months (IRR 1.67, 95% CI 1.44 – 1.95), 12 months (primary outcome) (IRR 2.10, 95% CI 1.79 – 2.48), and 24 months post-baseline (IRR 2.42, 95% CI 1.80 – 3.24). An IRR greater than 1 indicates an increase in the total number of people in the intervention group relative to the comparison group, whereas an IRR less than 1 indicates a decrease. For example, for the primary outcome, the IRR (2.10) at 12 months post-baseline suggests that the intervention is associated with a 110% increase in the total number of people compared to the comparison group.

6.4.2 Physical activity and other wellbeing behaviours

Table 6.4 shows the results for all physical activity levels and other wellbeing behaviours. Compared to the comparison group, Walking behaviour increased in the intervention site at all three follow-ups; Vigorous activity increased at 7 months (p = .009) and 24 months (p = .002), but not at 12 months (p = .96) post-baseline (Table 6.4); Connect behaviour increased at 12 months (p = .03) and 24 months (p = .006), but not at 7 months (p = .42) post-baseline (Table 6.4); and Take Notice behaviour increased at 24 months (p = .001), but not at 7 months (p = .07) or 12 months (p = .58) post-baseline (Table 6.4).

6.4.3 Sensitivity analyses

Removing observation periods with high precipitation did not change the direction or statistical significance of the results for the secondary and exploratory outcomes (see Appendix P).

Table 6.3. Mixed-effects negative binomial regression results for the total number of people. *Models adjusted for day, time of day and precipitation.*

Time point	Total number of people (median per observation period)		Model results at each follow-up					
	Intervention group	Comparison group	Total number of observation periods in model	IRR	95% CI	Robust standard error	*p*-value	Random effect variance (site)
Baseline	51 (4)	65 (2)	-	-	-	-	-	-
7 months	148 (13.5)	117 (5)	72	1.67	1.44 – 1.95	0.13	< 0.001*	5.16e-34
12 months [a]	142 (11.5)	88 (3)	72	2.10	1.79 – 2.48	0.17	< 0.001*	8.75e-34
24 months	154 (13.5)	99 (4)	70 [b]	2.42	1.80 – 3.24	0.36	< 0.001*	5.76e-33

[a] Primary outcome; [b] Two observation periods were cancelled at the intervention site due to ethical concerns regarding observer safety;

* Statistically significant at $p < 0.05$ (z-test, two-tailed);

CI confidence interval; *IRR* incidence rate ratio

Table 6.4. Mann-Whitney U test results for all physical activity levels and other wellbeing behaviours.

Outcome		**Intervention group**				**Comparison group**				**Effect (difference between the change in the two groups)**	***p*-value** [a]
		Number of observation periods, *n*	**Total count**	**Median (IQR)**	**Change in median (from baseline)**	**Number of observation periods, *n***	**Total count**	**Median (IQR)**	**Change in median (from baseline)**		
Walking (secondary)	Baseline	12	48	4 (1.75)	-	24	64	2 (1.75)	-	-	-
	7 months	12	135	12.5 (8.5)	8.5	24	116	5 (3)	3	5.5	.005*
	12 months	12	140	11.5 (10)	7.5	24	88	3 (3)	1	6.5	< 0.001*
	24 months	10 [b]	141	12 (9.5)	8	24	98	4 (3.75)	2	6	< 0.001*
Vigorous (secondary)	Baseline	12	3	0 (0.75)	-	24	1	0 (0)	-	-	-
	7 months	12	13	0.5 (2)	0.5	24	1	0 (0)	0	0.5	.009*
	12 months	12	7	0 (0.75)	0	24	1	0 (0)	0	0	.96
	24 months	10 [b]	13	1 (2.25)	1	24	0	0 (0)	0	1	.002*
Sedentary (exploratory)	Baseline	12	2	0 (0)	-	24	4	0 (0)	-	-	-
	7 months	12	24	1.5 (4)	1.5	24	17	0 (1)	0	1.5	.07
	12 months	12	31	2 (1)	2	24	9	0 (0.75)	0	2	< 0.001*
	24 months	10 [b]	24	2.5 (2.25)	2.5	24	8	0 (0)	0	2.5	< 0.001*

Connect (exploratory)	Baseline	12	17	1.5 (2)	-	24	24	0.5 (2)	-	-	-
	7 months	12	49	4 (5.25)	2.5	24	41	2 (2)	1.5	1	.42
	12 months	12	40	3 (5.75)	1.5	24	21	0 (2)	-0.5	2	.03*
	24 months	10 [b]	49	3.5 (5.25)	2	24	25	0 (2)	-0.5	2.5	.006*
Take Notice (exploratory)	Baseline	12	3	0 (0.75)	-	24	5	0 (0)	-	-	-
	7 months	12	11	1 (1.75)	1	24	6	0 (0.75)	0	1	.07
	12 months	12	8	0 (1)	0	24	7	0 (0)	0	0	.58
	24 months	10 [b]	15	1.5 (1)	1.5	24	3	0 (0)	0	1.5	.001*

[a] Mann-Whitney U tests were carried out to determine if there were significant differences in the change in counts of behaviours from baseline to each follow-up between intervention and comparison groups; [b] Two observation periods were cancelled at the intervention site due to ethical concerns regarding observer safety;
* Statistically significant at $p < 0.05$ (z-test, two-tailed);
IQR interquartile range

6.4.4 Process evaluation

6.4.4.1 Systematic observation

To assess whether displacement could account for the main findings, we assessed the impact of the intervention on use of the unchanged canal path. As shown in Table 6.5, the total number of people accessing the unchanged canal path decreased at 12 and 24 months post-baseline. This indicates there was some displacement to the intervention path from existing routes along the unchanged side of the canal. However, there was an overall increase in the combined total number of people using the canal at the intervention site (Table 6.5), which suggests that displacement from this particular route cannot explain all of the observed increases in use of the intervention path.

Table 6.5. Total number of people using the intervention path and unchanged path at the intervention site.

Time point	Total number of people (median per observation period)		
	Intervention path	**Unchanged path**	**Combined total**
Baseline	51 (4)	103 (7)	154 (10.5)
7 months	148 (13.5)	143 (7.5)	291 (22.5)
12 months	142 (11.5)	88 (7)	230 (18.5)
24 months	154 (13.5)	63 (5)	217 (18)

6.4.4.2 Intercept surveys

A total of 53 participants completed intercept surveys, 58.9% of those approached: 21 in June/ July 2018 (7-8 months post-baseline); 15 in Nov/ Dec 2018 (12-13 months post-baseline); 17 in Nov/ Dec 2019 (24-25 months post-baseline). There were 30 females (56.6%) and 52 white participants (98.1%). The observed primary activity of these participants was walking (47.2%), dog walking (41.5%), running (7.5%), and cycling

(3.8%). Most participants reported that they walked for at least 10 minutes at a time on several days a week (41.5%) or every day (39.6%) in the past month.

Twenty participants (37.7%) were already using the intervention path before the intervention was implemented. Thirty participants (56.6%) only started using the intervention path post-intervention: most participants (n = 23) were displaced from other existing routes along the canal, but some participants (n = 6) were new canal users; five of these six participants had moved house after the intervention.

Some participants reported that the intervention encouraged them to spend more time using the canal per visit (n = 10, 18.9%) or visit the canal more often per week (n = 3, 5.7%). Results of the intercept surveys are provided in Appendix O.

6.5 Discussion

6.5.1 Key findings

The new walking infrastructure and green space improvements were associated with a statistically significant increase in the number of people using the intervention canal path compared to the comparison group at 7 months (IRR 1.67), 12 months (IRR 2.10) and 24 months (IRR 2.42) post-baseline. These increases were observed immediately at just one-month post-intervention (7 months post-baseline), with the largest effects observed at 18 months post-intervention (24 months post-baseline), suggesting that use of the canal path continued to increase over time. Examination of the IRR values suggests that these increases were sizable, especially at 12 and 24 months, where there were over twice as many people from baseline to follow-up using the intervention canal path relative to the comparison group i.e. IRR values greater than 2. The process evaluation suggested that there was displacement of activity from the existing unchanged canal path to the intervention path, but displacement from this particular canal route cannot explain all of the observed increases in use of the intervention path. Some survey participants reported that the intervention encouraged them to spend more time using the canal during each visit or visit the canal more often across the week.

There were some observed increases in walking and vigorous physical activity, social interactions, and people taking notice of the environment at the intervention canal path compared to the comparison group. The largest effects were observed on walking behaviour at all follow-ups, which is unsurprising given that the majority of people using the canal were walking rather than engaging in vigorous physical activity such as running or cycling. There were also notable increases in social interactions at 12 months and 24 months at the intervention path, with over twice as many social interactions compared to baseline (albeit from low baseline counts), whereas there were no observed changes in social interactions in the comparison group.

6.5.2 Interpretation

There are two key reasons why improving access to this particular part of the canal may have caused increased use of this route. First, the resurfaced footpath and removal of encroaching vegetation provided better and safer access to 'wild' and interesting green corridors along the canal, thus providing a more pleasant walking and visual experience compared to the existing canal path. This accords with previous research which suggests that physical activity in natural environments is generally more enjoyable [56], and can encourage people to be active for longer and at higher intensities [57]. It is therefore plausible that the intervention may act as a catalyst to encourage and sustain physical activity.

Second, improving access to this side of the canal created a more attractive separate route for canal users (especially for people walking), which may have enabled better segregation of canal users engaging in different types of physical activity (e.g. walking, jogging, cycling), thus reducing shared space conflict. For example, some of the intercept survey participants who were walking reported that they started using the improved intervention path to avoid collisions with cyclists on the existing well-surfaced towpath on the other side of the canal. Previous research suggests that segregating pedestrians and cyclists on shared use routes is an effective way to reduce collisions and conflict [58], which is very relevant to canals paths as they are generally quite narrow but are used by a variety of users. It is also possible that reduction in shared space conflict between canal users may have contributed to the observed increases in social interactions at the intervention path 12 months and 24 months post-baseline, thus further enhancing the physical activity experience due to the increased enjoyment and wellbeing benefits associated with more social contact [59,60]. In light of the current COVID-19 pandemic, the pertinence of reducing shared space conflict is heightened because of social distancing rules to reduce virus transmission [61].

Whilst this is the first natural experimental study of an urban canal intervention on physical activity to date, these findings are consistent with a similar recent controlled natural experimental study in the UK [62]. They found that a comparable intervention to improve access to urban woodlands (e.g. installing new footpaths, clearing overgrown vegetation) caused people to visit the woods more often, and increased physical activity levels, social cohesion and connectedness with nature. They also found that the intervention enhanced people's experience of the woodlands, as shown by improvements in perceived restorativeness measures (e.g. spending time away from a day-to-day routine, fascination). These findings, together with the present study, suggest that interventions which improve access to existing urban green spaces, especially spaces that enable close physical and visual contact with natural features (e.g. trees, shrubs, water), are effective in causing positive behaviour change and improving people's experience of urban green spaces. Importantly, the intervention providers in the present study included funding for 20 years of ongoing maintenance, which will increase the likelihood of providing long-term benefits and reduce negative effects on wellbeing associated with unmanaged green spaces (e.g. fear of crime) [63].

6.5.3 Strengths and limitations

The practical and methodological challenges associated with evaluating natural experiments are well-documented [64–66]. This natural experimental study offers a feasible exemplar of how to address major methodological weaknesses causing high risk of bias in previous studies in this area. Key strengths of the present study include identifying multiple matched comparison sites using important objective variables (e.g. population density); using methods recommended by MRC guidance [18] to minimise the risk of confounding; conducting a sample size calculation; publication of a study protocol with a priori analyses specified and reporting any key deviations; clear reporting of samples and interventions in line with standardised checklists; and a mixed-methods approach. These are substantial improvements on many previous natural experimental studies that have often used poorly matched single comparison sites (rarely matched on objective variables); not controlled for key confounding factors by study design or in the analyses; not attempted sample size calculations; not published study protocols; often relied on single outcome measures (sometimes unvalidated); and are poorly reported [16,17,19].

We discovered the opportunity to evaluate the natural experiment just six weeks prior to implementation of the intervention. This demonstrates the feasibility of applying these robust methods even with short time frames to collect baseline data, which is crucial in the absence of suitable routinely collected data. Similar methods have been applied in two other natural experimental studies [47,48].

Identifying closely matched comparison sites for environmental interventions is difficult, especially as these interventions tend to be implemented in a unique local context.

Although we had to relax the matching criteria slightly (Appendix K), our newly developed process of comparison site matching provided well-matched sites (Table 6.1) and the samples were similar at baseline (Table 6.2). This matching process has been used in two other recent natural experimental studies [47,48], therefore suggesting that this is a feasible method for future studies using matched intervention and comparison groups.

A process evaluation enabled us to strengthen causal inferences from the observations, including possible displacement effects; a key limitation of previous natural experimental studies in this area is not assessing for possible displacement effects [16]. However, we were unable to precisely estimate the effect of the intervention on net population-levels of physical activity, as we did not include measures that can assess within-person change over time (e.g. surveys, accelerometers). Nonetheless, within-person measures do suffer from certain methodological challenges (e.g. low response rates [19]), which is why triangulation of different outcome measures is optimal to strengthen causal inferences.

Statistical power was an issue for some of the wellbeing behaviours due to low counts (especially Vigorous and Take Notice behaviours), which prevented us from fitting appropriate regression models that could adjust for the total number of people and account for clustering. It is important to acknowledge that the Mann-Whitney U tests for each of the wellbeing behaviours could not account for clustering and therefore erroneously assumed that the outcomes were independent, which may have increased the risk of Type I errors ('false positives') due to artificially narrow confidence intervals [67]. Nevertheless, we powered the study for the primary outcome (i.e. total number of people), for which there was sufficient data to use regression modelling. MOHAWk data from this study will be valuable in informing sample size calculations and analytical plans in future studies of environmental interventions.

6.5.4 Implications for policy and practice

We have provided robust practice-based evidence that urban canals offer promising settings for interventions to encourage green space usage and potentially enhance population health and wellbeing. This evidence will be particularly useful, given the increasing interest in evidence-informed urban design policies and practices [68,69]. This study suggests that even relatively small-scale interventions which improve access to green corridors along canals and provide separate routes for different types of physical activities can have measurable impacts on the use of urban green spaces. However, current national UK policy and industry guidelines for planning and designing the physical environment to improve health and wellbeing do not emphasise the potential role of urban canals (e.g. [70–72]). This is an important gap as over 8 million people in England and Wales (14% of the population) live within 1 km of a canal or river [26]. There are also larger proportions of the population living near a canal or river in areas with high levels of deprivation in the UK, such as Greater Manchester (25% of the population) and West Midlands (51% of the population) [26]. Hence, urban canals have the potential to produce wide-reaching effects and address health inequalities.

Urban canals are particularly promising settings for urban green space interventions for a number of reasons. Canal paths offer pleasant recreational settings and also traffic-free utilitarian routes that connect suburban areas with city or town centres, as well as links between urban centres. They can facilitate a range of physical activities, including walking, jogging, cycling, and less physically exertive but increasingly popular 'e-bikes' and 'e-scooters' [73]. Canal paths tend to be predominantly flat, making them well suited to a multitude of users with varying physical abilities. Canals provide free access to green and blue space in dense built-up urban areas where proximity to local green and blue spaces may be limited, especially for populations that cannot access larger green and blue spaces such as national parks, coastlines and beaches. Crucially, as demonstrated in the present study, enhancement of a canal can increase use and only requires relatively small-scale change and maintenance that can be readily replicated at low cost.

6.5.5 Implications for research

The potential for urban canals to influence population health and wellbeing is understudied and the present research has suggested directions for future intervention research. Future studies should explore how urban canal interventions could reduce shared space conflict between different types of canal users. As well as finding ways to improve physical activity experiences of existing canal users, it is also important to examine the most effective interventions to encourage more new canal users from local communities, especially people from Black and Minority Ethnic (BAME) groups who are underrepresented users of canals [26], despite canals being more equally geographically distributed than parks and other urban green spaces [75]. A previous systematic review found that combining a physical change with a specific awareness/ promotion program (e.g. community events, outreach activities) is the most effective way to increase urban green space use and physical activity of users [16]. However, research needs to explore the most effective interventions that address specific barriers preventing use of urban canals (e.g. perceptions of safety [26]) and potential unintended consequences of such activities (e.g. exacerbating inequalities among BAME groups).

While the intervention in the present study seemed to be effective, more well-developed theory is needed to improve our understanding of whether findings can be generalised to different settings. In particular, it is important to understand how physical and social contexts impact intervention effectiveness, such as topography, deprivation and population demographics. A recent systematic review has begun to develop context-sensitive theory of how environmental interventions work to influence physical activity using realist evaluation methods [76]. This review found that environmental interventions that improve access to places or facilities (e.g. by installing new or improved walking routes) may provide an effective approach for increasing physical activity that is less sensitive to contextual factors. The results of this review therefore suggest that the intervention in the present study that improved access to the canal (e.g. by improving walking infrastructure) may also be effective in other contexts; for example, the intervention was implemented in a relatively affluent area, but the intervention may also be effective in more deprived areas. However, theoretical models need to incorporate other wellbeing behaviours beyond physical activity, especially social interactions given the role that these behaviours can have in supporting physical activity [77]. This mixed methods study provides robust evidence of intervention effectiveness required to advance theoretical understanding.

6.6 Conclusions

Using rigorous natural experimental methods, this mixed-methods study shows that urban canals are promising settings for replicable small-scale interventions to increase green space usage and potentially increase physical activity and other wellbeing behaviours. This study has suggested avenues for future intervention research to further examine which urban canal interventions can effectively influence population health and wellbeing; including interventions that reduce shared space conflict, improve physical activity experiences for existing canal users and attract new canal users (particularly underrepresented populations). This study provides an exemplar to improve the internal validity of future natural experimental studies of environmental interventions, by addressing major methodological weaknesses causing high risk of bias in many previous natural experimental studies of environmental interventions.

6.7 References

1. World Health Organization. Urban Green Space Interventions and Health: A review of impacts and effectiveness. Copenhagen; 2017. Available from: http://www.euro.who.int/__data/assets/pdf_file/0010/337690/FULL-REPORT-for-LLP.pdf?ua=1

2. Lee ACK, Maheswaran R. The health benefits of urban green spaces: a review of the evidence. J Public Health. 2010;33(2):212–22.

3. Houlden V, Weich S, Porto de Albuquerque J, Jarvis S, Rees K. The relationship between greenspace and the mental wellbeing of adults: A systematic review. PLoS One. 2018;13(9):e0203000.

4. Bell SL, Phoenix C, Lovell R, Wheeler BW. Green space, health and wellbeing: making space for individual agency. Health Place. 2014;30:287–92.

5. United Nations. World Urbanization Prospects: The 2014 Revision, Highlights. New York; 2015. Available from: https://esa.un.org/unpd/wup/publications/files/wup2014-report.pdf

6. World Health Organization. Urban green spaces and health: A review of evidence. Copenhagen: World Health Organization; 2016. Available from: http://www.euro.who.int/__data/assets/pdf_file/0005/321971/Urban-green-spaces-and-health-review-evidence.pdf?ua=1

7. Markevych I, Schoierer J, Hartig T, Chudnovsky A, Hystad P, Dzhambov AM, et al. Exploring pathways linking greenspace to health: Theoretical and methodological guidance. Environ Res. 2017;158:301–17.

8. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ. 2006;174(6):801–9.

9. Penedo FJ, Dahn JR. Exercise and well-being: a review of mental and physical health benefits associated with physical activity. Curr Opin Psychiatry. 2005;18(2):189–93.

10. Pretty JN, Barton J. The mental and physical health outcomes of green exercise. Int J Environ Health Res. 2005;15(5):319–37.

11. Maas J, van Dillen SME, Verheij RA, Groenewegen PP. Social contacts as a possible mechanism behind the relation between green space and health. Health Place. 2009;15(2):586–95.

12. Helliwell JF, Putnam RD. The Social Context of Well-being. In: Huppert FA, Baylis N, Keverne B, editors. The Science of Well-being. Oxford: Oxford University Press; 2005. p. 435–59.

13. Anderson J, Ruggeri K, Steemers K, Huppert F. Lively Social Space, Well-Being Activity, and Urban Design: Findings From a Low-Cost Community-Led Public Space Intervention. Environ Behav. 2017;49(6):685–716.

14. McEwan K, Richardson M, Sheffield D, Ferguson FJ, Brindley P. A Smartphone App for Improving Mental Health through Connecting with Urban Nature. Int J Environ Res Public Health. 2019;16(18):3373.

15. Seligman MEP, Steen TA, Peterson C. Positive Psychology Progress: Empirical Validation of Interventions. Am Psychol. 2005;60(5):410–21.

16. Hunter RF, Christian H, Veitch J, Astell-Burt T, Hipp JA, Schipperijn J. The impact of interventions to promote physical activity in urban green space: A systematic review and recommendations for future research. Soc Sci Med. 2015;124:246–56.

17. World Health Organization. Urban Green Space Interventions and Health: A

review of evidence. Copenhagen; 2017. Available from: http://www.euro.who.int/__data/assets/pdf_file/0005/321971/Urban-green-spaces-and-health-review-evidence.pdf?ua=1

18. Craig P, Gunnell D, Ogilvie D, Sutton M, Thompson S. Using natural experiments to evaluate population health interventions: guidance for producers and users of evidence Prepared on behalf of the Medical Research Council by. 2012. Available from: www.mrc.ac.uk/naturalexperimentsguidance

19. Benton JS, Anderson J, Hunter RF, French DP. The effect of changing the built environment on physical activity: a quantitative review of the risk of bias in natural experiments. Int J Behav Nutr Phys Act. 2016;13(1):107.

20. Roberts H, McEachan R, Margary T, Conner M, Kellar I. Identifying Effective Behavior Change Techniques in Built Environment Interventions to Increase Use of Green Space: A Systematic Review. Environ Behav. 2018;50(1):28–55.

21. Kärmeniemi M, Lankila T, Ikäheimo T, Koivumaa-Honkanen H, Korpelainen R. The Built Environment as a Determinant of Physical Activity: A Systematic Review of Longitudinal Studies and Natural Experiments. Ann Behav Med. 2018;52(3):239–51.

22. MacMillan F, George ES, Feng X, Merom D, Bennie A, Cook A, et al. Do Natural Experiments of Changes in Neighborhood Built Environment Impact Physical Activity and Diet? A Systematic Review. Int J Environ Res Public Health. 2018;15(2):217.

23. Sallis JF, Cerin E, Conway TL, Adams MA, Frank LD, Pratt M, et al. Physical activity in relation to urban environments in 14 cities worldwide: a cross-sectional study. Lancet. 2016;387(10034):2207–17.

24. New Economics Foundation. Five ways to wellbeing: The evidence.. London; 2008. Available from: http://b.3cdn.net/nefoundation/8984c5089d5c2285ee_t4m6bhqq5.pdf

25. Benton JS, Anderson J, Pulis M, Cotterill S, Hunter RF, French DP. Method for Observing pHysical Activity and Wellbeing (MOHAWk): validation of an observation tool to assess physical activity and other wellbeing behaviours in urban spaces. Cities Health. 2020;

26. Canal & River Trust. Waterways & Wellbeing. Building the Evidence Base. First Outcomes Report. Milton Keynes; 2017. Available from: https://canalrivertrust.org.uk/refresh/media/thumbnail/33802-canal-and-river-trust-outcomes-report-waterways-and-wellbeing-full-report.pdf

27. Gascon M, Zijlema W, Vert C, White MP, Nieuwenhuijsen MJ. Outdoor blue spaces, human health and well-being: A systematic review of quantitative studies. Int J Hyg Environ Health. 2017;220(8):1207–21.

28. Canal & River Trust. Annual Report & Accounts 2018/19. Milton Keynes; 2019. Available from: https://canalrivertrust.org.uk/media/original/40488-annual-report-and-accounts-2018-2019.pdf?v=f4ac07

29. Scottish Canals. Funding boost keeps Scotland’s canals flowing into the future. 2020. Available from: https://www.scottishcanals.co.uk/news/funding-announcement/

30. Agency EE. Rivers and Lakes in European Cities—Past and Future Challenges. Copenhagen, Denmark; 2016. Available from: https://www.eea.europa.eu/publications/rivers-and-lakes-in-cities

31. Benton J, Anderson J, Cotterill S, Dennis M, French DP. The impact of new

walking infrastructure and changes to green space along an urban canal on physical activity and wellbeing: protocol for a natural experimental study. Open Sci Framew. 2018; https://osf.io/zcm7v/

32. Salford City Council. Indices of Deprivation 2015 - the City of Salford’s position. 2015. Available from: https://www.salford.gov.uk/media/388062/index_of_multiple_deprivation_report_2015.pdf

33. Salford City Council. Boothstown and Ellenbrook Ward profile - January 2020. 2020. Available from: https://www.salford.gov.uk/media/395071/b-and-e-ward-profile-jan-2020.pdf

34. Peel Land and Property Group. Improvements to green spaces and recreational routes in Boothstown Salford. 2017. Available from: https://peellandp.co.uk/boothstown/

35. Benton JS, Anderson J, Cotterill S, Dennis M, Lindley SJ, French DP. Evaluating the impact of improvements in urban green space on older adults’ physical activity and wellbeing: protocol for a natural experimental study. BMC Public Health. 2018;18(1):923.

36. Humpel N, Owen N, Leslie E. Environmental factors associated with adults’ participation in physical activity. A review. Vol. 22, American Journal of Preventive Medicine. Elsevier Inc.; 2002. p. 188–99.

37. Owen N, Humpel N, Leslie E, Bauman A, Sallis JF. Understanding environmental influences on walking: Review and research agenda. Vol. 27, American Journal of Preventive Medicine. Elsevier; 2004. p. 67–76.

38. Duncan MJ, Spence JC, Mummery WK. Perceived environment and physical

activity: A meta-analysis of selected environmental characteristics. Vol. 2, International Journal of Behavioral Nutrition and Physical Activity. BioMed Central; 2005. p. 1–9.

39. Wendel-Vos W, Droomers M, Kremers S, Brug J, Van Lenthe F. Potential environmental determinants of physical activity in adults: A systematic review. Obes Rev. 2007;8(5):425–40.

40. Saelens BE, Handy SL. Built Environment Correlates of Walking: A Review. Med Sci Sports Exerc. 2008;40(7 Suppl):S550-66.

41. Panter JR, Jones A. Attitudes and the environment as determinants of active travel in adults: What do and don't we know? J Phys Act Heal. 2010;7(4):551–61.

42. Moran M, Van Cauwenberg J, Hercky-Linnewiel R, Cerin E, Deforche B, Plaut P. Understanding the relationships between the physical environment and physical activity in older adults: a systematic review of qualitative studies. Int J Behav Nutr Phys Act. 2014;11(1):79.

43. Barnett DW, Barnett A, Nathan A, Van Cauwenberg J, Cerin E. Built environmental correlates of older adults' total physical activity and walking: a systematic review and meta-analysis. Int J Behav Nutr Phys Act. 2017;14(1):103.

44. Cerin E, Nathan A, van Cauwenberg J, Barnett DW, Barnett A. The neighbourhood physical environment and active travel in older adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act. 2017;14(1):15.

45. Cain KL, Gavand KA, Conway TL, Geremia CM, Millstein RA, Frank LD, et al. Developing and validating an abbreviated version of the Microscale Audit for Pedestrian Streetscapes (MAPS-Abbreviated). J Transp Heal. 2017;5:84–96.

46. Gidlow CJ, Ellis NJ, Bostock S. Development of the Neighbourhood Green Space

Tool (NGST). Landsc Urban Plan. 2012;106(4):347–58.

47. Benton JS, Cotterill S, Anderson J, Macintyre VG, Gittins M, Dennis M, et al. Impact of a low-cost urban green space intervention on wellbeing behaviours in older adults: a natural experimental study. Wellbeing, Sp Soc. Under Review.

48. Anderson J, French DP, Benton JS, Dennis M, Rothwell J. The effects of the Grow Green urban greening improvements on physical health and wellbeing behaviours in Manchester UK: A natural experimental study. Open Sci Framew. 2020; https://osf.io/zqgcn

49. Cohen DA, Setodji C, Evenson KR, Ward P, Lapham S, Hillier A, et al. How Much Observation Is Enough? Refining the Administration of SOPARC. J Phys Act Heal. 2011;8(8):1117–23.

50. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med. 2016;15(2):155–63.

51. Panter J, Guell C, Prins R, Ogilvie D. Physical activity and the environment: conceptual review and framework for intervention research. Int J Behav Nutr Phys Act. 2017;14(1):156.

52. Donner A, Klar N. Design and Analysis of Cluster Randomization Trials in Health Research. John Wiley & Sons; 2000. 178 p.

53. Cohen DA, Han B, Derose KP, Williamson S, Marsh T, McKenzie TL. Increasing Physical Activity in Parks: Results of a Randomized Controlled Intervention Trial Using Community-Based Participatory Research. Am J Prev Med. 2013;45(5):590–7.

54. Cameron AC, Trivedi PK. Regression analysis of count data. Cambridge University Press; 1998.

55. Craig P, Vittal Katikireddi S, Leyland A, Popham F. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. Annu Rev Public Heal. 2017;38:39–56.

56. Gladwell VF, Brown DK, Wood C, Sandercock GR, Barton JL. The great outdoors: How a green exercise environment can benefit all. Vol. 2, Extreme Physiology and Medicine. BioMed Central Ltd.; 2013. p. 3.

57. Lahart I, Darcy P, Gidlow C, Calogiuri G. The effects of green exercise on physical and mental wellbeing: A systematic review. Int J Environ Res Public Health. 2019;16(8).

58. Atkins. Shared Use: Operational Review. 2012. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/9181/atkins-shared-use-operational-review.pdf

59. Howick J, Kelly P, Kelly M. Establishing a causal link between social relationships and health using the Bradford Hill Guidelines. SSM - Popul Heal. 2019;8:100402.

60. Francis J, Giles-Corti B, Wood LJ, Wood L, Knuiman M. Creating Sense of Community: The role of public space The Financing of Homelessness Programs View project Creating sense of community: The role of public space. Artic J Environ Psychol. 2012.

61. De Vos J. The effect of COVID-19 and subsequent social distancing on travel behavior. Transp Res Interdiscip Perspect. 2020;5:100121.

62. Ward Thompson C, Silveirinha de Oliveira E, Tilley S, Elizalde A, Botha W, Briggs A, et al. Health impacts of environmental and social interventions designed to increase deprived communities' access to urban woodlands: a mixed-methods

study. Public Heal Res. 2019;7(2):1–172.

63. Jorgensen A, Hitchmough J, Calvert T. Woodland spaces and edges: Their impact on perception of safety and preference. Landsc Urban Plan. 2002;60(3):135–50.

64. Crane M, Bohn-Goldbaum E, Grunseit A, Bauman A. Using natural experiments to improve public health evidence: a review of context and utility for obesity prevention. Heal Res Policy Syst. 2020;18(1):48.

65. Ogilvie D, Adams J, Bauman A, Gregg EW, Panter J, Siegel KR, et al. Using natural experimental studies to guide public health action: turning the evidence-based medicine paradigm on its head. J Epidemiol Community Heal. 2019;74(2):203–8.

66. Veitch J, Salmon J, Giles-Corti B, Crawford D, Dullaghan K, Carver A, et al. Challenges in conducting natural experiments in parks—lessons from the REVAMP study. Int J Behav Nutr Phys Act. 2017;14(1):5.

67. Clarke P. When can group level clustering be ignored? Multilevel models versus single-level models with sparse data. J Epidemiol Community Health. 2008;62:752–8.

68. Le Gouais A, Foley L, Ogilvie D, Guell C. Decision-making for Active Living Infrastructure in New Communities: A Qualitative Study in England. J Public Health. 2019;105.

69. Public Health England. Improving access to greenspace A new review for 2020. London; 2020.

70. Public Health England. Spatial Planning for Health An evidence resource for planning and designing healthier places. 2017. Available from: https://www.gov.uk/government/publications/spatial-planning-for-health-

evidence-review

71. National Institute for Health and Clinical Excellence (NICE). Physical activity and the environment. 2018. Available from: www.nice.org.uk/guidance/ng90

72. Ministry of Housing C& LG. National Design Guide. Planning practice guidance for beautiful, enduring and successful places.. London; 2019. Available from: https://www.gov.uk/government/publications/national-design-guide

73. Fishman E, Cherry C. Transport Reviews E-bikes in the Mainstream: Reviewing a Decade of Research E-bikes in the Mainstream: Reviewing a Decade of Research. 2015.

74. Britton E, Kindermann G, Domegan C, Carlin C. Blue care: a systematic review of blue space interventions for health and wellbeing. Health Promot Int. 2020;35(1):50–69.

75. Tieges Z, McGregor D, Georgiou M, Smith N, Saunders J, Millar R, et al. The Impact of Regeneration and Climate Adaptations of Urban Green–Blue Assets on All-Cause Mortality: A 17-Year Longitudinal Study. Int J Environ Res Public Health. 2020;17(12):4577.

76. Panter J, Guell C, Humphreys D, Ogilvie D. Can changing the physical environment promote walking and cycling? A systematic review of what works and how. Health Place. 2019;58:102161.

77. McNeill LH, Kreuter MW, Subramanian S V. Social Environment and Physical activity: A review of concepts and evidence. Soc Sci Med. 2006;63:1011–22.

78. Department for Communities and Local Government. English Indices of Deprivation 2015 [computer file]. Downloaded from: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015. Licensed under:

https://www.nationalarchives.gov.uk/doc/open-government-licence/ version/3.

2015.

SECTION 5. General Discussion

Section overview

This section includes the final chapter (Chapter 7), which discusses the extent to which this PhD successfully addressed my eight recommendations to improve natural experimental studies. Suggestions for future research are then made, followed by a discussion of the implications for research, policy and practice.

Chapter 7. Discussion

7.1 Introduction

There is a dearth of robust natural experimental studies examining the causal effects of urban green space interventions on physical activity and other wellbeing behaviours [1–6]. Prior to this PhD, I made eight recommendations to address methodological weaknesses causing high risk of bias in natural experimental studies of environmental interventions [6]. The central achievement of this PhD has been to implement seven of these recommendations in two natural experimental studies of urban green space interventions. Key methodological improvements developed and implemented in this PhD include a new process for systematically identifying matched comparison sites; appropriate adjustment for confounders to minimise the risk of confounding; publication of study protocols with a priori analyses specified and reporting any deviations; the development and validation of an observation tool for assessing a range of wellbeing behaviours; using mixed-methods evaluations that include both outcome and process measures; conducting sample size calculations; and clearly reporting the studies in line with standardised checklists. These natural experimental studies provide exemplars of how to use methods with substantially lower risk of bias than previous research.

7.2 Contributions of this PhD to implementing the eight recommendations to address methodological weaknesses

This section will discuss how this PhD successfully addressed my eight recommendations and some of the challenges encountered when addressing these recommendations.

7.2.1 Recommendation 1: Better matching of comparison sites

The commonly used methods for identifying matched comparison sites when studying the effects of environmental interventions on physical activity are typically ad hoc and often involve an element of subjectivity [6]. To address this issue in this PhD, we developed a new systematic five-step process for identifying matched comparison sites, based on several pre-defined objective and subjective variables at the neighbourhood (e.g. population density) and site level (e.g. street lighting). As shown in Chapters 4 and 6, this matching process can produce multiple closely matched sites at baseline. This matching process has recently been used in a further natural experimental study outside this PhD [7], suggesting that it is a robust and replicable method for future studies that use matched comparison sites.

This new method of matching has a number of important strengths. First, we prioritised variables that are the most consistent correlates of physical activity, based on several systematic reviews [8]. Ensuring that comparison sites are closely matched on the most important known confounders reduces the risk of baseline differences between intervention and comparison sites that may bias the effect of the intervention [9]. Second, we used objective neighbourhood-level data that have rarely been used in previous studies, including population density, street connectivity and area-level deprivation [8]. Using these objective data reduces subjectivity in identifying sites. It also ensures that researchers can replicate this method in future studies, especially as these spatial data are likely to be included in routinely collected datasets that are freely available in many countries. Finally, steps within this process can be adapted; for example, we relaxed certain criteria in one of the natural experimental studies to increase the number of potential matched sites within our geographical boundary (Chapter 6).

7.2.2 Recommendation 2: Use multiple comparison sites

We included multiple comparison sites in each natural experimental study: eight comparison sites in the first study (Chapter 4) and two comparison sites in the second study (Chapter 6). Using multiple comparison sites provides increased confidence that any variation in confounding variables across comparison sites is offset, thereby increasing confidence that any observed intervention effects are attributable to the intervention rather than other site-specific variables from a single comparison site [4]. While including multiple comparison sites required considerably more effort and resources when coordinating and conducting data collection, using multiple comparison sites was vital in this PhD to ensure that any unplanned changes or problems with individual comparison sites did not significantly reduce the validity of the study. The importance of using multiple comparison sites was illustrated in the first natural experimental study, as one of the comparison sites had to be removed from the study due to safety concerns for the observer (Chapter 4).

7.2.3 Recommendation 3: Control for confounders in statistical analyses

Even with highly rigorous matching of multiple comparison sites, it is unlikely that intervention and comparison sites will be closely matched on all important baseline characteristics, as there will always be some confounders that were unknown or unmeasured and therefore not included in the matching process. As a result, it is important to use statistical analyses that can deal with these unobserved confounders.

As recommended by Medical Research Council (MRC) guidance for natural experiments [9], we used a difference-in-differences approach. This approach estimates the intervention effect by comparing the pre- and post-intervention change in outcomes between the intervention and comparison group (see Figure 1) [10]. The strength of this relatively simple differencing procedure is that it controls for the unobserved differences in the fixed (i.e. time-invariant) characteristics of the groups that cannot be balanced when matching intervention and comparison sites, which reduces bias due to unmeasured or inaccurate measurement of confounders [11]. Another advantage of this approach is that it only requires observations from an exposed intervention group and an unexposed comparison group at one time-point before and after the intervention [12], which suited this PhD given that there was limited time pre-intervention to collect multiple waves of data. Alternative approaches for evaluating natural experiments, such as interrupted time series, often require multiple pre-intervention data points to establish pre-intervention trends [13].

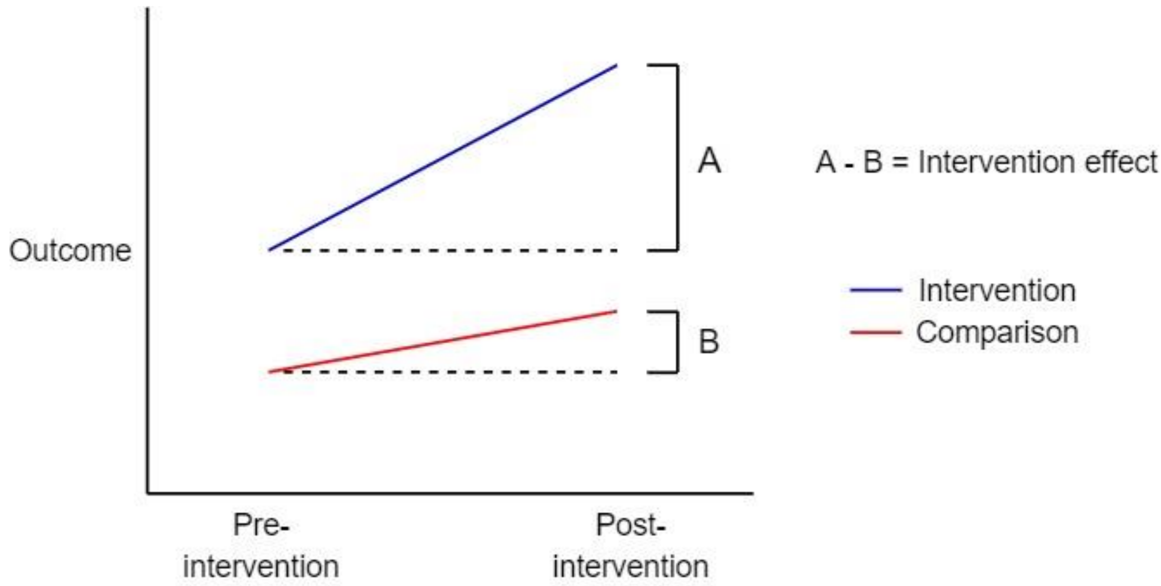


Figure 7.1. Simplified graphical explanation of the difference-in-differences approach

The difference-in-differences approach relies on a key assumption that if the intervention had not occurred, the average change in outcomes for the intervention and comparison groups would have been the same: this is known as the ‘parallel trends’ assumption [14]. We were unable to test whether the parallel trends assumption was violated, as this would require more than one wave of baseline data to assess the stability of pre-intervention trends [15]. However, it is unlikely that trends in use of outdoor spaces, physical activity and the other wellbeing behaviours would have differed between the intervention and comparison sites in our studies given that we used multiple closely matched comparison sites located within the same geographical region (Greater Manchester), over a relatively short period of time (< 2 years). Therefore, it is likely that the parallel trends assumption holds.

Our data had a hierarchical structure, in which observation periods were grouped (or 'clustered') within larger units (i.e. sites), with each site grouped according to exposure to the intervention (i.e. intervention or comparison). We used multilevel regression modelling because it accounts for this clustering in hierarchically structured data by incorporating cluster-specific random effects [16]. In contrast, traditional single-level analysis techniques often used in previous natural experimental studies, such as t-tests and linear regression models, ignore this clustering. When analysing clustered data, these single-level analyses lead to an increased risk of Type I errors ('false positives'). This is because they erroneously assume that the outcomes are independent of each other, which results in standard errors that are too small and generates artificially narrow confidence intervals [17]. The multilevel analyses used in this PhD have provided less biased effect estimates by accounting for this clustering.

7.2.4 Recommendation 4: Publish study protocols with a priori analyses specified

We submitted study protocols for publication before follow-up data collection began for each natural experimental study (Chapters 3 and 5) [18,19]. We specified all primary, secondary and exploratory outcomes and the planned statistical analyses for each outcome. Importantly, both protocols were published on open access platforms: the first protocol was registered with ClinicalTrials.gov and published in an open access journal [18], and the second protocol was published in the Open Science Framework [19].

When it came to conducting the analyses, we had to make some alterations to the initial pre-planned analyses. For example, in the first natural experimental study (Chapter 4), counts of the primary outcome (Take Notice behaviours) were too sparse to fit regression models to the data, so we used the overall count of older adults per observation period as the primary outcome instead (the original planned secondary outcome). We reported any deviations from the originally planned analyses and provided justification for the decisions we made. Without precise a priori analyses specified in a protocol, these deviations would not have been known, thus increasing the risk of selective reporting and data dredging. This demonstrates the importance of publishing planned analyses prior to follow-up data collection to improve transparency.

The alterations we made to the original plan in the protocol were due to limited use of the MOHAWk observation tool in previous studies, which made it difficult to accurately estimate the expected use of urban spaces to inform the sample size calculations. A total of 562 hours of MOHAWk data were collected during this PhD across 18 unique sites, including amenity green spaces, canal paths and urban squares. This substantial normative data will provide future researchers with more data to improve the accuracy of sample size estimations, therefore reducing the risk of having to deviate from the original planned analyses due to insufficient statistical power.

7.2.5 Recommendation 5: Use adequate outcome measurements

One of the aims of this PhD was to develop and validate a new systematic observation tool, to address a recommendation concerning unvalidated outcome measures used in previous natural experimental studies of environmental interventions [6]. We developed MOHAWk (Chapter 2) [20], which is a systematic observation tool with many strengths: there is evidence of reliability, validity and sensitivity to change; it captures a range of wellbeing behaviours beyond physical activity; and it can capture activity in urban spaces that typically have lower numbers of users (e.g. amenity green spaces) or that people pass through (e.g. residential streets), thus providing higher-powered studies that require fewer observations due to increased sensitivity.

We decided to use systematic observation because this method provides many advantages over alternative measures of physical activity, including self-report (e.g. questionnaires) and device-based measures (e.g. accelerometers). The main advantage is that observations are less obtrusive and without participant burden, which reduces issues of participant reactivity [21], low response rates [22] and survey recall bias [23].

Importantly, systematic observation does not require recruitment of participants, therefore enabling researchers to respond quickly to natural experiment opportunities by collecting data within a short period of time, which proved valuable for the studies in this PhD.

Although using systematic observation offers many advantages, observations cannot assess individual-level behaviour change and are therefore unable to provide estimates of net changes in behaviour within participants. Due to time constraints and the resources available, we were unable to include an additional outcome measure that assessed individual-level behaviour change in a cohort of participants (e.g. household surveys). Assessing within-person behaviour change would have been particularly useful in the second natural experimental study, as this would have provided a clearer idea of whether the canal intervention caused genuine net increases in green space usage and overall physical activity (Chapter 6).

A major limitation in previous natural experimental studies of environmental interventions is a lack of process evaluation [6]. MRC guidance for complex interventions recommends three process evaluation themes to provide a more detailed understanding of: (1) intervention implementation (intervention delivery and post-evaluation scale-up); (2) mechanisms of impact (how interventions trigger change); and (3) context (external factors that influence intervention delivery and effectiveness) [24]. In our studies, we conducted a process evaluation of two of these themes: intervention implementation and mechanisms of impact. To assess intervention implementation, we liaised with the intervention providers to check that the intervention was delivered in accordance with the prescribed intervention plan and whether any external events occurred that could have impacted the outcomes. To assess mechanisms of impact, we collected interview and survey data that could help explain how the intervention affected the outcomes (e.g. potential displacement effects - see Chapter 6) and thereby strengthen causal inferences. This was particularly important given that systematic observation cannot provide information on why people use, or do not use, the spaces; therefore highlighting the importance of using different methods to triangulate findings. Although we did not assess the third theme (i.e. how contextual factors impacted intervention effectiveness), the studies in this PhD are a vast improvement compared to many previous studies that have often relied on single outcome measures without any process evaluation [6].

7.2.6 Recommendation 6: Conduct sample size calculations

We used an approach suggested by Donner and Klar ([25], p.66) for calculating the sample size for a matched-pair design. This involves firstly calculating the number of clusters required for a completely randomised cluster design. Then, this value is multiplied by 1 minus the correlation between the mean outcomes in the intervention and comparison groups. This adjustment is made because matching intervention and comparison groups on baseline variables strongly related to the outcome can reduce variance and therefore increase the power of the study, so it is presumed that a matched-pair design has increased statistical power compared to a completely randomised design [26]. This approach suited our natural experimental studies because it not only accounts for clustering, but also allows for unequal numbers of clusters between intervention and comparison groups; both of our studies had more sites in the comparison group than the intervention group.

However, conducting accurate sample size calculations was challenging for two reasons. First, there is generally a lack of research dedicated toward calculating sample size for cluster randomised trials compared to individually randomised trials [27]. There is even less research dedicated to calculating sample size for cluster non-randomised trials. As a result, there were limited established methods that were suitable for our study design.

Second, there were a lack of relevant normative data to inform the calculations. Specifically, there were very few existing natural experimental studies of small-scale environmental interventions, since most studies were of much larger interventions, especially park-based interventions in the United States (US) [28]. The primary outcome in the first natural experimental study also focused on older adults, for which there was only one previous natural experimental study of an environmental intervention on physical activity specifically among older adults [29]. Further, we used a newly developed observation tool, which exacerbated this issue of limited normative data. Due to this considerable uncertainty, the assumptions around expected counts and the likely effect size from the interventions were inaccurate, which caused difficulties when fitting regression models.

Despite the inaccuracies of the estimated values in our sample size calculations, the methods used in this PhD provide a suitable approach to determine sample size in future studies that use matched intervention and comparison sites. This is a vast improvement on many previous natural experimental studies of environmental interventions that have rarely considered statistical power [6]. MOHAWk data from this study, and other recent natural experimental studies that have used MOHAWk [7], will provide valuable normative data for sample size determination in the planning of future natural experimental studies to ensure that studies are adequately powered.

7.2.7 Recommendation 7: Better reporting of samples and interventions

The natural experimental studies in this PhD were reported in line with established guidelines for the reporting of non-randomised studies (Transparent Reporting of Evaluations with Nonrandomized Designs (TREND) statement [30]) and describing interventions (Template for Intervention Description and Replication (TIDieR) checklist [31]). Using these guidelines for reporting was a simple way to increase transparency and ensure that the quality of the study can be assessed.

The TREND statement was the most appropriate study reporting guideline available as it covers non-randomised public health interventions. However, these guidelines did not specifically refer to some of the common issues associated with evaluating natural experiments, such as reporting how suitable comparison sites were identified, the timings of data collection and intervention delivery, and any disruptions to data collection [32]. This indicates a need for specific reporting guidelines for natural experimental studies of environmental interventions.

Reporting guidelines are increasingly being encouraged by journals to improve reporting standards. For example, the results of the second natural experimental study have been submitted for publication in the International Journal of Behavioral Nutrition and Physical Activity (IJBNPA) (Chapter 6) [33], which is a journal that requires the completion of intervention and study reporting checklists before submission. Journal endorsement of reporting guidelines is an effective way to ensure that studies are sufficiently reported [34].

7.2.8 Recommendation 8: Measure intervention exposure at the individual level

Due to the unobtrusive nature of observations, we were unable to obtain detailed information on participants' individual-level exposure to the intervention. Therefore, this was the only recommendation that this PhD did not attempt to address. The rise in ownership of smartphones and GPS devices offers unique opportunities to examine how different doses of intervention exposure relate to specific wellbeing behaviours. For example, smartphone apps can objectively assess intervention exposure using Global Positioning System (GPS) data [35] and can also obtain information on real-time perceptions and experiences of urban environments [36]. However, there are still important conceptual and analytical issues to overcome when assessing spatial exposure to environmental changes, such as how to interpret large and complex datasets from smartphones and GPS devices [35].

7.2.9 Summary of methodological improvements

This PhD has addressed all but one of my pre-PhD recommendations to reduce the risk of bias in natural experimental studies of environmental interventions. Minimising the risk of bias when evaluating natural experiments is undoubtedly challenging due to researchers' lack of control over the intervention [37]. However, the natural experimental studies undertaken during this PhD demonstrate that, with careful planning and methodological creativity, it is possible to conduct robust natural experimental research. Additional methodological and theoretical work is now needed to extend the work in this PhD and advance the field; this next section offers suggestions for future research.

7.3 Future research

7.3.1 Sample size calculations

Out of the eight recommendations, conducting sample size calculations was the most challenging to address. More research is needed to identify the most suitable way of calculating sample size calculations for non-randomised studies using matched intervention and comparison groups, which account for clustering and are suitable for count data. Given the substantial costs associated with primary data collection, future research should also explore ways of designing natural experimental studies that minimise costs without decreasing power. For example, how to optimise the ratio of the sample size in the intervention versus comparison group to increase statistical power [38].

To better inform future sample size calculations, more normative MOHAWk data are now needed for a wider range of environmental interventions and outside the UK. Researchers have recently been using MOHAWk to evaluate natural experiments in Valencia, Spain (M. Vallés-Planells, personal communication) and Suzhou, China (Y. Chang, personal communication). Studies should report the means, standard deviations and intracluster correlation coefficients to enable future researchers to conduct more accurate sample size calculations.

7.3.2 Individual-level outcome measures

Future studies should include outcome measures that can assess individual-level behaviour change within participants, such as household surveys or data from smartphone apps. Using individual-level outcome measures would provide better insight into the nuances around intervention effects that systematic observation cannot provide, including estimates of within-person behaviour change and potential inequity effects of interventions between subgroups of the population. It would also allow researchers to include exposure-based comparison groups. For example, sampling residents who live at different distances from the intervention site and are therefore exposed to different doses of the intervention (e.g. [39]). Triangulating population-level outcome measures (e.g. systematic observation) and individual-level outcome measures (e.g. household surveys) is the optimal approach to provide reassurance that findings are robust to the different types of bias associated with each outcome measure.

7.3.3 Routine data

Given the substantial costs and practical difficulties associated with collecting primary data around natural experiments, future studies should make use of routine data where possible i.e. data that are already being collected for regular monitoring or other evaluation purposes. Examples of route data include national survey data, smartphone data and automatic count data. If spatially sensitive and reliable routine data are available, this would provide a cost-effective way of collecting data on a much large number of individuals, thus improving statistical power. Using routine data would also make it easier to collect data at multiple time points before and after an intervention, therefore facilitating alternative study designs like interrupted time series that make use of repeated time points pre- and post-intervention [13], or synthetic controls using weighted data from several areas [10]. Routine data are also likely to be of more relevance to governments because data are often collected at administrative and spatial units of interest to government decision-making priorities. Using a combination of existing routine data and bespoke primary data would be the optimal approach for understanding the effectiveness of interventions at different scales, including site-specific, neighbourhood and city scales.

7.3.4 Camera-based observations

Conducting in-person observations is costly and labour-intensive, mainly due to recruiting and training multiple observers. As a result, it can be difficult to obtain the amount of data needed to achieve sufficient statistical power. One way to overcome this issue is to use cameras to collect video recordings in public spaces, which can then be transcribed by a researcher. Using cameras would provide many significant benefits over in-person observations, including scientific benefits (e.g. more reliable estimates of activity due to the possibility of pausing and rewinding video footage); practical benefits (e.g. overcome issues of observer availability and safety); and cost benefits (e.g. reduce costs associated with employing multiple researchers to conduct observations).

While there is emerging research on the use of wearable cameras [40] and surveillance cameras [41], limited research to date has explored the deployment of purposefully-placed cameras to capture data in outdoor public settings. If the ethical and regulatory concerns associated with deploying cameras can be addressed (e.g. complying with data protection legislation), this could enable much larger, and thus higher-powered studies, with the potential to assess a wider range of behaviours beyond those examined in this PhD, such as anti-social behaviour.

7.3.5 Multifaceted approach to evaluation

Urban green space interventions can simultaneously affect important outcomes across multiple domains [42]. Given that different stakeholders will value different outcomes [43], future research should aim to assess a broader range of impacts beyond behavioural outcomes. Examples of outcomes pertinent to urban green space interventions include other wellbeing outcomes (e.g. subjective wellbeing), more distal health outcomes (e.g. Body Mass Index), and environmental outcomes like air quality and flood mitigation. Collecting data on these other outcomes would provide a more comprehensive understanding of the inter-related potential benefits and harms of urban green space interventions, therefore appealing to a wider range of stakeholders.

7.3.6 Importance of developing theory

This PhD focussed on improving the internal validity of natural experimental studies of environmental interventions, as this has consistently been identified as a key weakness in the evidence base [1–4]. Therefore, the primary aim of this PhD was to improve the rigour of evidence for estimating urban green space intervention effectiveness (causal estimation), which meant there was less focus on developing an in-depth theoretical understanding of explaining how intervention effects occur (causal explanation) [44].

This PhD has raised important theoretical questions around causal mechanisms and how contextual factors influence intervention effectiveness. For example, contextual factors may help explain intervention ‘failure’ in the first natural experimental study (Chapter 4); that is, a similar intervention implemented in an area with less existing green space (e.g. city centres) may have bigger effects on people’s behaviour. Also, in the second natural experimental study (Chapter 6), there are uncertainties as to whether a similar urban canal intervention would produce the same effects in a less affluent area.

Despite repeated calls for more advanced theory to understand how green space interventions might work, current theoretical understanding is underdeveloped [44]. Most existing theoretical frameworks are often presented as a ‘list’ of variables and outcomes without details on the links between these variables and outcomes, the strengths of those links, and the influence of external contextual factors [45]. The development of theory can play a critical role to increase the utility of existing and new evidence for policy and practice, by producing more generalisable causal inferences based on specific intervention studies [44]. Given the length of time it takes to accumulate natural experimental research, more theory is now urgently needed to accelerate our understanding of how changes to urban green spaces might affect behaviour to inform the design and evaluation of urban green space interventions.

One promising approach to formally developing context-sensitive theory is through a realist evaluation perspective [46,47]. The main aim of realist evaluation is to test and refine programme theory, by understanding if and how an intervention worked in a particular setting for particular people i.e. “what works for whom, in what circumstances and in what respects, and how?” ([47], p.2). The realist approach can be incorporated in evaluation studies to go beyond providing overall estimates of effectiveness to show how and why intervention effects vary for different people and in different contexts [48]. Using realist evaluation has the potential to provide more nuanced insights to better inform decision makers about whether to roll out, scale up or try new urban green space interventions in different contexts.

Since the beginning of this PhD, one recent review attempted to theorise the impact of environmental interventions on physical activity using principles of realist evaluation [44]. Although this review offers an important starting point for the development of theory, the researchers relied on formal published evaluation studies, which are not only sparse and of low quality, but rarely report context in detail. As a result, this review was unable to produce the significant amount of new knowledge required to advance research and practice. Hence, theory needs to be developed using other forms of ‘evidence’ beyond formally published evaluations. For example, recent

National Institute for Health Research (NIHR) guidance on taking account of context in public health interventions recommends using stakeholders' practical experience to understand intervention effectiveness in different contexts [48]; this could be done using interviews or obtaining 'dirtier' policy or practice-oriented evaluations that are not necessarily written up and published in the public domain.

An alternative, albeit similar approach to developing theory, is 'systems thinking'. Taking a systems thinking approach acknowledges that public health outcomes are derived from a dynamic and complex system of interdependent "linkages, relationships, feedback loops and interactions" ([49], p. 269) that operate at multiple levels; interventions are conceptualised as events that occur within this system. Systems thinking approaches are similar to realist evaluation in the sense that they acknowledge that interventions operate within a 'system', which is somewhat analogous to context in realist approaches. However, there is currently limited guidance on how to translate systems thinking concepts into theory [50], which makes the task of trying to accurately map the whole system extremely challenging. In contrast, realist evaluation provides a more refined methodology that can focus on specific dimensions of context that are most important to the success of interventions.

7.4 Implications for research

7.4.1 Feasibility of these methods for evaluating natural experiments

It is important to examine the factors that facilitated the robust methods used in this PhD to consider the feasibility of these methods for future natural experimental research. The following discussion focuses on natural experimental studies set up prospectively before the intervention was implemented, rather than retrospective evaluations of natural experiments that have already occurred.

Identifying and responding to natural experiment opportunities is a major challenge. Environmental changes are often introduced quickly, which results in limited time to plan a robust study and obtain ethical clearance before the intervention begins. This issue is exacerbated by a research funding system that is less suited to respond to the fast-paced nature of natural experimental research and prioritises RCTs for evaluating intervention effectiveness [37]. Whilst part of the onus is on research funding bodies to increase flexibility and speed to respond to natural experiment opportunities (e.g. rapid funding schemes), the methods used in this PhD provide a template study design that can be used to facilitate rapid data collection. For example, we have provided examples of reliable observation schedules and normative data in a range of urban environments to assist with power calculations. The efficiency of using these methods was demonstrated in the second natural experimental study (Chapter 6), as baseline data were collected within just six weeks between initially discovering the natural experiment opportunity and the intervention being implemented. However, it must be acknowledged that ethical approval was granted within a matter of days as an amendment to an existing application, which is not always feasible as long delays are common for obtaining ethical approval [51].

Meticulous methodological planning was crucial to ensure rigour was minimally affected by unforeseen external events; publishing a study protocol encouraged us to think about contingency procedures for such events. Fortunately, many of the potential issues that can affect internal validity in natural experimental studies did not occur: the interventions went ahead as planned; there were limited external events which compromised internal validity (e.g. unplanned changes to comparison sites); and there were few disruptions to data collection (e.g. cancelled observations due to safety threats to the observer). Nevertheless, our methods were resistant to such unforeseen circumstances. For example, one of the comparison sites had to be removed from the study due to safety concerns for the observer (Chapter 4), but having multiple comparison sites ensured we still had sufficient sites for analysis.

Another source of methodological rigour was having strong relationships with the intervention providers. Having strong partnerships is a common theme in producing robust practice-based natural experimental research [52]. We developed close relationships with stakeholders, who could keep us informed of upcoming natural experiment opportunities and contributed to reducing risk of bias by communicating any important changes to the intervention or other external factors that may impact findings. Importantly, the intervention providers had no financial or other involvement in the collection, analyses or interpretation of data, or in the writing of the papers and the decision to publish the results, which reduced the risk of bias due to selective reporting or data fabrication.

7.4.2 Importance of evaluating small-scale interventions

This PhD has highlighted the potential for relatively small-scale environmental changes to bring about measurable effects on green space usage and wellbeing behaviours (Chapter 6). There are a growing number of recent natural experimental studies that have also found positive effects of other types of small-scale green space interventions on physical activity and other wellbeing outcomes; examples include greening vacant lots [53], improving access to local woodlands [54] and revitalising urban wastelands [55]. However, the null findings in our first natural experimental study (Chapter 4) highlight the need for a more in-depth theoretical understanding to find the ‘sweet spot’ between interventions that are low cost but still bring about measurable effects on behaviour. The methods used in this PhD are suitable for evaluating small-scale interventions, as they were developed to capture changes in behaviour at the site level.

7.5 Implications for policy and practice

Decision makers in policy and practice are becoming increasingly aware of the potential of using urban green space interventions to improve health and wellbeing outcomes. Intervention providers in this PhD believed that the interventions would increase the use of specific areas targeted for environmental improvements and promote health and wellbeing. A recent qualitative study that looked at local level decision-making found that these expectations are likely to be partly based on policy and industry guidance documents [56], such as guidance from Public Health England [57] and the Town and Country Planning Association [58]. However, a key issue with these guidance documents is that there is a shortage of robust evidence to underpin recommendations, which means that these recommendations are predominantly based on expert consensus. As shown in this PhD, robust research is crucial to test the validity of these recommendations to avoid producing guidance that may overclaim (as shown in Chapter 4) or underclaim (as shown in Chapter 6) potential health and wellbeing benefits of urban green space interventions. Without any robust evidence or theory to guide expert consensus, it is unsurprising that predictions about what interventions are effective are inaccurate, let alone be able to consider how intervention effects vary for different subpopulations by different types of green space interventions in different local contexts [59]. Although it is unrealistic to recommend that no action is taken until 'sufficient' robust evidence has been cumulated [37], recommendations ought to be more modest due to the considerable current uncertainty. Demand for robust practice-based evidence like this is growing in policy and practice [60], but how to translate research into urban planning policy and decision-making is a longstanding issue that is being explored in more recent research [56,61,62].

7.6 The impact of COVID-19

The COVID-19 pandemic has caused unprecedented worldwide impacts on urban mobility and the use of urban green spaces [63]. Whilst the planning, design and conduct of this PhD research was completed before the COVID-19 pandemic, it feels apt to briefly reflect how the research from this PhD is likely to be influenced by, and influence responses to, COVID-19.

This pandemic has exacerbated the need to understand how people can safely access urban green spaces to support population wellbeing in conjunction with COVID-19 physical distancing strategies. With major disruptions to day-to-day living (e.g. likely reductions in physical activity trips for commuting purposes [64]), neighbourhood urban green spaces will play a vital role in improving health and wellbeing outcomes and alleviate demand on the healthcare system. Hence, there is an urgent need to incorporate more good quality green space within cities and towns, not only because of COVID-19, but also to address other major societal issues such as mental ill-health and climate change. The methods used in this PhD will now be of even higher importance to ensure researchers can react quickly to intervention opportunities and provide robust evidence on how novel post-COVID-19 temporary and permanent interventions affect the use of urban green spaces.

7.7 Conclusions

This PhD has provided robust evidence regarding the causal effects of urban green space interventions on physical activity and other behaviours important for wellbeing. These natural experimental studies provide exemplars of how to use methods with substantially lower risk of bias than previous research, by addressing important methodological weaknesses prevalent in previous natural experimental studies of environmental interventions. More robust natural experimental studies like these are now needed to investigate the (in)effectiveness of a wider range of urban green space interventions in different contexts to better inform policy and practice recommendations, especially in Europe where there is a dearth of robust evidence. We have provided evidence of the reliability and validity of MOHAWk; an unobtrusive and inexpensive observation tool that will facilitate more robust natural experimental studies, particularly of small-scale urban green space interventions, on a range of wellbeing behaviours beyond physical activity.

7.8 References

1. Hunter RF, Christian H, Veitch J, Astell-Burt T, Hipp JA, Schipperijn J. The impact of interventions to promote physical activity in urban green space: A systematic review and recommendations for future research. Soc Sci Med. 2015;124:246–56.

2. Roberts H, McEachan R, Margary T, Conner M, Kellar I. Identifying Effective Behavior Change Techniques in Built Environment Interventions to Increase Use of Green Space: A Systematic Review. Environ Behav. 2018;50(1):28–55.

3. MacMillan F, George ES, Feng X, Merom D, Bennie A, Cook A, et al. Do Natural Experiments of Changes in Neighborhood Built Environment Impact Physical Activity and Diet? A Systematic Review. Int J Environ Res Public Health. 2018;15(2):217.

4. Moore THM, Kesten JM, López-López JA, Ijaz S, McAleenan A, Richards A, et al. The effects of changes to the built environment on the mental health and well-being of adults: Systematic review. Health Place. 2018;53:237–57.

5. Stappers NEH, Van Kann DHH, Ettema D, De Vries NK, Kremers SPJ. The effect of infrastructural changes in the built environment on physical activity, active transportation and sedentary behavior – A systematic review. Heal Place. 2018;53:135–49.

6. Benton JS, Anderson J, Hunter RF, French DP. The effect of changing the built environment on physical activity: a quantitative review of the risk of bias in natural experiments. Int J Behav Nutr Phys Act. 2016;13(1):107.

7. Anderson J, French DP, Benton JS, Dennis M, Rothwell J. The effects of the Grow Green urban greening improvements on physical health and wellbeing behaviours in Manchester UK: A natural experimental study. Open Sci Framew.

2020; https://osf.io/zqgcn

8. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJ, Martin BW. Correlates of physical activity: why are some people physically active and others not? Lancet. 2012;380(9838):258–71.

9. Craig P, Gunnell D, Ogilvie D, Sutton M, Thompson S. Using natural experiments to evaluate population health interventions: guidance for producers and users of evidence Prepared on behalf of the Medical Research Council by. 2012. Available from: www.mrc.ac.uk/naturalexperimentsguidance

10. O'Neill S, Kreif N, Grieve R, Sutton M, Sekhon JS. Estimating causal effects: considering three alternatives to difference-in-differences estimation. Heal Serv Outcomes Res Methodol. 2016;16(1–2):1–21.

11. Craig P, Vittal Katikireddi S, Leyland A, Popham F. Natural Experiments: An Overview of Methods, Approaches, and Contributions to Public Health Intervention Research. Annu Rev Public Heal. 2017;38:39–56.

12. Zeldow B, Hatfield LA. Confounding and Regression Adjustment in Difference-in-Differences. arXiv e-prints. 2019;1–33.

13. Bernal JL, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: A tutorial. Int J Epidemiol. 2017;46(1):348–55.

14. Butsic V, Lewis DJ, Radeloff VC, Baumann M, Kuemmerle T. Quasi-experimental methods enable stronger inferences from observational data in ecology. Basic Appl Ecol. 2017;19:1–10.

15. Lindner S, McConnell KJ. Difference-in-differences and matching on outcomes: a tale of two unobservables. Heal Serv Outcomes Res Methodol. 2019;19(2–3):127–44.

16. Austin PC, Leckie G. The effect of number of clusters and cluster size on statistical power and Type I error rates when testing random effects variance components in multilevel linear and logistic regression models The effect of number of clusters and cluster size on statisti. J Stat Comput Simul. 2018.

17. Clarke P. When can group level clustering be ignored? Multilevel models versus single-level models with sparse data. J Epidemiol Community Health. 2008;62:752–8.

18. Benton JS, Anderson J, Cotterill S, Dennis M, Lindley SJ, French DP. Evaluating the impact of improvements in urban green space on older adults' physical activity and wellbeing: protocol for a natural experimental study. BMC Public Health. 2018;18(1):923.

19. Benton J, Anderson J, Cotterill S, Dennis M, French DP. The impact of new walking infrastructure and changes to green space along an urban canal on physical activity and wellbeing: protocol for a natural experimental study. Open Sci Framew. 2018; https://osf.io/zcm7v/

20. Benton JS, Anderson J, Pulis M, Cotterill S, Hunter RF, French DP. Method for Observing pHysical Activity and Wellbeing (MOHAWk): validation of an observation tool to assess physical activity and other wellbeing behaviours in urban spaces. Cities Health. 2020;

21. Miles LM, Elbourne D, Farmer A, Gulliford M, Locock L, McCambridge J, et al. Bias due to MEasurement Reactions In Trials to improve health (MERIT): protocol for research to develop MRC guidance. Trials. 2018;19(1):653.

22. McKenzie TL, van der Mars H. Top 10 Research Questions Related to Assessing Physical Activity and Its Contexts Using Systematic Observation. Res Q Exerc Sport. 2015;86(1):13–29.

23. French DP, Sutton S. Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? Br J Health Psychol. 2010;15(3):453–68.

24. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ. 2015;350.

25. Donner A, Klar N. Design and Analysis of Cluster Randomization Trials in Health Research. John Wiley & Sons; 2000. 178 p.

26. Freedman LS, Green SB, Byar DP. Assessing the gain in efficiency due to matching in a community intervention study. Stat Med. 1990;9(8):943–52.

27. Cunningham T. Power and Sample Size for Three-Level Cluster Designs. Virginia Commonwealth University; 2010. Available from: https://core.ac.uk/download/pdf/51291682.pdf

28. World Health Organization. Urban Green Space Interventions and Health: A review of evidence. Copenhagen; 2017. Available from: http://www.euro.who.int/__data/assets/pdf_file/0005/321971/Urban-green-spaces-and-health-review-evidence.pdf?ua=1

29. Ward Thompson C, Curl A, Aspinall P, Alves S, Zuin A. Do changes to the local street environment alter behaviour and quality of life of older adults? The “DIY Streets” intervention. Br J Sports Med. 2014;48(13):1059–65.

30. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health. 2004;94(3):361–6.

31. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication

(TIDieR) checklist and guide. Res Methods Report. 2014;348:g1687.

32. Veitch J, Salmon J, Giles-Corti B, Crawford D, Dullaghan K, Carver A, et al. Challenges in conducting natural experiments in parks—lessons from the REVAMP study. Int J Behav Nutr Phys Act. 2017;14(1):5.

33. Benton JS, Cotterill S, Anderson J, Macintyre VG, Gittins M, Dennis M, et al. A natural experimental study of improvements along an urban canal: impact on canal usage, physical activity and other wellbeing behaviours. Int J Behav Nutr Phys Act. Under Review.

34. Caron JE, March JK, Cohen MB, Schmidt RL. A Survey of the Prevalence and Impact of Reporting Guideline Endorsement in Pathology Journals. Am J Clin Pathol Oct. 2017;148(4):314–22.

35. Humphreys DK, Panter J, Sahlqvist S, Goodman A, Ogilvie D. Changing the environment to improve population health: a framework for considering exposure in natural experimental studies. J Epidemiol Community Health. 2016;70(9):941–6.

36. McEwan K, Richardson M, Brindley P, Sheffield D, Tait C, Johnson S, et al. Shmapped: development of an app to record and promote the well-being benefits of noticing urban nature. Transl Behav Med. 2020;10(3):723–33.

37. Ogilvie D, Adams J, Bauman A, Gregg EW, Panter J, Siegel KR, et al. Using natural experimental studies to guide public health action: turning the evidence-based medicine paradigm on its head. J Epidemiol Community Heal. 2019;74(2):203–8.

38. Mcconnell B, Vera-Hernández M. Going beyond simple sample size calculations: A practitioner’s guide. London; 2015. (IFS Working Papers). Available from: http://hdl.handle.net/10419/119557

39. Ogilvie D, Bull F, Cooper A, Rutter H, Adams E, Brand C, et al. Evaluating the travel, physical activity and carbon impacts of a "natural experiment" in the provision of new walking and cycling infrastructure: methods for the core module of the iConnect study. BMJ Open. 2012;2(1):e000694.

40. Doherty AR, Hodges SE, King AC, Smeaton AF, Berry E, Moulin CJA, et al. Wearable cameras in health: The state of the art and future possibilities. Am J Prev Med. 2013;44(3):320–3.

41. Carlson J, Liu B, Sallis J, Kerr J, Hipp J, Staggs V, et al. Automated Ecological Assessment of Physical Activity: Advancing Direct Observation. Int J Environ Res Public Health. 2017;14(12):1487.

42. Shanahan DF, Astell–Burt T, Barber EA, Brymer E, Cox DTC, Dean J, et al. Nature–Based Interventions for Improving Health and Wellbeing: The Purpose, the People and the Outcomes. Sports. 2019;7(6):141.

43. Egan M, McGill E, Penney T, Anderson de Cuevas R, Er V, Orton L, et al. NIHR SPHR Guidance on Systems Approaches to Local Public Health Evaluation. Part 1: Introducing systems thinking. London; 2019.

44. Panter J, Guell C, Humphreys D, Ogilvie D. Can changing the physical environment promote walking and cycling? A systematic review of what works and how. Health Place. 2019;58:102161.

45. Panter J, Guell C, Prins R, Ogilvie D. Physical activity and the environment: conceptual review and framework for intervention research. Int J Behav Nutr Phys Act. 2017;14(1):156.

46. Pawson R, Tilley N. Realistic Evaluation. London: SAGE Publications; 1997.

47. Pawson R, Tilley N. Realistic evaluation. In: Mathison S, editor. Encyclopedia of evaluation. Thousand Oaks: Sage; 2004. p. 362–7.

48. Craig P, Di Ruggiero E, Frohlich KL, Mykhalovskiy E, White M, Campbell R, et al. Taking account of context in population health intervention research: guidance for producers, users and funders of research. Southampton; 2018. Available from: https://www.ncbi.nlm.nih.gov/books/NBK498645/pdf/Bookshelf_NBK498645.pdf

49. Hawe P, Shiell A, Riley T. Theorising Interventions as Events in Systems. Am J Community Psychol. 2009;43:267–76.

50. Chen HT. Interfacing theories of program with theories of evaluation for advancing evaluation practice: Reductionism, systems thinking, and pragmatic synthesis. 2016;

51. Lewis JC, Tomkins S, Sampson JR. Ethical approval for research involving geographically dispersed subjects: unsuitability of the UK MREC/LREC system and relevance to uncommon genetic disorders. J Med Ethics. 2001;27:347–51.

52. Giles-Corti B, Whitzman C. Active living research: Partnerships that count. Heal Place. 2012;18(1):118–20.

53. South EC, Hohl BC, Kondo MC, MacDonald JM, Branas CC. Effect of Greening Vacant Land on Mental Health of Community-Dwelling Adults: A Cluster Randomized Trial. JAMA Netw open. 2018;1(3):e180298.

54. Ward Thompson C, Silveirinha de Oliveira E, Tilley S, Elizalde A, Botha W, Briggs A, et al. Health impacts of environmental and social interventions designed to increase deprived communities' access to urban woodlands: a mixed-methods study. Public Heal Res. 2019;7(2):1–172.

55. Unt AL, Bell S. The impact of small-scale design interventions on the behaviour patterns of the users of an urban wasteland. Urban For Urban Green. 2014;13(1):121–35.

56. Le Gouais A, Foley L, Ogilvie D, Guell C. Decision-making for Active Living Infrastructure in New Communities: A Qualitative Study in England. J Public Health. 2019;105.

57. Public Health England. Spatial Planning for Health An evidence resource for planning and designing healthier places. 2017. Available from: https://www.gov.uk/government/publications/spatial-planning-for-health-evidence-review

58. Ross A, Chang M. Planning healthy-weight environments - a TCPA reuniting health with planning project. London; 2014. Available from: https://www.tcpa.org.uk/Handlers/Download.ashx?IDMF=363c8a41-e45e-46cf-9692-1ebceebc94d5

59. Douglas O, Lennon M, Scott M. Green space benefits for health and well-being: A life-course approach for urban planning, design and management. Cities. 2017;66:53–62.

60. Leatherdale ST. Natural experiment methodology for research: a review of how different methods can support real-world research. Int J Soc Res Methodol. 2018;22(1):19–35.

61. Pineo H, Zimmermann N, Davies M. Integrating health into the complex urban planning policy and decision-making context: a systems thinking analysis. Palgrave Commun. 2020;6(1):1–14.

62. Lge-Elegbede J, Pilkington P, Bird EL, Gray S, Mindell JS, Chang M, et al. Exploring the views of planners and public health practitioners on integrating health evidence into spatial planning in England: a mixed-methods study. J Public Health. 2020;1–9.

63. Ugolini F, Massetti L, Calaza-Martínez P, Cariñanos P, Dobbs C, Ostoić SK, et al.

Effects of the COVID-19 pandemic on the use and perceptions of urban green space: An international exploratory study. Urban For Urban Green. 2020;56:126888.

64. Laverty AA, Millett C, Majeed A, Vamos EP. COVID-19 presents opportunities and threats to transport and health. J R Soc Med. 2020;113(7):251–4.

Appendices

Appendix A. MOHAWk instruction manual

MOHAWk

(Method for Observing pHysical Activity and Wellbeing)

Instruction Manual

Revised 12/11/2019

1. OVERVIEW OF MOHAWk

Introduction to MOHAWk

Method for Observing Physical Activity and Wellbeing (MOHAWk) is an observation tool designed for assessing three levels of physical activity (Sedentary, Walking and Vigorous) and two other wellbeing behaviours (Connect: socially interacting with others; and Take Notice: taking notice of the environment) in urban spaces. Using direct observations of people and their behaviour, MOHAWk assesses the number of people that enter a predetermined area ('target area'), their characteristics, physical activity levels, and whether they engage in social interactions or take notice of the environment. It also measures the presence of incivilities in the environment where observations are carried out. MOHAWk has been designed to be used in a wide variety of urban spaces, particularly spaces that have lower numbers of users or that people pass through, such as pocket parks, residential streets, civic squares, and canal towpaths.

MOHAWk is freely available for use. The tool consists of an instruction manual, a standardised observation form, and a summary form – all of which should be provided separately (or alternatively available in the Appendices in this manual).

Summary of how it works

MOHAWk uses continuous scanning to record the characteristics and behaviours of each person entering a pre-determined boundary ('target area') during hour-long time periods ('observation periods'). Data are recorded using pen and paper. Observers use a standardised observation form to record the following information for each person that enters the target area during each observation period:

- Physical activity level (Sedentary, Walking, Vigorous)
- Other wellbeing behaviours (Connect, Take Notice)
- Activity type (Cycling, Using phone, Dog walking or other pre-determined activities)
- Gender (Male or Female)
- Age group (Infant, Child, Teen, Adult or Older Adult)
- Ethnicity (White or Non-White)
- Mobility assistance required (Yes or No)

The following contextual information is also recorded:

- Extent of the following incivilities in the target area before the observation period:
    - General litter
    - Evidence of alcohol use (empty bottles/ cans)
    - Evidence of drug taking (e.g. needles, syringes)
    - Graffiti
    - Broken glass
    - Vandalism
    - Dog mess
    - Noise
- Weather (duration of any precipitation during the observation period)

2. OBSERVATION PREPARATION

Defining the target area

Before data collection begins, the site should be visited by all observers, who should agree on the precise boundaries of the target area. This is important so that all observers can agree on whether an individual falls within the boundary of that area or not. Any individual and their behaviour should only be recorded within this target area.

A target area should be large enough to accommodate activity, but small enough so that observers can accurately count everyone who moves into the target area. The target area should be located where activity is most likely to occur. The target area should be small enough so that observers can reliably view and record the characteristics and activity of all people that enter the target area.

A location should be identified where the observer should position themselves during observations, with a good view of the target area and no visual obstructions. Ideally, this should include an area with good cover for adverse weather. If an area that affords a good view with good cover is not possible, an area that affords a reasonable view with good cover should be identified for particularly severe weather. This is important because observations are carried out regardless of weather conditions, unless weather conditions become so extreme that they compromise the observer’s safety.

Target areas boundaries are not physically drawn out. However, existing boundaries (e.g. fences, road junctions, pavement markings) can help define a target area. Although, these boundaries need to be the same throughout all observation periods.

At least one “safe” area should be identified, where the observer can go if anyone being observed becomes aggressive e.g. café, public library, busy public space.

Before observing the target area

All observers should be fully trained in using MOHAWk. Ideally, all observers should be trained together and should carry out practice observations prior to the ‘real’ observations. This will help calibrate observers and allow observers to agree on any ambiguities; thus improving agreement between observers i.e. ‘inter-rater reliability’.

Inter-rater reliability calculations should be conducted to formally assess whether observers are achieving acceptable agreement with each other, using intraclass correlation coefficients (ICC). Observers should aim to achieve inter-rater reliability of at least ICC = 0.75. Achieving high inter-rater reliability means that different observers can use MOHAWk and still produce very similar data, which is essential when using multiple observers. Any discrepancies between observers were resolved by discussion.

Print out a copy of the site map so that the target area can be clearly defined. Also, print several copies of a letter detailing information about the study (including details of ethical approval), in the event that any member of the public requests more information about what observers are doing.

It is recommended that photographs and environmental audits of the target area are taken prior to observations. An environmental audit of the target area should be carried out using a validated tool appropriate for the particular area under investigation and the purposes of the evaluation e.g. the Environmental Assessment of Public Recreation Spaces (EAPRS) tool[1] is suitable for assessing the quality of public open spaces.

Observers should arrive at the site at least 15 minutes prior to the official start of coding. In this time before observations begin, observers should record the extent of incivilities for each target area (General litter, Evidence of alcohol use, Evidence of drug taking, Graffiti, Broken glass, Vandalism, Dog mess, Noise – see next section).

3. CODES AND RECORDING

[1] Saelens BE, Frank LD, Auffrey C, Whitaker RC, Burdette HL, Colabianchi N. Measuring physical environments of parks and playgrounds: EAPRS instrument development and inter-rater reliability. Journal of Physical Activity and Health. 2006 Feb 1;3(s1):S190-207.

Target area incivilities [2]

Tick “None”, “Hardly noticeable”, “Noticeable”, or “Very Noticeable” to describe specific conditions for the target area.

Scoring for the extent of incivilities should be based on whether people using the green space would be generally aware of its presence (observers should not have to forensically examine all parts of the green space).

Incivility	None	Hardly noticeable	Noticeable	Very noticeable
General litter	None visible	Hardly noticeable A few items visible on the ground	Noticeable Several items are on the ground	Very noticeable Many items are on the ground
Evidence of alcohol use (bottles, cans, or bottle caps visible)	None visible	Hardly noticeable A few items	Noticeable Several items	Very noticeable
Evidence of drug taking (e.g. paint cans, rags, baggies, rolling papers)	None visible	None visible	One or more clear examples of drug taking	Very noticeable
Graffiti	None visible	Hardly noticeable, but it appears on up to a few pieces of furniture/ equipment	Noticeable Several small or large pieces, clearly visible	Very noticeable Several large pieces, on much of the park furniture/ equipment
Broken glass	None visible	A few pieces of broken glass, does not really spoil enjoyment of space	Several pieces of broken glass, affecting enjoyment of area	Many pieces of broken glass, affecting enjoyment of area

[2] Target area incivilities based on: Gidlow CJ, Ellis NJ, Bostock S. Development of the neighbourhood green space tool (NGST). Landscape and Urban Planning. 2012;106(4):347-58.

Vandalism	None visible	Hardly noticeable, but some evidence on up to a few pieces of furniture/ equipment	Noticeable, more than a few pieces of equipment **OR** an area of the space has been rendered unusable by vandalism	Very noticeable, more equipment in disrepair than in good order because of vandalism. Signs of vandalism are obvious
Dog mess	None visible	Hardly noticeable, perhaps a single example	Noticeable/ several dog refuse piles, affecting enjoyment of area	Very noticeable, seriously affecting enjoyment of area
Noise	Not aware of any	Some sound but hardly noticeable, not annoying	Sound(s) is (are) noticeable and interfere(s) with enjoyment of area	Noticeable sounds which are unpleasant. Seriously affecting enjoyment of area

MOHAWk observations

Gender

Individuals are recorded as either Male or Female.

Age Group

Individuals are coded by age group according to the following criteria:

Infant = Babies or toddlers in a pram, or other baby carrier/ sling.

NOTE: Do not record gender, ethnicity or activity level for infants.

Child = Individuals from infancy to 12 years of age.

NOTE: Children can be identified by general appearance, especially if they are wearing a school uniform. Children are more likely to be accompanied by a parent.

Teen = Individuals aged 13 to 19 years of age

NOTE: Teens are likely to dress more 'extravagant' than children under 12 years of age. Any individual that appears like a secondary school student, college student or university undergraduate can be recorded as a Teen.

Adult = Individuals aged 20 to 64 years of age.

Older Adult = Individuals aged 65 years of age and older.

NOTE: Individuals recorded as an Older Adult will be those who appear to be 65 years or older based on their general appearance and mobility. The main criteria observers should take note of are gait and general movement, as older adults may show signs of aging with stiffer, slower, or simply inhibited movements. Observers should also look for physical attributes traditionally associated with age, such as grey hair, wrinkles, or lack of hair, to decide if issues with gait or mobility appear to be linked to advancing age or other causes of impairment. People who are wearing work-related uniforms may be indicative of an adult rather than an older adult.

Ethnicity

Record whether the ethnicity for each person is White or Non-White (Asian, black, mixed, or other). Whilst there may be ambiguities in deciding whether some ethnicities are White or Non-White (e.g. mixed ethnicity), such ambiguities should be agreed before observations begin and observers should be consistent in recording ethnicity.

Physical Activity Level [3]

Record the activity level based on the following criteria (more than one activity level can be recorded for each person):

Sedentary (S) = Individuals are lying down, sitting, or standing in place.

> NOTE: Observers should consider whether that behaviour would be recorded as Sedentary if they conducted one brief scan of the target area; if the answer is yes, then it is likely that the behaviour can be recorded as Sedentary.

> NOTE: Very brief or incidental instances of sitting or standing should not be recorded as Sedentary behaviour. For example, the following examples usually would not be recorded as Sedentary as these are brief or incidental behaviours: briefly stopping to pick up a piece of litter, briefly waiting for a car to pass before crossing the road. Whereas the following examples would usually be recorded as Sedentary as these are prolonged deliberative behaviours: stopping to talk to someone, lying down, sitting on a bench.

Walking (W) = Individuals are walking (or moving) at a casual pace.

Vigorous (V) = Individuals are currently engaged in an activity more vigorous than an ordinary walk (e.g. increasing heart rate causing them to sweat, such as jogging, power walking, doing cart wheels, skipping). All cyclists should be recorded as Vigorous, unless they are walking with their bicycle in which case they should be recorded as Walking.

> NOTE: When a person cycles into the observation area, dismounts, locks cycle and then walks away, then it may be appropriate for all three physical activity behaviours to be recorded i.e., Sedentary, Walking and Vigorous.

> NOTE: Individuals in a mobility scooter or electric wheelchair should be recorded as Sedentary, whereas those in a manual wheelchair should be recorded as Walking or Vigorous depending on how intensely they are moving.

---

[3] Physical activity intensities based on SOPARC: McKenzie TL, Cohen DA, Sehgal A, Williamson S, Golinelli D. System for Observing Play and Recreation in Communities (SOPARC): reliability and feasibility measures. Journal of Physical Activity and Health. 2006 Feb;3(s1):S208-22.

NOTE: If there are any uncertainties for what physical activity intensity an activity should be recorded as, then consult the 2011 compendium of physical activities for a list of the energy cost of a wide variety of physical activity [4]. Use the following validated metabolic equivalent (MET) conversion for each physical activity level to identify the closest level:

Sedentary = 1.5 METs (i.e. <2.9 METs)

Walking = 3 METs (i.e. ≫3 METs and <6 METs)

Vigorous = 6 METs (i.e. ≫6 METs)

Other wellbeing behaviours

Take Notice (TN) = Individuals stop or slow down, and appear as if they are making a conscious decision to appreciate their surroundings. Examples of this could be (a) extended viewing of the scenery, (b) an intentional pause in activity to look at or photograph something in the vicinity, or (c) a pronounced head swivel to look at a specific object, view or person.

The following scenarios are not to be recorded as Take Notice behaviours:

- The person does not stop or slow down.
- Staring into space (e.g. when smoking or looking closely at phones).
- Looking around when crossing a road, junction, pathway or other crossing.
- A person is on a phone call.
- Taking notice of the observer.

NOTE: If there is uncertainty whether a person engaged in Take Notice behaviour, then it is unlikely that this is sufficient to be recorded as Take Notice.

NOTE: Take Notice behaviours are sometimes momentary behaviours, in a similar way as Sedentary behaviours. Therefore, observers should consider whether that behaviour would be recorded as Take Notice if they conducted one brief scan of the Target area; if the answer is yes, then it is likely that the behaviour can be recorded as Take Notice.

[4] https://www.ncbi.nlm.nih.gov/pubmed/21681120.

Connect (C) = Individuals are engaging/ interacting with a person or the people around them in some way. The activity must involve either (a) conversing (e.g. talking and listening, or using sign language) with other users, (b) being physically linked with someone (e.g. holding hands, linked arms, being carried on shoulders), (c) smiling and making eye contact when passing through a door or other narrow space, or (d) participation in a group activity. Note that when two people are interacting, this should be recorded as two behaviours i.e. one for each individual person.

The following scenarios are <u>not</u> to be recorded as Connect behaviours:

- Individuals who are not physically or verbally interacting with others in the same vicinity (e.g. phone call, video call).
- Walking or cycling side by side, unless socially interacting with each other (e.g. talking)
- Interacting with the observer.

Activity Type

There are three predefined types of activity that observers can record during observations:

- Cycling (riding, walking or standing with a bicycle)
- Using Phone (holding and using a phone in some way e.g. texting, phone call, taking a picture or recording a video)
- Dog Walking (walking with a dog, whether it be on or off a lead)

  NOTE: multiple people walking the same dog can all be recorded as Dog Walking e.g. if two individuals are walking together with one dog then both individuals should be recorded as Dog Walking

There is space on the MOHAWk data collection form to add further activities that researchers may be interested in recording. For example, researchers may be interested in activities specific to a particular site e.g. fishing may be a prominent activity of interest when observing along a canal waterway (see Appendices 3 and 4 for an example). Alternatively, researchers might want to record specific Connect (e.g. holding hands) or Take Notice (e.g. taking photographs) behaviours.

Group/ Busy

Large groups

When there is a large group of people (~ >10 people together), observers should only record the number of people in that group and record estimates of the frequency (or percentage if the overall count is known) of each activity level and category for age group, gender, ethnicity, and mobility assistance (see Appendices 3 and 4 for an example).

When using this procedure for large groups, observers should prioritise estimates for the most important variables depending on the primary focus of the evaluation. For example, if the primary aim of the evaluation is to assess the impact of an intervention specifically on women's' physical activity levels, then prioritise estimates of gender and physical activity levels.

Busy periods

The procedure for coping with large groups can also be used during extremely busy periods as a last resort i.e. when it is impossible to reliably record the characteristics and activity of each separate individual. For example, if it is extremely busy between 3pm and 3:10pm, then observers should record the overall number of people and an estimate of the most important variables. If this happens regularly then it is likely that MOHAWk is an unsuitable tool for that particular site. Consider reducing the size of the target area or consider a different observation tool.

Mobility assistance

Record whether the individual requires assistance to move. For example, use of a walking stick, crutches, wheelchair, mobility scooter or a helper to assist with movement. Individuals who are limping should not be classed as requiring mobility assistance.

Notes

If needed, there is a column for the observer to make notes on each observed individual. This should only be used if needed to keep track of an individual, particularly if they remain in the target area for a prolonged period of time to avoid double counting that individual and their activities - see Appendix 3 for an example.

Weather

Record the duration (approximate start and end time) of any precipitation that occurs during the observation period. Ensure to do this throughout the observation period e.g. “rain from 10-10.15am and 10.40-10.55am”.

Comments

Make a note of any other relevant observations that occur during the observation period e.g. if anyone spoke to the observer during the observation period, if any activities were particularly popular during the observation period, reasons for any missing data, toilet breaks etc.

4. OBSERVATION PROCEDURES

How to observe

Observers should continuously scan the target area for the full hour. When a person enters the target area, observers should record the following for that person: gender (Male or Female), ethnicity (White or Non-white), age group (Infant, Child, Teen, Adult or Older Adult), physical activity level (Sedentary, Walking, Vigorous), social interaction (Connect), taking notice of the environment (Take Notice), activity type (Cycling, Using phone, Dog walking, or other pre-determined activities), and if mobility assistance is required (Yes or No).

Record all individuals entering the target area, apart from those in a vehicle (e.g. car, motorbike).

If an observed person re-enters the target area in the same observation period, do not record a second time.

The unit of coding is the behaviour, so that the number of people performing each behaviour should be counted. Therefore, the same person can be recorded as engaging in multiple behaviours. However, each behaviour cannot be recorded more than once for the same person within the same observation period e.g. if a person speaks to someone and then hugs someone else within the same observation period, this would only be recorded as one Connect behaviour for that person.

Observers can make a mark on the observation form if they want to split the hour-long observation period into smaller blocks (e.g. 15 minute blocks) - see Appendix 3 for an example. This provides researchers with the option to analyse data by smaller blocks as well as hour-long observation periods.

If necessary, observers can move around the target area, as long as observers do not interfere with activity and only record individuals entering the target area.

It may be more difficult for observers to reliably record individual's characteristics and their activity levels during adverse weather conditions e.g. people are more likely to wear protective clothing, such as hooded jackets, when it is cold or raining. However, to avoid missing data, observers should record their best estimate of individual's characteristics and activity levels.

Observation procedure

1. Prior to the start of the observation period, record the Date/ Day, Site name, Observer initials, Start time and End time on the top of the observation form
2. Score the extent of incivilities in each area (General litter, Evidence of alcohol use, Evidence of drug taking, Graffiti, Broken glass, Vandalism, Dog mess, Noise) using the MOHAWk summary form (see Appendix 2 for the form and Appendix 4 for an example).
3. Record the gender, ethnicity, age group, physical activity level, whether they engage in social interaction or take notice of the environment, activity type and whether mobility assistance is required for all individuals that enter the target area using the MOHAWk observation form (see Appendix 1 for the observation form and Appendix 3 for an example). Record the duration of any precipitation that occurs during the observation period.
4. After the observation period has finished, record total counts of gender, age group, ethnicity, activity type, activity level, and the number of individuals requiring mobility assistance onto the MOHAWk summary form. Also, record the total duration of any precipitation that occurred.

Toilet breaks

A designated place for toilet breaks should be determined during the observation preparation stage. Toilet breaks should be planned to occur before or after observation periods. However, should a toilet break be necessary then this should ideally be taken in the final minutes of the observation period. Any toilet breaks should be documented on the comments section of the MOHAWk data collection form.

Respecting the public

Whilst observers should avoid standing out, observers should be overt about what they are doing (e.g. wear a high visibility vest, use a clipboard) and should be willing to respond honestly to any questions from members of the public during observations. Observers should ensure they respect members of the public and politely engage with them, even if it means that they miss individuals entering the target area. If any members of the public (particularly local residents) request that the observer should stop, then the observer should immediately stop and postpone all planned future observations at that particular site.

5. OBSERVATION PERIODS

Timing and frequency of observation periods

Observation periods are one hour long. Observation periods can be carried out at any time of the day depending on the aims and requirements of the evaluation, and available time and resources. For example, MOHAWk was recently used for the Green Infrastructure and the Health and Wellbeing Influences on an Ageing Population (GHIA) study: a study evaluating the impact of improvements in urban green space on older adults' physical activity and wellbeing. Observations for this study were conducted at four set observation periods per day: morning (10-11am), lunchtime (12-1pm), afternoon (3-4pm), and evening (5-6pm). These times were found to capture the biggest variation in older adults' activity across the day in a feasibility study, whilst also providing sufficient time for breaks and possible travel to other sites in between observation periods.

Missing or postponed observation periods

Any missed observations (e.g. due to illness) should be rescheduled for the same day of the next available week. For example, if a 3-4pm observation period is missed on Monday, it should be made up on the next available Monday at 3-4pm.

Example observation period

9:45am - Check target area, prepare data forms and rate the extent of any incivilities in the target area

10am - Start observing in target area

11am - Stop observing in target area

11:01am - Transfer data from the MOHAWk observation form onto the MOHAWk summary form

6. DATA ANALYSIS

Summarising data

Observers should use the MOHAWk summary form to summarise overall counts of gender, ethnicity, age group, physical activity level, whether they engaged in social interaction or took notice of the environment, activity type and whether mobility assistance was required for all individuals.

The MOHAWk summary form currently links age group with the characteristics and activity of each individual. However, researchers can amend this form depending on the aims of the evaluation e.g. researchers can edit the form to link gender and activity level, for instance, if they specifically want to know how many males and females engage in vigorous activity.

Sensitivity analysis

MOHAWk requires observers to carry on observing regardless of weather conditions, unless weather conditions become so extreme that they compromise the observer's safety. Therefore, to control for the potential bias associated with weather, it is recommended that a sensitivity analysis is carried out to assess the impact of weather (specifically precipitation).

The duration of any precipitation that occurs during an observation period is recorded by the observer. It is recommended that observation periods are removed for the sensitivity analysis if the accumulated duration of any precipitation lasts for 50% or more of the observation period i.e. 30 minutes or more. Alternatively, weather can be included as a covariate in the analysis.

APPENDIX 1 – MOHAWk observation form

START TIME: ________________ END TIME: ________________

DATE / DAY: ____ SITE: ____ OBSERVER: _____________

WEATHER / COMMENTS: ____

Note any key observations. For weather, include the duration of any precipitation e.g. 'Rain from 10.20-10.45am'

Person	Gender		Age Group					Ethnicity		Activity Type(s)				Activity Level(s)					Mobility assistance ?	Group (record head count in group)	Notes
	Female	Male	Infant	Child	Teen	Adult	Older Adult	White	Non-White	Cycling	Using phone	Dog walking	……….	S	W	V	TN	C			
1																					
2																					
3																					
4																					
5																					
6																					
7																					
8																					
9																					
10																					
11																					

Person	Gender		Age Group					Ethnicity		Activity Type(s)				Activity Level(s)					Mobility assistance ?	Group (record head count in group)	Notes
	Female	Male	Infant	Child	Teen	Adult	Older Adult	White	Non-White	Cycling	Using phone	Dog walking	………	S	W	V	TN	C			
12																					
13																					
14																					
15																					
16																					
17																					
18																					
19																					
20																					
21																					
22																					
23																					
24																					
25																					
26																					
27																					
28																					
29																					
30																					
31																					
32																					
33																					
34																					
35																					
36																					

Person	Gender		Age Group					Ethnicity		Activity Type(s)				Activity Level(s)					Mobility assistance ?	Group (record head count in group)	Notes
	Female	Male	Infant	Child	Teen	Adult	Older Adult	White	Non-White	Cycling	Using phone	Dog walking	……….	S	W	V	TN	C			
37																					
38																					
39																					
40																					
41																					
42																					
43																					
44																					
45																					
46																					
47																					
48																					
49																					
50																					
51																					
52																					
53																					
54																					
55																					
56																					
57																					
58																					
59																					
60																					
61																					

Person	Gender		Age Group					Ethnicity		Activity Type(s)				Activity Level(s)					Mobility assistance ?	Group (record head count in group)	Notes
	Female	Male	Infant	Child	Teen	Adult	Older Adult	White	Non-White	Cycling	Using phone	Dog walking	……….	S	W	V	TN	C			
62																					
63																					
64																					
65																					
66																					
67																					
68																					
69																					
70																					
71																					
72																					
73																					
74																					
75																					
76																					
77																					
78																					
79																					
80																					
81																					
82																					
83																					
84																					
85																					
86																					

Person	Gender		Age Group					Ethnicity		Activity Type(s)				Activity Level(s)					Mobility assistance ?	Group (record head count in group)	Notes
	Female	Male	Infant	Child	Teen	Adult	Older Adult	White	Non-White	Cycling	Using phone	Dog walking	……….	S	W	V	TN	C			
87																					
88																					
89																					
90																					
91																					
92																					
93																					
94																					
95																					
96																					
97																					
98																					
99																					
100																					
101																					
102																					
103																					
104																					
105																					
106																					
107																					
108																					
109																					
110																					
111																					

Person	Gender		Age Group					Ethnicity		Activity Type(s)				Activity Level(s)					Mobility assistance ?	Group (record head count in group)	Notes
	Female	Male	Infant	Child	Teen	Adult	Older Adult	White	Non-White	Cycling	Using phone	Dog walking	……….	S	W	V	TN	C			
112																					
113																					
114																					
115																					
116																					
117																					
118																					
119																					
120																					
121																					
122																					
123																					
124																					
125																					
126																					
127																					
128																					
129																					
130																					
131																					
132																					
133																					
134																					
135																					
136																					

Person	Gender		Age Group					Ethnicity		Activity Type(s)				Activity Level(s)					Mobility assistance ?	Group (record head count in group)	Notes
	Female	Male	Infant	Child	Teen	Adult	Older Adult	White	Non-White	Cycling	Using phone	Dog walking	……….	S	W	V	TN	C			
137																					
138																					
139																					
140																					
141																					
142																					
143																					
144																					
145																					
146																					
147																					
148																					
149																					
150																					
151																					
152																					
153																					
154																					
155																					
156																					
157																					
158																					
159																					
160																					
161																					

APPENDIX 2 – MOHAWk summary form

Data summary form

START TIME: ______________ **END TIME:** ______________

DATE / DAY: ______________________________ **SITE:** ______________________________ **OBSERVER:** ____________

WEATHER/ COMMENTS: ____

Note any key observations. For weather, include any precipitation (type, intensity, duration) e.g., light rain from 10.20-10.45am

INCIVILITIES IN TARGET AREA	None	Hardly Noticeable	Noticeable	Very Noticeable
General litter	☐	☐	☐	☐
Evidence of alcohol use (empty bottles/ cans)	☐	☐	☐	☐
Evidence of drug taking (e.g., needles, syringes)	☐	☐	☐	☐
Graffiti	☐	☐	☐	☐

	None	Hardly Noticeable	Noticeable	Very Noticeable
Broken Glass	☐	☐	☐	☐
Vandalism	☐	☐	☐	☐
Dog mess	☐	☐	☐	☐
Noise	☐	☐	☐	☐

AGE GROUP	GENDER		ETHNICITY		ACTIVITY LEVEL					MOBILITY ASSISTANCE?
	Female	Male	White	Non-White	S	W	V	TN	C	
Infant										
Child										
Teen										
Adult										
Older Adult										
Group / Busy										
	Frequencies by age group: Infant:	Child:		Teen:		Adult:		Older Adult:		

ACTIVITY TYPE	FREQUENCY
Cycling	
Using phone	
Dog walking	

APPENDIX 3 – Example of a completed MOHAWK observation form

Data collection form **START TIME:** 10am **END TIME:** 11am **Page** 1

DATE / DAY: Monday 8th January **SITE:** Intervention site 1 (Boothstown) **OBSERVER:** JB

WEATHER / COMMENTS: Rain: 10-10.15, 10.35-10.50

Note any key observations. For weather, include the duration of any precipitation e.g., 'Rain from 10.20-10.45am'

Person	Gender		Age Group					Ethnicity		Activity Type(s)				Activity Level(s)					Mobility assistance ?	Group (record head count in group)	Notes
	Female	Male	Infant	Child	Teen	Adult	Older Adult	White	Non-White	Cycling	Using phone	Dog walking	Fishing	S	W	V	TN	C			
1	✓					✓		✓			✓				✓						
2		✓				✓		✓							✓			✓	✓		
3		✓		✓				✓							✓			✓			
4		✓			✓			✓		✓					✓	✓					
5 (↑10.15am)	✓						✓		✓			✓			✓		✓	✓	✓		
6		✓					✓		✓			✓			✓			✓			
7		✓				✓		✓							✓						
8	✓					✓		✓			✓			✓	✓						Red cap
9		✓					✓	✓							✓						
10	INFANT		✓																		
11 (↑10.30am)		✓				✓		✓		✓		✓				✓					
12	✓					✓			✓			✓			✓			✓			
13	✓						✓		✓						✓			✓			
14	8	10		16		2		15	3						18	1	2	18		18	
15		✓				✓		✓							✓						
16 (↑10.45am)		✓				✓		✓			✓	✓			✓		✓				
17	✓				✓			✓							✓						
18		✓					✓		✓						✓			✓			
19		✓		✓					✓						✓			✓			
20		✓				✓		✓					✓	✓	✓		✓	✓			Blue hat
21		✓				✓		✓					✓	✓	✓		✓	✓			Red t-shirt
22		✓				✓		✓					✓	✓	✓		✓	✓			Long hair

APPENDIX 4 – Example of a completed MOHAWK summary form

Data summary form **START TIME:** 10am **END TIME:** 11am

DATE / DAY: Monday 8th January **SITE:** Intervention site 1 (Boothstown) **OBSERVER:** JB

WEATHER/ COMMENTS: Rain: 10 - 10.15, 10.35 - 10.50

Note any key observations. For weather, include any precipitation (type, intensity, duration) e.g., light rain from 10.20-10.45am

INCIVILITIES IN TARGET AREA

	None	Hardly Noticeable	Noticeable	Very Noticeable
General litter	☐	☑	☐	☐
Evidence of alcohol use (empty bottles/ cans)	☑	☐	☐	☐
Evidence of drug taking (e.g., needles, syringes)	☑	☐	☐	☐
Graffiti	☐	☐	☐	☑

	None	Hardly Noticeable	Noticeable	Very Noticeable
Broken Glass	☑	☐	☐	☐
Vandalism	☑	☐	☐	☐
Dog mess	☑	☐	☐	☐
Noise	☐	☑	☐	☐

AGE GROUP	GENDER		ETHNICITY		ACTIVITY LEVEL					MOBILITY ASSISTANCE?
	Female	Male	White	Non-White	S	W	V	TN	C	
Infant	1									
Child	-	2	1	1	-	2	-	-	2	-
Teen	1	1	2	-	-	2	1	-	-	-
Adult	3	8	10	1	4	10	1	4	5	1
Older Adult	2	3	1	4	-	5	-	1	5	1
Group / Busy	8	10	15	3	-	18	1	2	18	-

Frequencies by age group

Infant: – Child: 16 Teen: – Adult: 2 Older Adult: –

ACTIVITY TYPE	FREQUENCY
Cycling	2
Using phone	3
Dog walking	5
Fishing	3

Appendix B. MOHAWk standardised observation form

MOHAWk standardised observation form **START TIME: ___________________** **END TIME: ___________________**

DATE / DAY: ____ **SITE: ____** **OBSERVER: ______________**

WEATHER / COMMENTS: ____

Include the duration of any precipitation e.g. 'Rain from 10.20-10.45am. Also make a note of any other potentially important observations.

Person	Gender		Age Group					Ethnicity		Activity Type(s)				Activity Level(s)					Mobility assistance ?	Group (record head count in group)	Notes
	Female	Male	Infant	Child	Teen	Adult	Older Adult	White	Non-White	Cycling	Using phone	Dog walking	………	S	W	V	TN	C			
1																					
2																					
3																					
4																					
5																					
6																					
7																					
8																					
9																					
10																					
11																					
12																					

Person	Gender		Age Group					Ethnicity		Activity Type(s)				Activity Level(s)					Mobility assistance ?	Group (record head count in group)	Notes
	Female	Male	Infant	Child	Teen	Adult	Older Adult	White	Non-White	Cycling	Using phone	Dog walking	………	S	W	V	TN	C			
13																					
14																					
15																					
16																					
17																					
18																					
19																					
20																					
21																					
22																					
23																					
24																					
25																					
26																					
27																					
28																					
29																					
30																					
31																					
32																					
33																					
34																					
35																					
36																					
37																					

Person	Gender		Age Group					Ethnicity		Activity Type(s)				Activity Level(s)					Mobility assistance ?	Group (record head count in group)	Notes
	Female	Male	Infant	Child	Teen	Adult	Older Adult	White	Non-White	Cycling	Using phone	Dog walking	……….	S	W	V	TN	C			
38																					
39																					
40																					
41																					
42																					
43																					
44																					
45																					
46																					
47																					
48																					
49																					
50																					
51																					
52																					
53																					
54																					
55																					
56																					
57																					
58																					
59																					
60																					
61																					
62																					

Person	Gender		Age Group					Ethnicity		Activity Type(s)				Activity Level(s)					Mobility assistance ?	Group (record head count in group)	Notes
	Female	Male	Infant	Child	Teen	Adult	Older Adult	White	Non-White	Cycling	Using phone	Dog walking	………	S	W	V	TN	C			
63																					
64																					
65																					
66																					
67																					
68																					
69																					
70																					
71																					
72																					
73																					
74																					
75																					
76																					
77																					
78																					
79																					
80																					
81																					
82																					
83																					
84																					
85																					
86																					
87																					

Person	Gender		Age Group					Ethnicity		Activity Type(s)				Activity Level(s)					Mobility assistance ?	Group (record head count in group)	Notes
	Female	Male	Infant	Child	Teen	Adult	Older Adult	White	Non-White	Cycling	Using phone	Dog walking	………	S	W	V	TN	C			
88																					
89																					
90																					
91																					
92																					
93																					
94																					
95																					
96																					
97																					
98																					
99																					
100																					
101																					
102																					
103																					
104																					
105																					
106																					
107																					
108																					
109																					
110																					
111																					
112																					

Person	Gender		Age Group					Ethnicity		Activity Type(s)				Activity Level(s)					Mobility assistance ?	Group (record head count in group)	Notes
	Female	Male	Infant	Child	Teen	Adult	Older Adult	White	Non-White	Cycling	Using phone	Dog walking	……….	S	W	V	TN	C			
113																					
114																					
115																					
116																					
117																					
118																					
119																					
120																					
121																					
122																					
123																					
124																					
125																					
126																					
127																					
128																					
129																					
130																					
131																					
132																					
133																					
134																					
135																					
136																					
137																					

Person	Gender		Age Group					Ethnicity		Activity Type(s)				Activity Level(s)					Mobility assistance ?	Group (record head count in group)	Notes
	Female	Male	Infant	Child	Teen	Adult	Older Adult	White	Non-White	Cycling	Using phone	Dog walking	………	S	W	V	TN	C			
138																					
139																					
140																					
141																					
142																					
143																					
144																					
145																					
146																					
147																					
148																					
149																					
150																					
151																					
152																					
153																					
154																					
155																					
156																					
157																					
158																					
159																					
160																					
161																					
162																					

Appendix C. MOHAWk data summary form

MOHAWk data summary form **START TIME: ________________** **END TIME: ________________**

DATE / DAY: ____ **SITE: ____________________________________** **OBSERVER: _____________**

WEATHER/ COMMENTS: ____

Include the duration of any precipitation e.g. 'Rain from 10.20-10.45am. Also make a note of any other potentially important observations.

INCIVILITIES IN TARGET AREA	None	Hardly Noticeable	Noticeable	Very Noticeable
General litter	☐	☐	☐	☐
Evidence of alcohol use (empty bottles/ cans)	☐	☐	☐	☐
Evidence of drug taking (e.g., needles, syringes)	☐	☐	☐	☐
Graffiti	☐	☐	☐	☐

	None	Hardly Noticeable	Noticeable	Very Noticeable
Broken Glass	☐	☐	☐	☐
Vandalism	☐	☐	☐	☐
Dog mess	☐	☐	☐	☐
Noise	☐	☐	☐	☐

AGE GROUP	GENDER		ETHNICITY		ACTIVITY LEVEL					MOBILITY ASSISTANCE?
	Female	Male	White	Non-White	S	W	V	TN	C	
Infant										
Child										
Teen										
Adult										
Older Adult										
Group / Busy										
	Frequencies by age group									
	Infant:		Child:		Teen:		Adult:		Older Adult:	

ACTIVITY TYPE	FREQUENCY
Cycling	
Using phone	
Dog walking	

Appendix D. Key characteristics and descriptive statistics for three separate natural experimental studies (baseline data only)

Study	Site	Type of site	Time of year	Days of observation (total hours)	Mean count of each behaviour per hour (SD) and total count of behaviours observed per site, *N* [c]				
					Sedentary	Walking	Vigorous	Take Notice	Connect
Benton et al. [25] [a]	Burton Road, Manchester, UK, M20 3GB (Intervention site 1)	Residential street	September 2017	Monday, Wednesday (8 hours)	0.5 (0.76) *N* = 4	100.25 (36.20) *N* = 802	26.63 (18.56) *N* = 213	0.13 (0.35) *N* = 1	29 (16.20) *N* = 232
	Burnage Lane, Manchester, UK, M19 1EN (Comparison site 1A)	Residential street	September 2017	Monday, Wednesday (8 hours)	2.88 (1.89) *N* = 23	81.38 (68.64) *N* = 651	14.5 (13.18) *N* = 116	1 (2.07) *N* = 8	36 (43.81) *N* = 288
	Hempshaw Lane, Stockport, UK, SK2 6DS (Comparison site 1B)	Residential street	September 2017	Monday, Wednesday (8 hours)	2.38 (2.67) *N* = 19	69.75 (41.30) *N* = 558	8.88 (7.41) *N* = 71	0.13 (0.35) *N* = 1	24.75 (25.46) *N* = 198
	Parbold Avenue, Manchester, UK, M20 1FU (Intervention site 2)	Residential street	September 2017	Thursday, Friday (8 hours)	0.75 (1.49) *N* = 6	27.38 (8.23) *N* = 219	2.5 (1.85) *N* = 20	0.25 (0.46) *N* = 2	5.88 (3.48) *N* = 47

Study	Site	Type of site	Time of year	Days of observation (total hours)	Mean count of each behaviour per hour (SD) and total count of behaviours observed per site, N^c				
					Sedentary	Walking	Vigorous	Take Notice	Connect
	Westcroft Road, Manchester, UK, M20 6FE (Comparison site 2A)	Residential street	September 2017	Thursday, Friday (8 hours)	0.63 (0.74) *N* = 5	24.88 (4.19) *N* = 199	5.13 (2.80) *N* = 41	1.13 (3.18) *N* = 9	5.88 (4.12) *N* = 47
	Leech Avenue, Ashton-under-Lyne, UK, OL6 8HH (Comparison site 2B)	Residential street	September 2017	Thursday, Friday (8 hours)	2.88 (3.40) *N* = 23	20.63 (5.97) *N* = 165	2.38 (1.92) *N* = 19	0.5 (1.07) *N* = 4	9.13 (5.08) *N* = 73
	Dennison Avenue, Manchester, UK, M20 1GF (Intervention site 3)	Residential street	September 2017	Tuesday, Friday (8 hours)	0.75 (1.49) *N* = 6	20.25 (9.84) *N* = 162	3.13 (2.53) *N* = 25	0 (0) *N* = 0	7.38 (6.72) *N* = 59
	Wynyard Road, Manchester, UK, M22 9PS (Comparison site 3A)	Residential street	September 2017	Tuesday, Friday (8 hours)	0.88 (2.1) *N* = 7	14.25 (4.23) *N* = 114	2.5 (1.77) *N* = 20	0.13 (0.35) *N* = 1	7.75 (5.04) *N* = 62
	Hadfield Crescent, Ashton-under-Lyne, UK, OL6 8HW (Comparison site 3B)	Residential street	September 2017	Tuesday, Friday (8 hours)	2 (1.77) *N* = 16	26.5 (6.78) *N* = 212	0.5 (0.76) *N* = 4	0.63 (0.74) *N* = 5	6.63 (3.81) *N* = 53
	Alford Avenue, Manchester, UK, M20 1AQ (Intervention site 4)	Residential street	September 2017	Tuesday, Thursday (8 hours)	1.63 (1.85) *N* = 13	9.5 (3.25) *N* = 76	0.38 (0.52) *N* = 3	0 (0) *N* = 0	2.25 (2.25) *N* = 18

Study	Site	Type of site	Time of year	Days of observation (total hours)	Mean count of each behaviour per hour (SD) and total count of behaviours observed per site, *N* [c]				
					Sedentary	Walking	Vigorous	Take Notice	Connect
	Pembury Close, Manchester, UK, M22 9PU (Comparison site 4A)	Residential street	September 2017	Tuesday, Thursday (8 hours)	0.13 (0.35) *N* = 1	8.13 (5.06) *N* = 65	0.88 (1.46) *N* = 7	0 (0) *N* = 0	2.75 (2.87) *N* = 22
Benton et al. [28] [a]	Moss House Lane, Salford, UK, M28 1JD (Intervention site 1)	Green corridor	November 2017	Tuesday, Wednesday, Saturday (12 hours)	0.17 (0.39) *N* = 2	4 (3.13) *N* = 48	0.25 (0.45) *N* = 3	0.25 (0.45) *N* = 3	1.42 (1.73) *N* = 17
	Rigby Lane, Bolton, UK, BL2 3EQ (Comparison site 1A)	Green corridor	November 2017	Tuesday, Wednesday, Saturday (12 hours)	0 (0) *N* = 0	2.5 (1.38) *N* = 30	0 (0) *N* = 0	0 (0) *N* = 0	1.08 (1.31) *N* = 13
	Common Lane, Wigan, UK, WN7 4QP (Comparison site 1B)	Green corridor	November 2017	Tuesday, Wednesday, Saturday (12 hours)	0.33 (0.65) *N* = 4	2.83 (1.75) *N* = 34	0.08 (0.29) *N* = 1	0.42 (0.67) *N* = 9	0.92 (1.08) *N* = 11
Anderson et al. [29] [b]	Bennett Street, Manchester, UK, M12 5AU (Intervention site 1)	Park	August/ September 2018	Wednesday, Thursday (8 hours)	3 (2.07) *N* = 24	15.88 (8.59) *N* = 127	4.5 (3.21) *N* = 36	2.13 (1.55) *N* = 17	5.25 (4.53) *N* = 42

Study	Site	Type of site	Time of year	Days of observation (total hours)	Mean count of each behaviour per hour (SD) and total count of behaviours observed per site, N^c				
					Sedentary	Walking	Vigorous	Take Notice	Connect
	Graythorpe Walk, Salford, UK, M5 4HL (Comparison site 1A)	Park	August/ September 2018	Wednesday, Thursday (8 hours)	10.88 (4.22) $N = 87$	37.5 (15.81) $N = 300$	7.88 (8.77) $N = 63$	2.13 (2.17) $N = 17$	17 (13.69) $N = 136$
	Buckingham Street, Salford, UK, M5 4FD (Comparison site 1B)	Park	August/ September 2018	Wednesday, Thursday (8 hours)	5.25 (6.48) $N = 42$	22.13 (10.68) $N = 177$	3 (4.31) $N = 24$	2.25 (2.61) $N = 18$	8.75 (7.11) $N = 70$
	Wenlock Way, Manchester, UK, M12 5TS (Intervention site 2)	Brownfield site	August/ September 2018	Wednesday, Saturday (8 hours)	4.25 (4.23) $N = 34$	32.13 (16.26) $N = 257$	6 (3.67) $N = 48$	1.75 (1.91) $N = 14$	11.75 (9.32) $N = 94$
	Rostron Avenue, Manchester, UK, M12 5NP (Comparison site 2A)	Brownfield site	August/ September 2018	Wednesday, Saturday (8 hours)	11 (4.78) $N = 88$	47.38 (13.03) $N = 379$	5.63 (2.88) $N = 45$	1 (1.20) $N = 8$	19 (10.58) $N = 152$
	Reabrook Avenue, Manchester, UK, M12 5LL (Comparison site 2B)	Brownfield site	August/ September 2018	Wednesday, Saturday (8 hours)	6.13 (6.42) $N = 49$	30.25 (28.18) $N = 242$	3.75 (2.38) $N = 30$	2.38 (3.58) $N = 19$	12.63 (14.55) $N = 101$

[a] Mean counts and standard deviations for each behaviour are based on the sum of adults and older adults;
[b] Mean counts and standard deviations for each behaviour are based on the sum of children, teens, adults and older adults;
[c] Each observed person can engage in more than one behaviour per observation period e.g. a person who is walking and talking to a friend would be coded as ‘Walking’ and ‘Connect’;
SD Standard deviation

Appendix E. A list of key refinements to the original MOHAWk

Original MOHAWk	Refinement
Age categories were split into 10-24 year olds, 25-44 year olds, 45-64 year olds and 65 + year olds.	Changed age categories to Infants (babies/ toddlers in a pram), Children (0-12 years old), Teens (13-19 years old), Adults (20-64 years old) and Older Adults (65 + years old).
Age and gender were the characteristics recorded for each observed person.	Added ethnicity (White or Non-white) and mobility assistance required (Yes or No) to the characteristics recorded for each observed person.
Be Active was coded as any physical activity.	Redefined Be Active as three physical activity levels, similar to previous observation tools [7–10]: Sedentary, Walking, Vigorous.
All types of activities that each person engaged in were recorded.	Added key predetermined activity types (Cycling, Using phone, Dog walking) rather than recording all types of behaviours. Researchers can edit these predetermined activity types.
No procedures in place for counting large groups or coping with busy periods.	Contingency procedures put in place to reduce the risk of missing data for large groups or when observing during busy periods.
Instances of social incivilities in the target area were recorded by the observer (e.g. anti-social behaviour).	The following incivilities in the target area are recorded using a checklist, based on an existing validated tool for measuring the quality of neighbourhood green space [15]: general litter, evidence of alcohol use, evidence of drug taking, graffiti, broken glass, vandalism, dog mess, noise.
No formal procedure for testing the potential confounding effect of precipitation.	Observers record the duration of any precipitation that occurs during an observation period. A sensitivity analysis controls for the confounding influence of weather (or by including weather as a covariate): observation periods are removed for the sensitivity analysis if the accumulated duration of any precipitation lasts for 50% or more of the observation period i.e. 30 minutes or more.

Appendix F. Inter-rater reliabilities across the three studies

Variable	Study	ICC (95% CI)	Number of observation periods	Total count (Observer 1)	Total count (Observer 2)
Total number of people	Study 2	0.98 (0.98 – 0.99)	80	2,403	2,337
	Study 3	0.99 (0.93 – 0.99)	13	174	172
Infant	Study 2	0.91 (0.86 – 0.94)	80	61	65
	Study 3	1 (1)	13	4	4
Child	Study 2	0.65 (0.50 – 0.76)	80	250	215
	Study 3	0.97 (0.90 – 0.99)	13	31	31
Teen	Study 2	0.89 (0.83 – 0.93)	80	166	178
	Study 3	0.74 (0.34 – 0.91)	13	8	12
Adult	Study 2	0.99 (0.98 – 0.99)	80	1,682	1,615
	Study 3	0.97 (0.90 – 0.99)	13	129	122
Older Adult	Study 2	0.84 (0.76 – 0.89)	80	244	264
	Study 3	0.77 (0.40 – 0.92)	13	2	3
Female	Study 2	0.98 (0.96 – 0.99)	80	1,057	1,056
	Study 3	0.95 (0.84 – 0.98)	13	87	88
Male	Study 2	(0.98 – 0.99)	80	1,303	1,270
	Study 3	(0.73 – 0.97)	13	85	79
White	Study 2	0.97 (0.96 – 0.98)	80	1,926	1,982
	Study 3	0.98 (0.93 – 0.99)	13	171	166

Non-white	Study 2	0.83 (0.75 – 0.89)	80	434	406
	Study 3	1 (1)	13	1	1
Sedentary	Study 1	Observer pair 1:			
		0.73 (0.50 – 0.86)	30	72	66
	Study 2	Observer pair 2:			
		0.80 (0.62 – 0.90)	32	93	88
		0.84 (0.76 – 0.89)	80	45	32
	Study 3	0.90 (0.70 – 0.97)	13	39	34
Walking	Study 1	Observer pair 1:			
		0.97 (0.94 – 0.99)	30	1,207	1,119
	Study 2	Observer pair 2:			
		0.998 (0.996 – 0.999)	32	1,494	1,526
		0.98 (0.97 – 0.99)	80	1,792	1,805
	Study 3	0.90 (0.92 – 0.99)	13	141	136
Vigorous	Study 1	Observer pair 1:			
		0.94 (0.87 – 0.97)	30	40	44
	Study 2	Observer pair 2:			
		0.99 (0.99 – 0.996)	32	47	46
		0.98 (0.97 – 0.99)	80	541	531
	Study 3	0.92 (0.76 – 0.98)	13	34	35
Take Notice	Study 1	Observer pair 1:			
		0.77 (0.57 – 0.88)	30	135	90

	Study 2	Observer pair 2:			
		0.87 (0.74 – 0.93)	32	153	123
		0.82 (0.73 – 0.88)	80	16	14
	Study 3	0.78 (0.43 – 0.93)	13	17	14
Connect	Study 1	Observer pair 1:			
		0.94 (0.97 – 0.99)	30	373	337
	Study 2	Observer pair 2:			
		0.99 (0.97 – 0.99)	32	428	422
		0.96 (0.94 – 0.98)	80	492	518
	Study 3	0.81 (0.48 – 0.94)	13	69	64
Mobility assistance required	Study 2	0.88 (0.81 – 0.92)	80	21	25
	Study 3	1 (1)	13	0	0

Appendix G. Comparison site matching process

The overall aim of the comparison site matching process was to identify two closely matched comparison sites for each intervention site, using several key objective and subjective environmental variables.

Step one: Matching at the neighbourhood (LSOA) level

Each intervention site is located along a residential street, so the overall aim of the process was to identify the most closely matched streets based on key variables; comparison sites could then be identified within these streets. The first step was to identify the most closely matched neighbourhoods in which the potential streets for comparison sites could be searched from.

Due to a lack of available walkability indices for Greater Manchester (GM), it was necessary to manually search for neighbourhoods based on available spatial data at the Lower Layer Super Output Area (LSOA) level (census reporting units containing between 1000 and 3000 individuals) [1,2]. Population density, defined as the number of persons per hectare, was used as a proxy measure of residential density. Street connectivity was measured using street intersection density; the number of 3-way junctions standardised by LSOA area. Normalised Difference Vegetation Index (NDVI) scores, a normalised scale of healthy vegetation cover [3], were used for presence of greenery. Socioeconomic status was measured using the Index of Multiple Deprivation (IMD) Score [4]; an area deprivation score that combines several indicators of deprivation including income, employment, health and crime. Spatial analyses were carried out using ArcGIS 10.4.

To identify the most closely matched LSOAs to the intervention LSOA, a systematic funnelling approach was used. All LSOAs in GM were firstly ranked in order of residential density and the 100 most closely matched LSOAs to each intervention site LSOA were extracted. These 100 LSOAs were then ranked in order of street connectivity and the 50 most closely matched LSOAs were selected. These two variables were matched first because they are the strongest and most consistent correlates of older adults' physical activity [5,6]. Next, the remaining 50 LSOAs were ranked in order of closeness of IMD score to the intervention LSOA and the most closely matched 25 LSOAs were selected. Finally, these 25 LSOAs were ranked in order of those most closely matched on NDVI scores to the intervention LSOA and the five most closely matched LSOAs were selected. Thus, by the end of this step there were a manageable total of five potential LSOAs for each intervention site.

Step two: Matching at the street level (access to/ availability of destinations and services)

As there were no data available for 'access to/ availability of destinations and services' at the LSOA level, the second step was to measure this variable at the street level. There is a lack of data on walkability scores in GM and manually calculating distances to nearby destinations and services for each street within each LSOA was beyond the scope of this project. Instead, Walk Score (www.walkscore.com) was the most appropriate objective and reliable measure readily available. Walk Score is a website that uses a Google search algorithm to calculate a weighted score (1-100) based on the number and accessibility of amenities (such as shops and parks) within a 1-mile radius of a user-entered postcode, whereby closer amenities with the most accessible walking routes are weighted more strongly. Walk Scores have shown good correlation with gold-standard measures of walkability using Geographic Information Systems (GIS) [7and have previously been used in studies in the US [e.g. 8]. Walk Score confirmed that data is available in in the UK (WalkScore, personal communication).

To calculate Walk Scores for each street in each LSOA, it was necessary to obtain all postcodes within each LSOA. Postcodes were obtained using FreeMapTools (www.doogal.co.uk/ukpostcodes.php): a free website that provides postcode data within LSOAs in the UK. Walk Scores were then calculated for each intervention site postcode and all postcodes extracted from the potential comparison site LSOAs. All postcodes were then ranked in order of closeness of Walk Score to each intervention site.

Step three: Matching at the street level (streetscape characteristics)

The aim of step three was to find at least three potential comparison sites located within streets (postcodes) most closely matched to the intervention site based on streetscape characteristics. All postcodes were remotely audited using Google Street View. Google Street View can be accessed via Google Maps (www.maps.google.com) and permits users to remotely navigate 360° through panoramic images of the streetscape environment from the internet. Virtual streetscape audits were preferred given the vast number of postcodes that needed to be examined (between 165 and 212 postcodes for each intervention site). Empirical research has previously demonstrated that Google Street View is mostly a reliable and efficient tool to measure the streetscape in comparison with physical on-site audits [9].

Starting with postcodes most closely matched to the intervention site based on Walk Scores, JB audited the following objective streetscape characteristics associated with older adults' physical activity for each postcode: road type (proxy measure of traffic), number of pedestrian access points (proxy measure of connectivity), presence of a pavement/ footpath, number of benches, presence of street lights, non-residential buildings, and presence of greenery. Presence of greenery, specifically a grass-covered area, was a required characteristic for potential comparison sites because all intervention sites contained a grass-covered area at baseline. For some of the intervention sites, it was necessary to iteratively search for more potential comparison sites; this was done by identifying more LSOAs and repeating the first three steps. By the end of step three, all intervention sites had three potential comparison sites.

Step four: On-site environmental audits

The aim of step four was to ensure potential comparison sites were closely matched in terms of quality and quantity of green space and pavements/ footpaths, and attractiveness of the buildings. These variables are more difficult to reliably judge using Google Street View and therefore required on-site audits of the environment.

JB visited each potential comparison site to systematically audit each site using two validated environmental audit tools: the Neighbourhood Green Space Tool (NGST) [10] and the 54-item abbreviated version of the Microscale Audit of Pedestrian Streetscapes (MAPS-Abbreviated) measure of street design [11]. NGST enabled an audit of the green space characteristics and was specifically developed in the UK; which was important due to the lack of environmental walkability tools that have been developed in the UK [12]. MAPS-Abbreviated was chosen because it measures the pedestrian streetscape (including pavements/ footpaths and buildings) and was partly based on a previous tool that has been modified by the Healthy Aging Network [13]. The scores on both tools in each potential comparison site and corresponding intervention site were used to inform selection of the two final comparison sites.

Step five: Matching on pedestrian traffic

The fifth and final step aimed to ensure that intervention and comparison sites were well matched in terms of pedestrian traffic; that is, the frequency of users passing through a site. It has been found that 15 minutes of observation can provide excellent reliability in estimating the frequency of users passing through a site across the whole hour [Benton et al., in press]. Therefore, 15 minute observations were carried out across the 12 sites (four intervention and eight comparison sites) during August 2017 to count pedestrian traffic. JB conducted 15 minute observations at two intervention sites on a Thursday between 12-2pm, and did the same at the other two intervention sites on a Friday between 12-2pm. Observations of the corresponding comparison sites were then carried out in the following week on Thursday and Friday between 12-2pm. All final comparison sites were, for the most part, similar to corresponding intervention sites in terms of pedestrian traffic and subjective 'feel' of the sites.

References

1. Office for National Statistics. 2001 Census: Digitised Boundary Data (England and Wales) [computer file]. UK Data Service Census Support. Downloaded from: http://edina.ac.uk/census. Licensed under the terms of the Open Government Licence [http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2]
2. Office for National Statistics. Super Output Areas: Office for National Statistics, 2011
3. Markevych I, Schoierer J, Hartig T, Chudnovsky A, Hystad P, Dzhambov AM, De Vries S, Triguero-Mas M, Brauer M, Nieuwenhuijsen MJ, Lupp G. Exploring pathways linking greenspace to health: Theoretical and methodological guidance. Environ Res. 2017;158:301-17.
4. Department for Communities and Local Government. 2015. English Indices of Deprivation 2015 [computer file]. Downloaded from: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015. Licensed under: https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3
5. Barnett DW, Barnett A, Nathan A, Van Cauwenberg J, Cerin E. Built environmental correlates of older adults' total physical activity and walking: a systematic review and meta-analysis. Int J Behav Nutr Phys Act. 2017;14(1):103.
6. Cerin E, Nathan A, Van Cauwenberg J, Barnett DW, Barnett A. The neighbourhood physical environment and active travel in older adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act. 2017;14(1):15.

7. Carr LJ, Dunsiger SI, Marcus BH. Validation of Walk Score for estimating access to walkable amenities. Br J Sports Med. 2011;45(14):1144-8.
8. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature. 2017;547(7663):336-9.
9. Badland HM, Opit S, Witten K, Kearns RA, Mavoa S. Can virtual streetscape audits reliably replace physical streetscape audits?. J Urban Health. 2010;87(6):1007-16.
10. Gidlow CJ, Ellis NJ, Bostock S. Development of the neighbourhood green space tool (NGST). Landsc and Urban Planning. 2012;106(4):347-58.
11. Cain KL, Gavand KA, Conway TL, Geremia CM, Millstein RA, Frank LD, Saelens BE, Adams MA, Glanz K, King AC, Sallis JF. Developing and validating an abbreviated version of the Microscale Audit for Pedestrian Streetscapes (MAPS-Abbreviated). J Transport & Health. 2017; 5: 84-96.
12. Brownson RC, Hoehner CM, Day K, Forsyth A, Sallis JF. Measuring the built environment for physical activity: state of the science. Am J Prev Med. 2009;36(4):S99-123.
13. Kealey M, Kruger J, Hunter R, Ivey S, Satariano W, Bayles C, Ramirez B, Bryant L, Johnson C, Lee C, Levinger D. Engaging older adults to be more active where they live: audit tool development. Prev Chronic Dis. 2005; 2:1–2.
14. Benton JS, Anderson J, Cotterill S, Dennis M, Lindley SJ, French DP. Evaluating the impact of improvements in urban green space on older adults' physical activity and wellbeing: protocol for a natural experimental study. BMC Public Health. 2018;18(1):923.

Appendix H. Target area boundaries for the intervention and comparison sites (Natural experimental study 1)

This document provides maps of the boundaries for the Target Areas in each intervention and comparison site. Each map also shows the primary location for the observer during observation periods and locations where the urban street greening interventions will be implemented at each intervention site.

Key

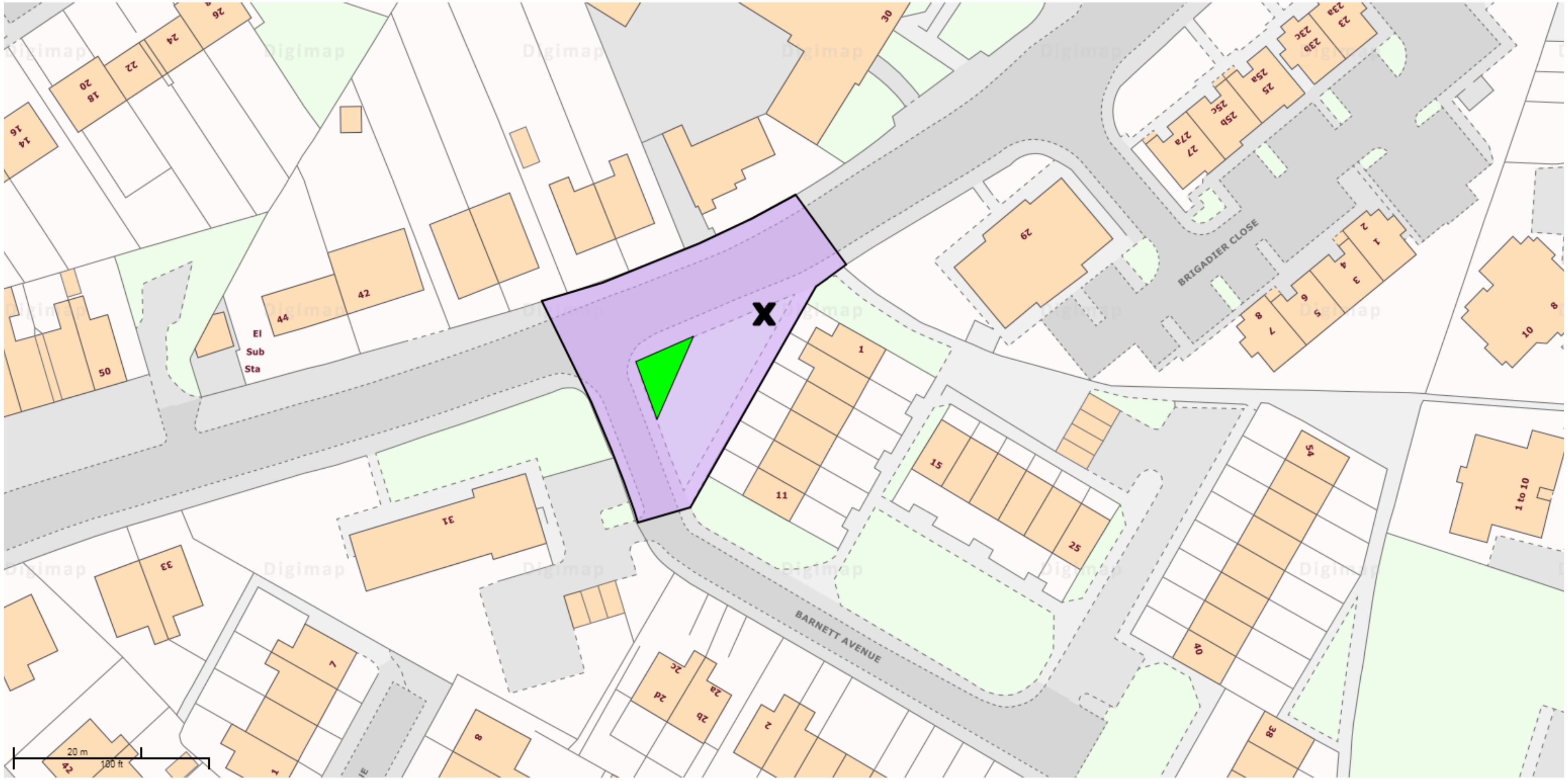
X = Observer location

= Target area boundary

= Location where the urban street greening interventions will be implemented

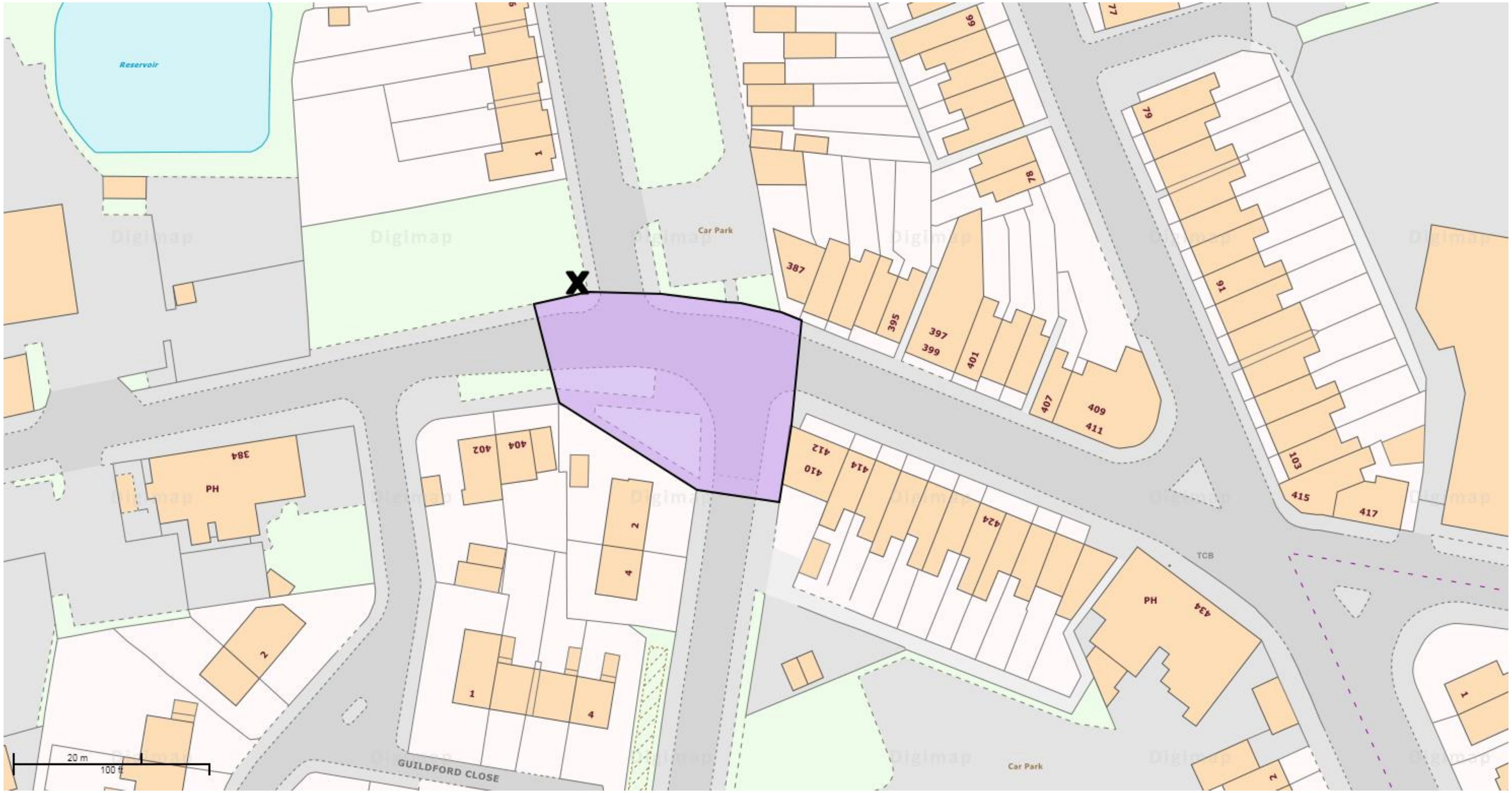
Maps drawn using: www.digimap.edina.ac.uk

Intervention site 1

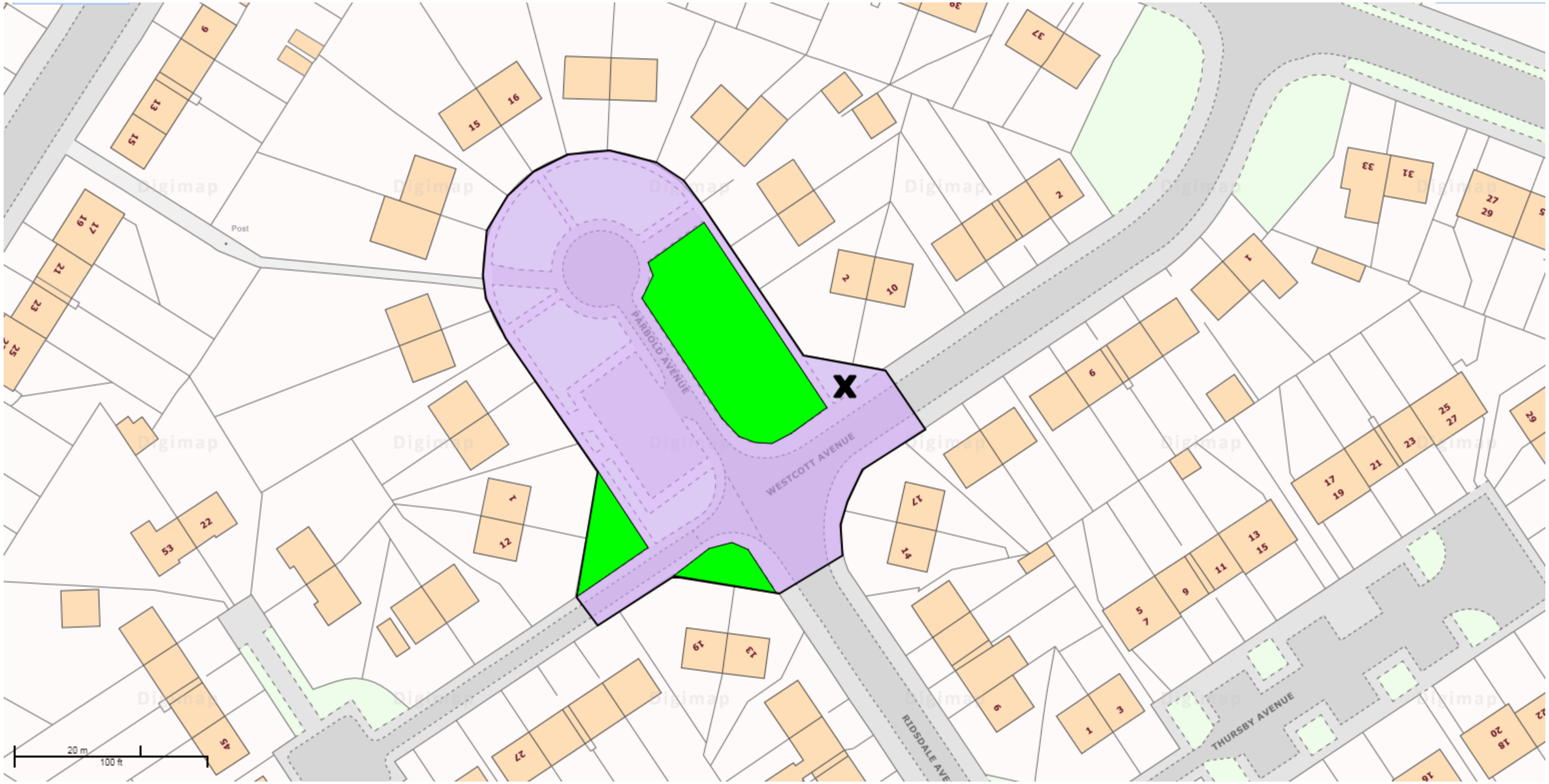


Comparison site 1A

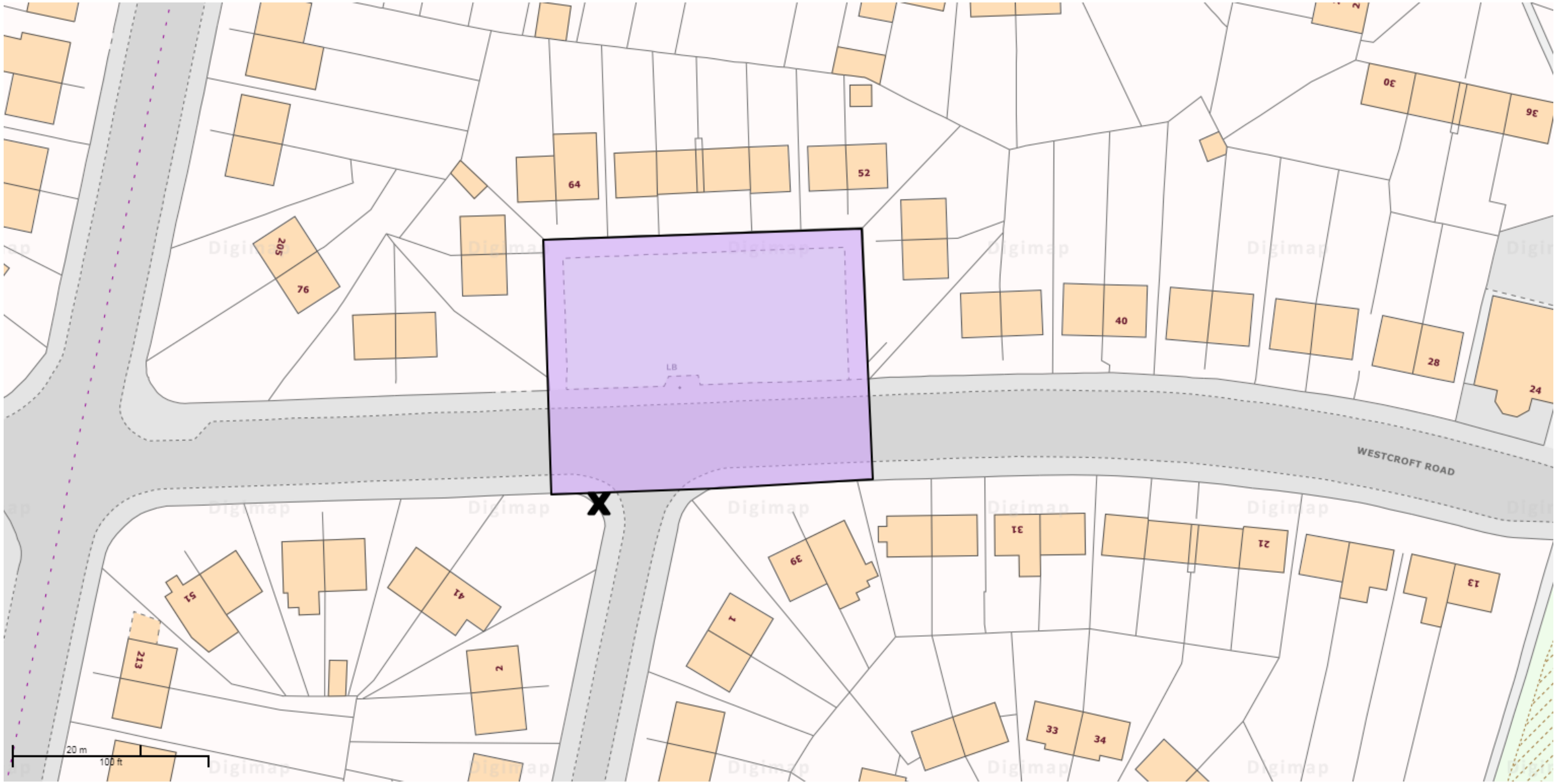
Comparison site 1B



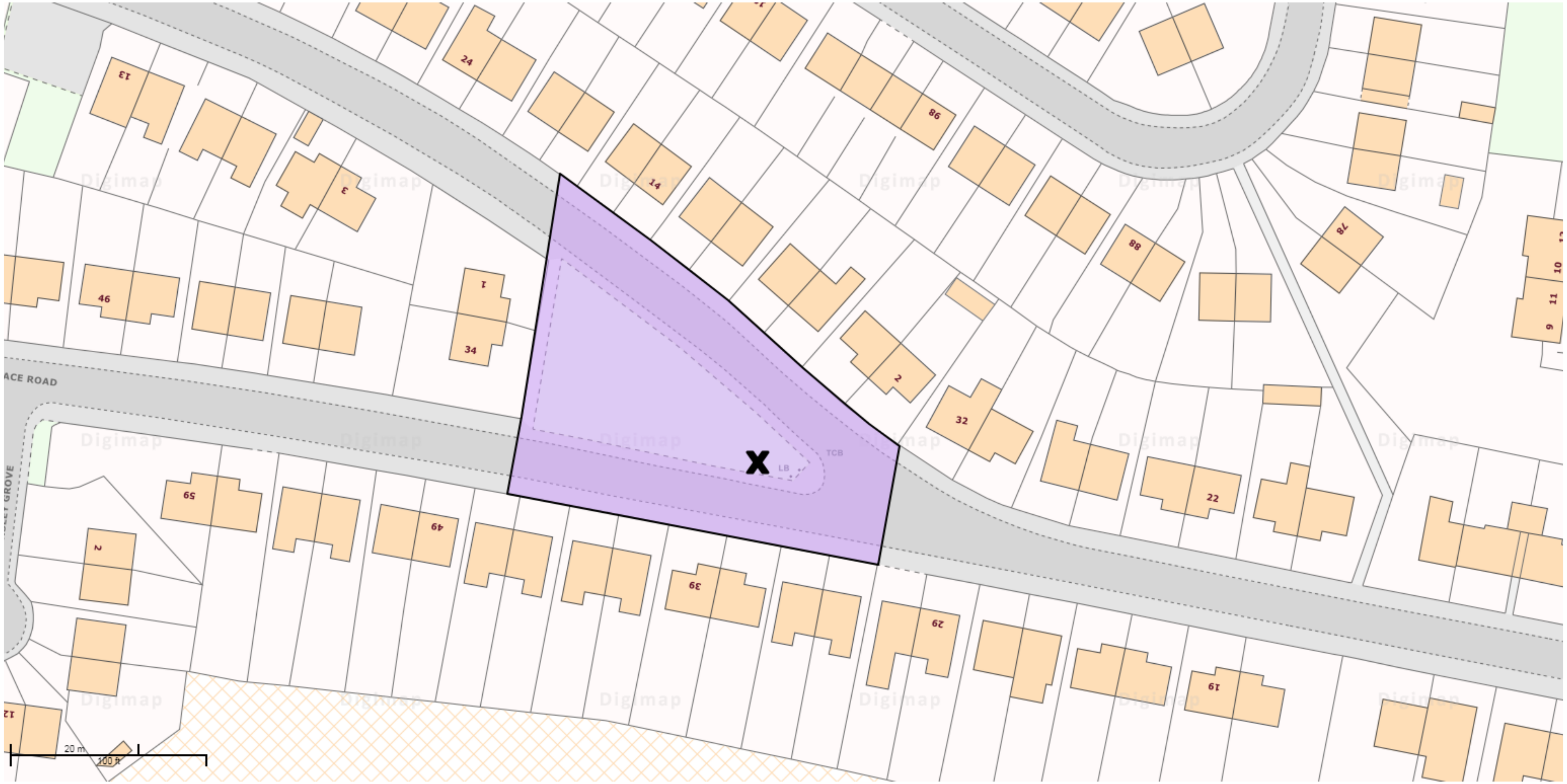
Intervention site 2



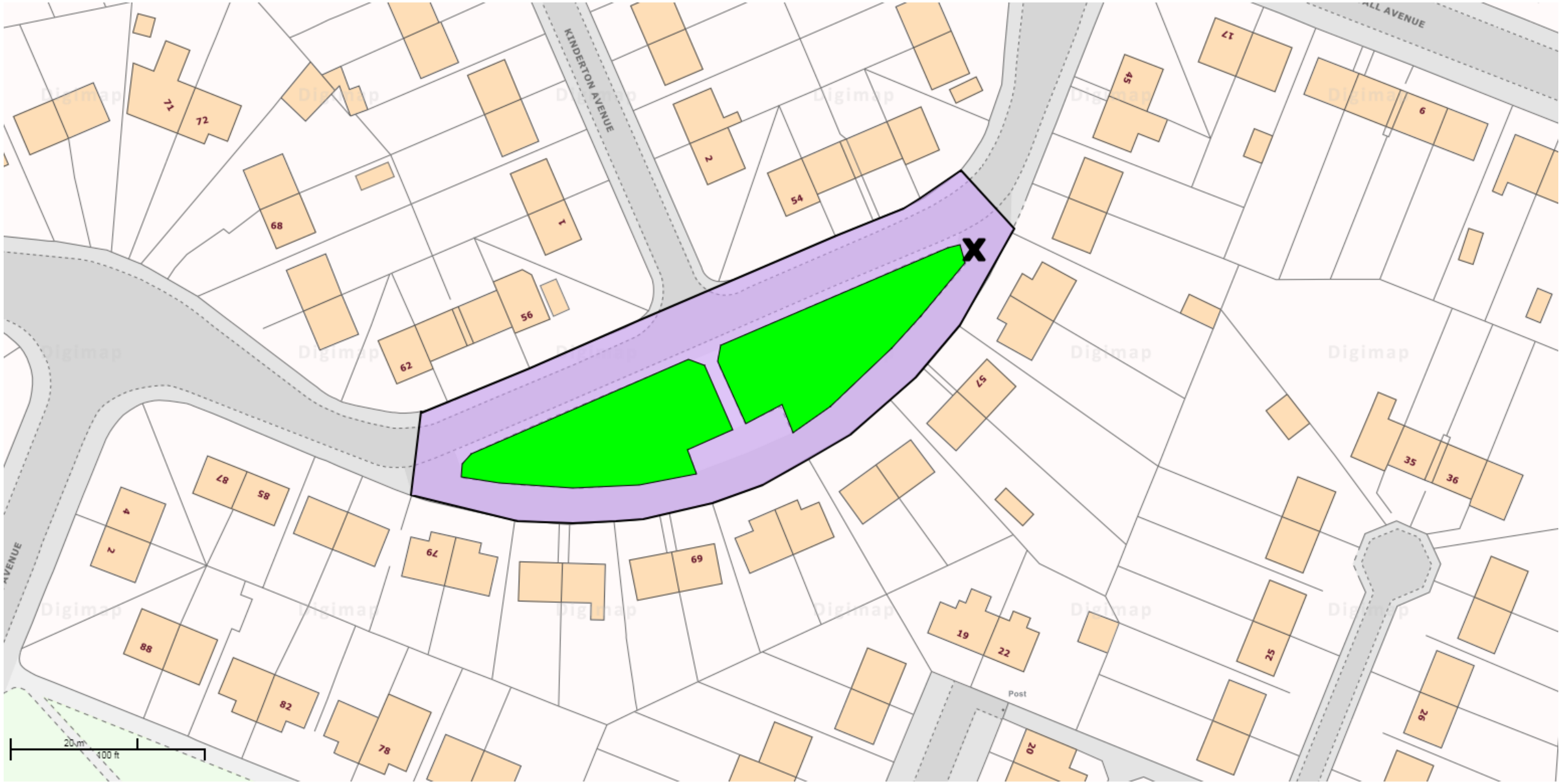
Comparison site 2A



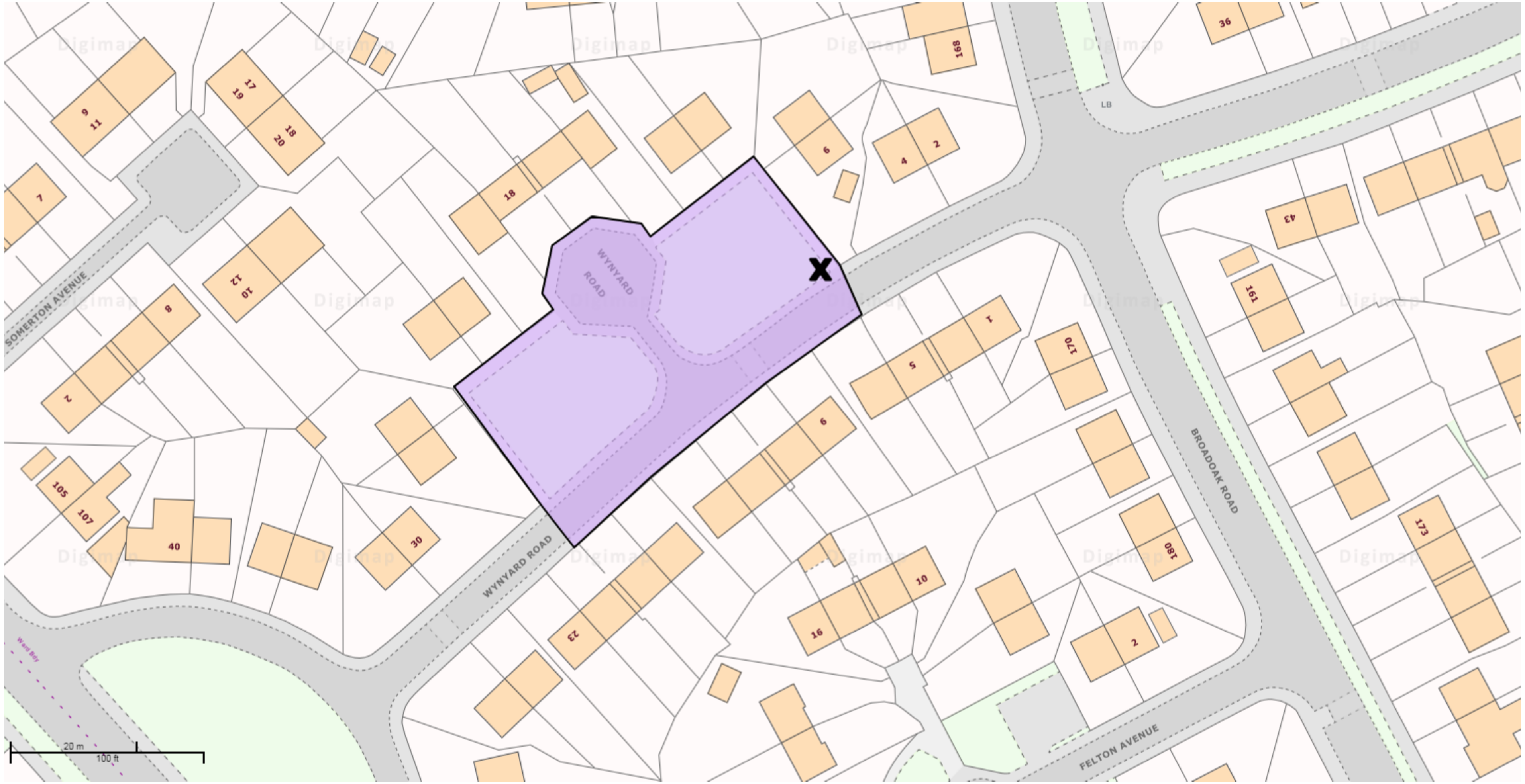
Comparison site 2B



Intervention site 3

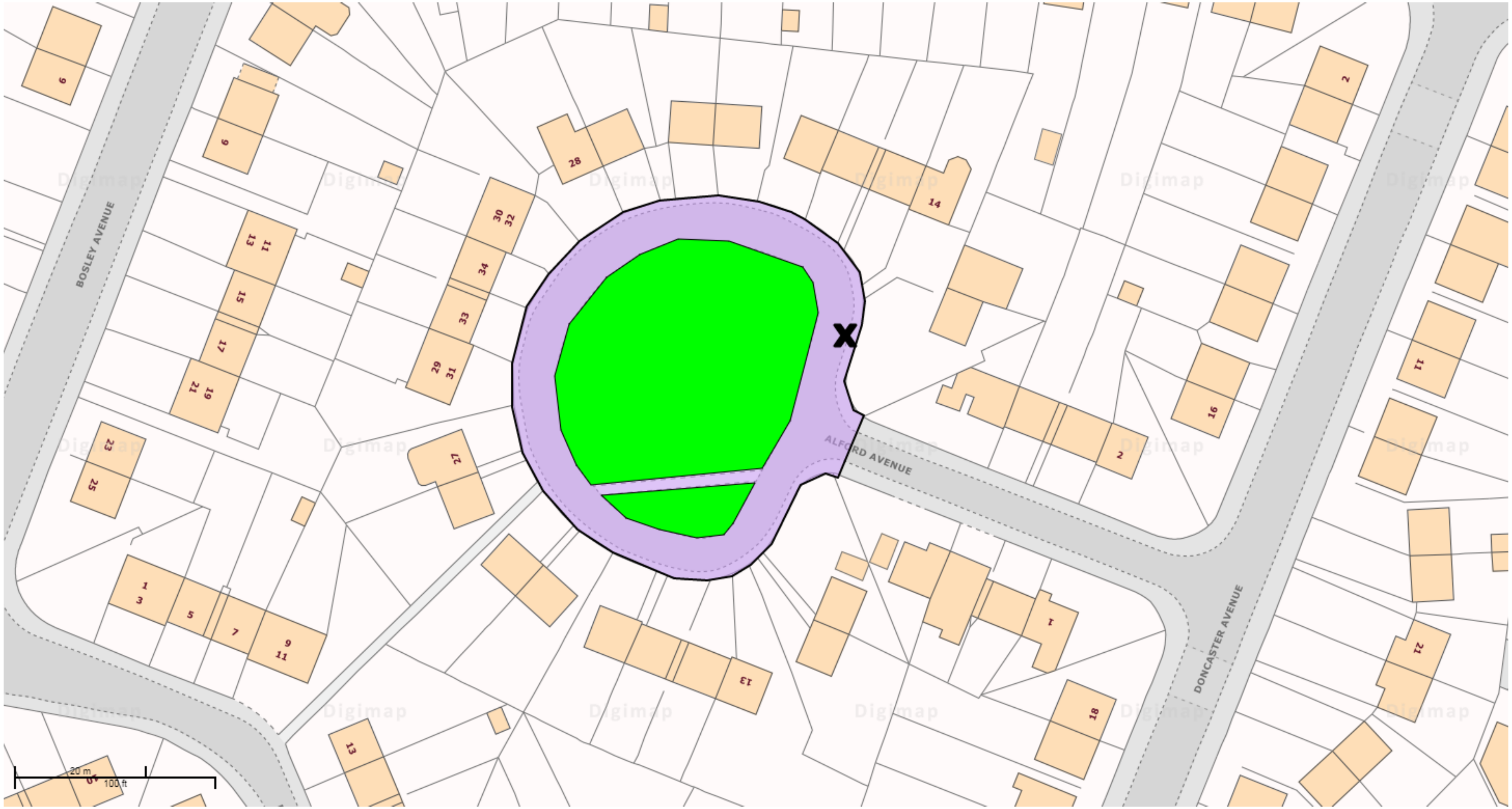


Comparison site 3A

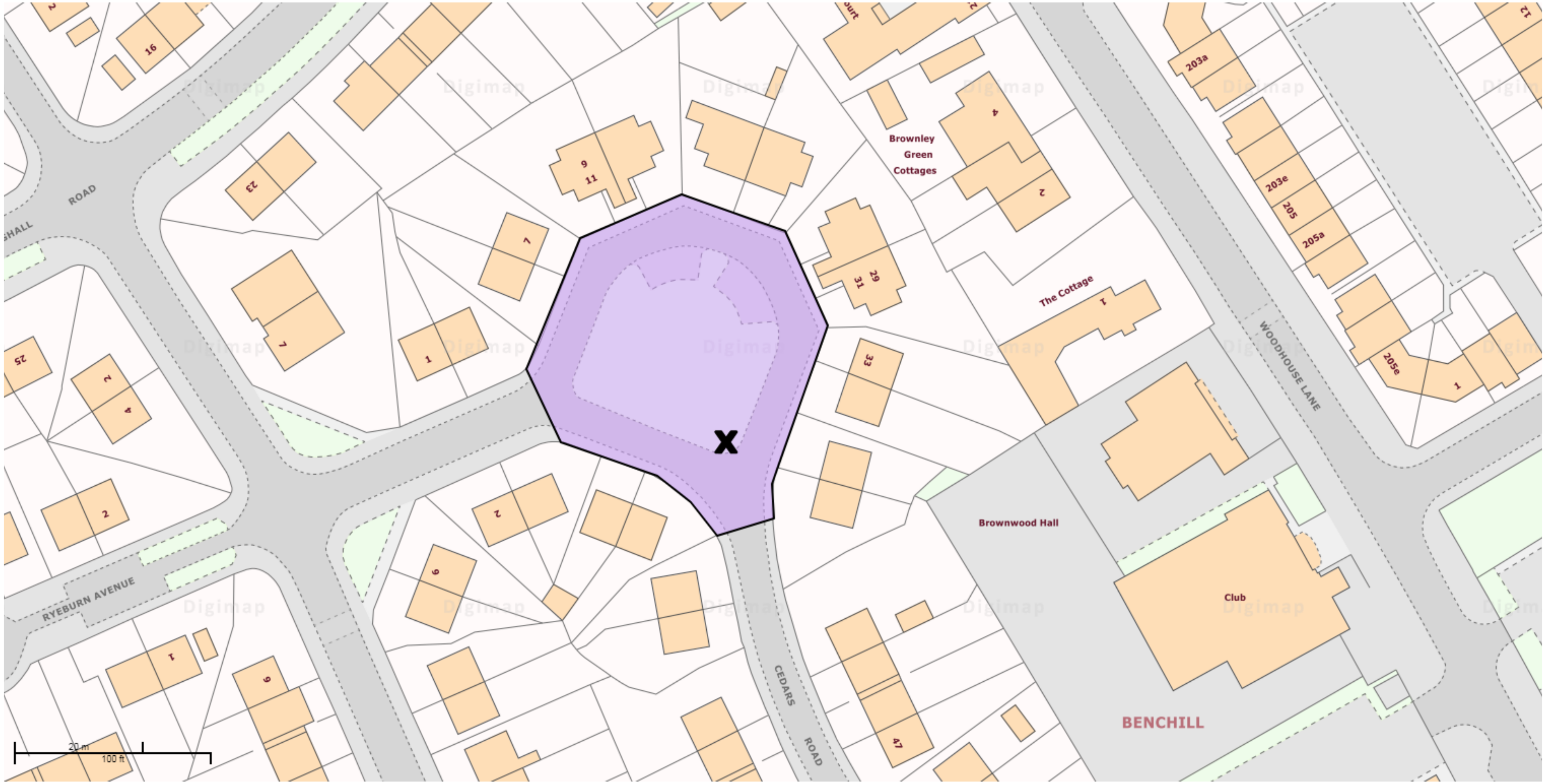


Comparison site 3B

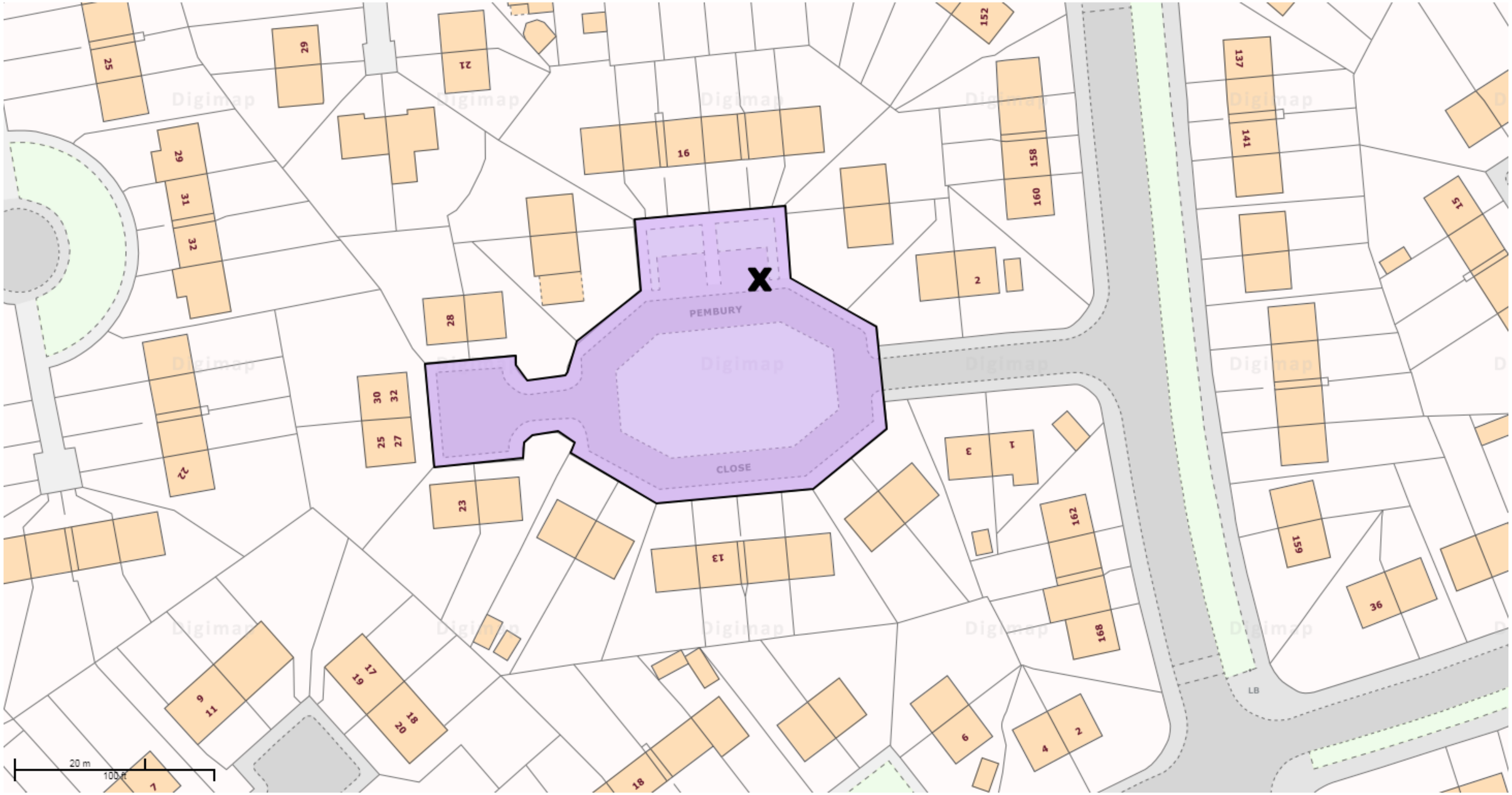
Intervention site 4



Comparison site 4A



Comparison site 4B



Appendix I. Total number of people (median) observed at each site at each time point

Site	Baseline		6 months post-baseline		12 months post-baseline	
	Older adults	**Adults**	**Older adults**	**Adults**	**Older adults**	**Adults**
Intervention site 1	70 (8.5)	885 (100)	63 (8)	1028 (109)	80 (10)	922 (110.5)
Comparison site 1A	88 (10)	652 (55)	120 (14)	711 (66.5)	97 (11)	735 (63)
Comparison site 1B	104 (11)	513 (56)	80 (9)	516 (55)	68 (8.5)	599 (64)
Intervention site 2	33 (4)	204 (26.5)	23 (2.5)	168 (18.5)	18 (2)	180 (19)
Comparison site 2A	43 (4.5)	200 (23)	22 (1.5)	170 (19)	21 (3)	201 (21.5)
Comparison site 2B	19 (1.5)	160 (19)	25 (2.5)	157 (19.5)	17 (2.5)	187 (24)
Intervention site 3	12 (1)	172 (17.5)	25 (2.5)	182 (22.5)	18 (2)	208 (21)
Comparison site 3A	17 (2)	116 (15.5)	5 (0.5)	157 (20)	15 (2)	174 (22.5)
Comparison site 3B	31 (4)	188 (24)	23 (2.5)	193 (21)	17 (2)	175 (21)
Intervention site 4	5 (0)	77 (10)	8 (1)	90 (11.5)	6 (1)	64 (9)
Comparison site 4B	5 (0.5)	71 (6.5)	8 (1)	95 (12.5)	8 (1)	118 (16.5)

All values to one decimal place

Appendix J. Results of the sensitivity analyses (Natural experimental study 1)

Table J1. Median counts of all five behaviours at baseline and 12 months, with high precipitation observation periods removed [a]

Outcome		**Intervention group observation periods (*n* = 51)**					**Comparison group observation periods (*n* = 87)**					**Effect (difference between the change in the two groups)**	**p-value** [b]
		Baseline		**12 months**		**Change in median**	**Baseline**		**12 months**		**Change in median**		
		Median (IQR)	**Total**	**Median (IQR)**	**Total**		**Median (IQR)**	**Total**	**Median (IQR)**	**Total**			
Older adults	Take Notice	0 (0)	1	0 (0)	3	0	0 (0)	5	0 (0)	7	0	0	.59
	Sedentary	0 (0)	5	0 (1)	10	0	0 (1)	16	0 (1)	34	0	0	.87
	Walking	2 (6)	95	2.5 (6.75)	0.5	0.5	4 (5.5)	240	3 (6)	194	-1	1.5	.14
	Vigorous	0 (0)	5	0 (0)	4	0	0 (0)	12	0 (0)	8	0	0	.33
	Connect	0 (1)	20	0 (1.75)	20	0	1 (2)	64	1 (2)	67	0	0	.89
Adults	Take Notice	0 (0)	1	0 (1)	11	0	0 (0)	19	0 (1)	22	0	0	.20
	Sedentary	0 (1)	20	3 (4)	95	3	1 (2)	67	4 (6)	230	3	0	.10
	Walking	19 (25)	871	22 (48.75)	990	3	18.5 (23.25)	1415	21 (26)	1652	2.5	0.5	.06

	Vigorous	2 (8)	162	4 (15.75)	231	2	3 (4)	217	4 (5)	273	1	1	.35
	Connect	7 (18)	299	8 (19)	334	1	7.5 (7.75)	568	9 (9)	626	1.5	-0.5	.39

[a] Observation periods were removed for the sensitivity analysis if there was any precipitation that lasted for more than 50% of the observation period i.e. an overall accumulated duration of 30 minutes or more (recorded by the observer);
[b] A Mann-Whitney U test was carried out to determine if there were differences in the change in counts of behaviours at baseline and 12 months between comparison and intervention groups;
* Statistically significant at $p < 0.05$ (z-test, two-tailed).

Appendix K. Comparison site matching process (Natural experimental study 2)

The overall aim of the comparison site matching process was to identify two closely matched comparison sites for the intervention site, using several key objective and subjective environmental correlates of physical activity.

Step one: Matching at the neighbourhood (LSOA) level

The intervention site is located along a canal waterway, so the overall aim of the process was to identify the most closely matched waterways based on key variables; comparison sites could then be identified within these waterways. The first step was to identify the most closely matched neighbourhoods in which the potential waterways for comparison sites could be searched from.

Due to a lack of available walkability indices for Greater Manchester (GM) it was necessary to manually search for neighbourhoods based on available spatial data at the Lower Layer Super Output Area (LSOA) level (census reporting units containing between 1000 and 3000 individuals) [1,2]. Population density, defined as the number of persons per hectare, was used as a proxy measure of residential density. Street connectivity was measured using street intersection density; the number of 3-way junctions standardised by LSOA area. Normalised Difference Vegetation Index (NDVI) scores, a normalised scale of healthy vegetation cover [3], were used for presence of greenery. Socioeconomic status was measured using the Index of Multiple Deprivation (IMD) Score [4]; an area deprivation score that combines several indicators of deprivation including income, employment, health and crime. Spatial analyses were carried out using ArcGIS 10.4.

To identify the most closely matched LSOAs to the intervention LSOA, a systematic funnelling approach was used. All LSOAs in GM were firstly ranked in order of residential density and the 100 most closely matched LSOAs to each intervention site

LSOA were extracted. These 100 LSOAs were then ranked in order of street connectivity and the 50 most closely matched LSOAs were selected. These two variables were matched first because they are the strongest and most consistent correlates of physical activity [5]. Next, the remaining 50 LSOAs were ranked in order of closeness of IMD score to the intervention LSOA and the most closely matched 25 LSOAs were selected. Finally, these 25 LSOAs were ranked in order of those most closely matched on NDVI scores to the intervention LSOA and the five most closely matched LSOAs were selected. Thus, by the end of this step there were a manageable total of five potential LSOAs for each intervention site.

Step two: Matching at the site level (access to/ availability of destinations and services)

As there were no data available for 'access to/ availability of destinations and services' at the LSOA level, the second step was to measure this variable at the site level. There is a lack of data on walkability scores in GM and manually calculating distances to nearby destinations and services for each site within each LSOA was beyond the scope of this project. Instead, Walk Score (www.walkscore.com) was the most appropriate objective and reliable measure readily available. Walk Score is a website that uses a Google search algorithm to calculate a weighted score (1-100) based on the number and accessibility of amenities (such as shops and parks) within a 1-mile radius of a user-entered postcode, whereby closer amenities with the most accessible walking routes are weighted more strongly. Walk Scores have shown good correlation with gold-standard measures of walkability using Geographic Information Systems (GIS) [6] and have previously been used in studies in the US (e.g. [7]). Walk Score confirmed that data is available in in the UK (WalkScore, personal communication).

To calculate Walk Scores for each street in each LSOA, it was necessary to obtain all postcodes within each LSOA. Postcodes were obtained using FreeMapTools (www.doogal.co.uk/ukpostcodes.php): a free website that provides postcode data within

LSOAs in the UK. Walk Scores were then calculated for each intervention site postcode and all postcodes extracted from the potential comparison site LSOAs. All postcodes were then ranked in order of closeness of Walk Score to each intervention site.

There were difficulties in identifying closely matched comparison sites. To increase the number of potential matches, we considered all types of inland linear waterways which may serve a similar recreational function as canals (e.g. rivers, brooks); which produced the first suitable comparison site (Comparison site 1A).

Step three: Matching at the site level (site level characteristics)

The aim of step three was to find at least three potential comparison sites located within sites (postcodes) most closely matched to the intervention site based on site level characteristics. All postcodes were remotely audited using Google Maps and, where possible, Google Street View. Google Street View can be accessed via Google Maps (www.maps.google.com) and permits users to remotely navigate 360° through panoramic images of the environment from the internet. Virtual environmental audits were preferred given the vast number of postcodes that needed to be examined. Empirical research has previously demonstrated that Google Street View is mostly a reliable and efficient tool to measure the streetscape in comparison with physical on-site audits [8].

Starting with postcodes most closely matched to the intervention site based on Walk Scores, JB audited the following objective streetscape characteristics associated with adults' and older adults' physical activity for each postcode: type of road nearest to the waterway (proxy measure of traffic), presence of a footpath and footbridge, number of pedestrian access points to the waterway (proxy measure of connectivity), number of benches, presence of lighting, presence of greenery, and non-residential buildings. For some of the intervention sites, it was necessary to iteratively search for more potential comparison sites; this was done by identifying more LSOAs and repeating the first three steps. By the end of step three, the intervention site had only one potential comparison site

To identify more potential comparison sites, the same matching process was used but potential neighbourhoods for step one was purposefully identified from the same canal route as the intervention site, within the boundaries of Greater Manchester. This resulted in a second potential comparison site.

Step four: On-site environmental audits

The aim of step four was to ensure the two potential comparison sites were closely matched in terms of quality and quantity of green space and footpaths. These variables are more difficult to reliably judge using Google Street View and therefore required on-site audits of the environment.

JB visited each potential comparison site to systematically audit each site using two validated environmental audit tools: the Neighbourhood Green Space Tool (NGST) [9] and the 54-item abbreviated version of the Microscale Audit of Pedestrian Streetscapes (MAPS-Abbreviated) measure of street design [10]. NGST enabled an audit of the green space characteristics and was specifically developed in the UK; which was important due to the lack of environmental walkability tools that have been developed in the UK [11]. MAPS-Abbreviated was chosen because it measures the quality of pedestrian footpaths and was partly based on a previous tool that has been modified by the Healthy Aging Network [12]. The scores on both tools in each potential comparison site and intervention site were used to ensure the two potential comparison sites were closely matched to the intervention site.

Step five: Matching on pedestrian traffic

The fifth and final step aimed to ensure that intervention and comparison sites were closely matched in terms of pedestrian traffic; that is, the frequency of users passing through a site. It has been found that 15 minutes of observation can provide excellent reliability in estimating the frequency of users passing through a site across the whole hour [13].

Therefore, 15-minute observations were carried out across the three sites (one intervention and two comparison sites) during November 2017 to count pedestrian traffic. JB conducted 15-minute observations at the intervention site and one comparison site on a Friday between 10am-12pm. Due to time restraints, JB conducted 15 minute observations at the other comparison site on a different day (Monday) between 10am-12pm. All final comparison sites were similar to the intervention site in terms of pedestrian traffic and subjective 'feel' of the sites.

Difficulties identifying multiple closely matched comparison sites

We could only identify one closely matched comparison site using this rigorous matching process. To identify a second comparison site, the same matching process was used but potential neighbourhoods in step one was identified from the same canal as the intervention site (i.e. Bridgewater Canal) within Greater Manchester, therefore not matching on neighbourhood level variables. Comparison site 1B was approximately 8.8 km walking distance from the intervention site and there are no direct footpath links between the sites, thus reducing the risk of contamination between intervention and comparison sites.

References

1. Office for National Statistics. 2001 Census: Digitised Boundary Data (England and Wales) [computer file]. UK Data Service Census Support. Downloaded from: http://edina.ac.uk/census. Licensed under the terms of the Open Government Licence [http://www.nationalarchives.gov.uk/doc/open-government-licence/version/2]
2. Office for National Statistics. Super Output Areas: Office for National Statistics, 2011
3. Markevych I, Schoierer J, Hartig T, Chudnovsky A, Hystad P, Dzhambov AM, De

Vries S, Triguero-Mas M, Brauer M, Nieuwenhuijsen MJ, Lupp G. Exploring pathways linking greenspace to health: Theoretical and methodological guidance. Environ Res. 2017;158:301-17.

4. Department for Communities and Local Government. 2015. English Indices of Deprivation 2015 [computer file]. Downloaded from: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015. Licensed under: https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3
5. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJ, Martin BW. Correlates of physical activity: why are some people physically active and others not? Lancet. 2012;380(9838):258–71.
6. Carr LJ, Dunsiger SI, Marcus BH. Validation of Walk Score for estimating access to walkable amenities. Br J Sports Med. 2011;45(14):1144-8.
7. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature. 2017;547(7663):336-9.
8. Badland HM, Opit S, Witten K, Kearns RA, Mavoa S. Can virtual streetscape audits reliably replace physical streetscape audits?. J Urban Health. 2010;87(6):1007-16.
9. Gidlow CJ, Ellis NJ, Bostock S. Development of the neighbourhood green space tool (NGST). Landsc and Urban Planning. 2012;106(4):347-58.
10. Cain KL, Gavand KA, Conway TL, Geremia CM, Millstein RA, Frank LD, Saelens BE, Adams MA, Glanz K, King AC, Sallis JF. Developing and validating an abbreviated version of the Microscale Audit for Pedestrian Streetscapes (MAPS-Abbreviated). J Transport & Health. 2017; 5: 84-96.
11. Brownson RC, Hoehner CM, Day K, Forsyth A, Sallis JF. Measuring the built environment for physical activity: state of the science. Am J Prev Med. 2009;36(4):S99-123.

12. Kealey M, Kruger J, Hunter R, Ivey S, Satariano W, Bayles C, Ramirez B, Bryant L, Johnson C, Lee C, Levinger D. Engaging older adults to be more active where they live: audit tool development. Prev Chronic Dis. 2005; 2:1–2.
13. Benton JS, Anderson J, Pulis M, Cotterill S, Hunter RF, French DP. Method for Observing pHysical Activity and Wellbeing (MOHAWk): validation of an observation tool to assess physical activity and other wellbeing behaviours in urban spaces. Cities Health. 2020.

Appendix L. Target area boundaries for the intervention and comparison sites (Natural experimental study 2)

This document provides maps of the boundaries for the Target Areas in the intervention and comparison sites. Each map also shows the primary location for the observer during observation periods at each site.

Key

X = Observer location

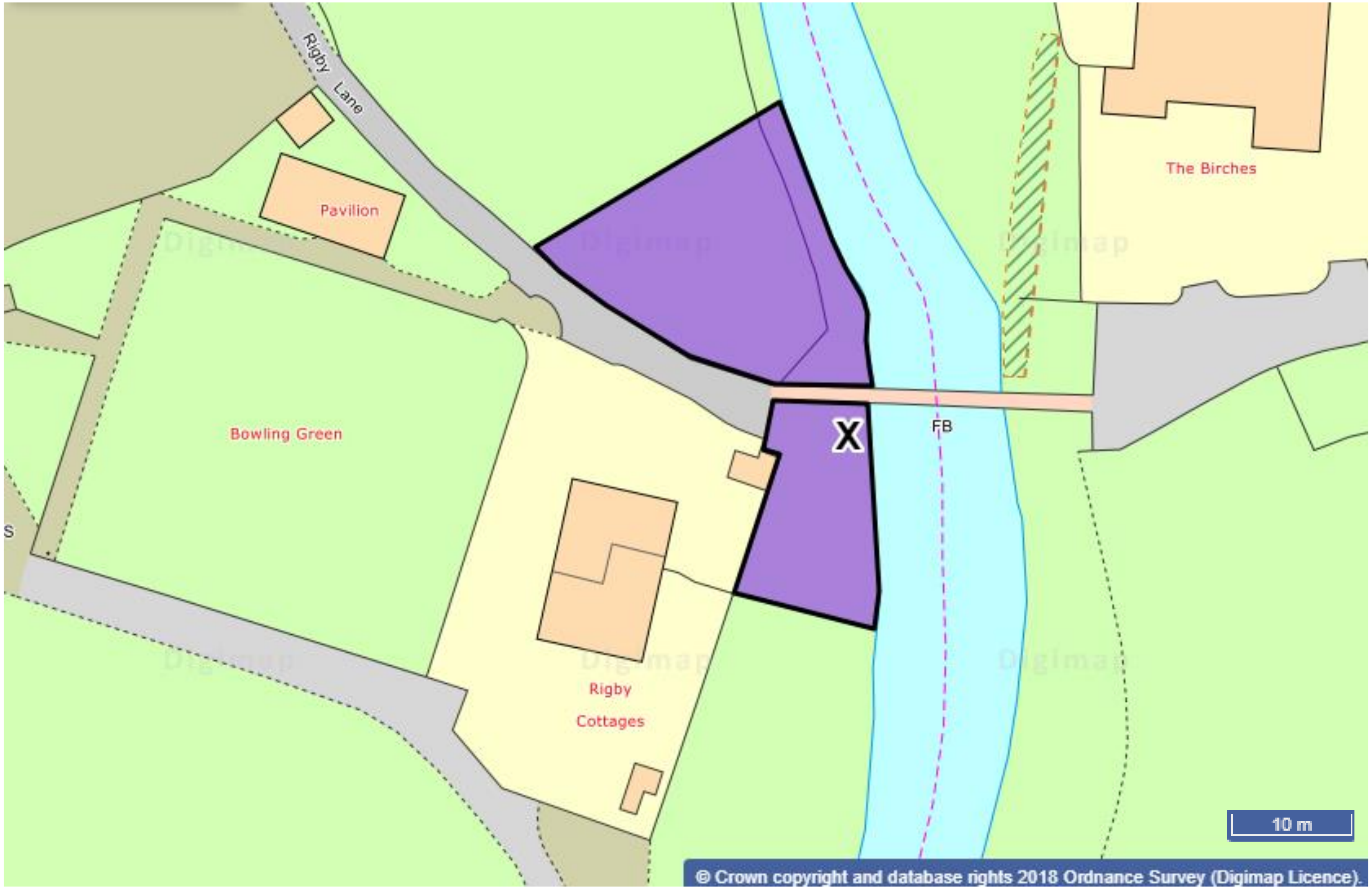
= Target area boundary

= Target area boundary for exploratory analysis (intervention site only)

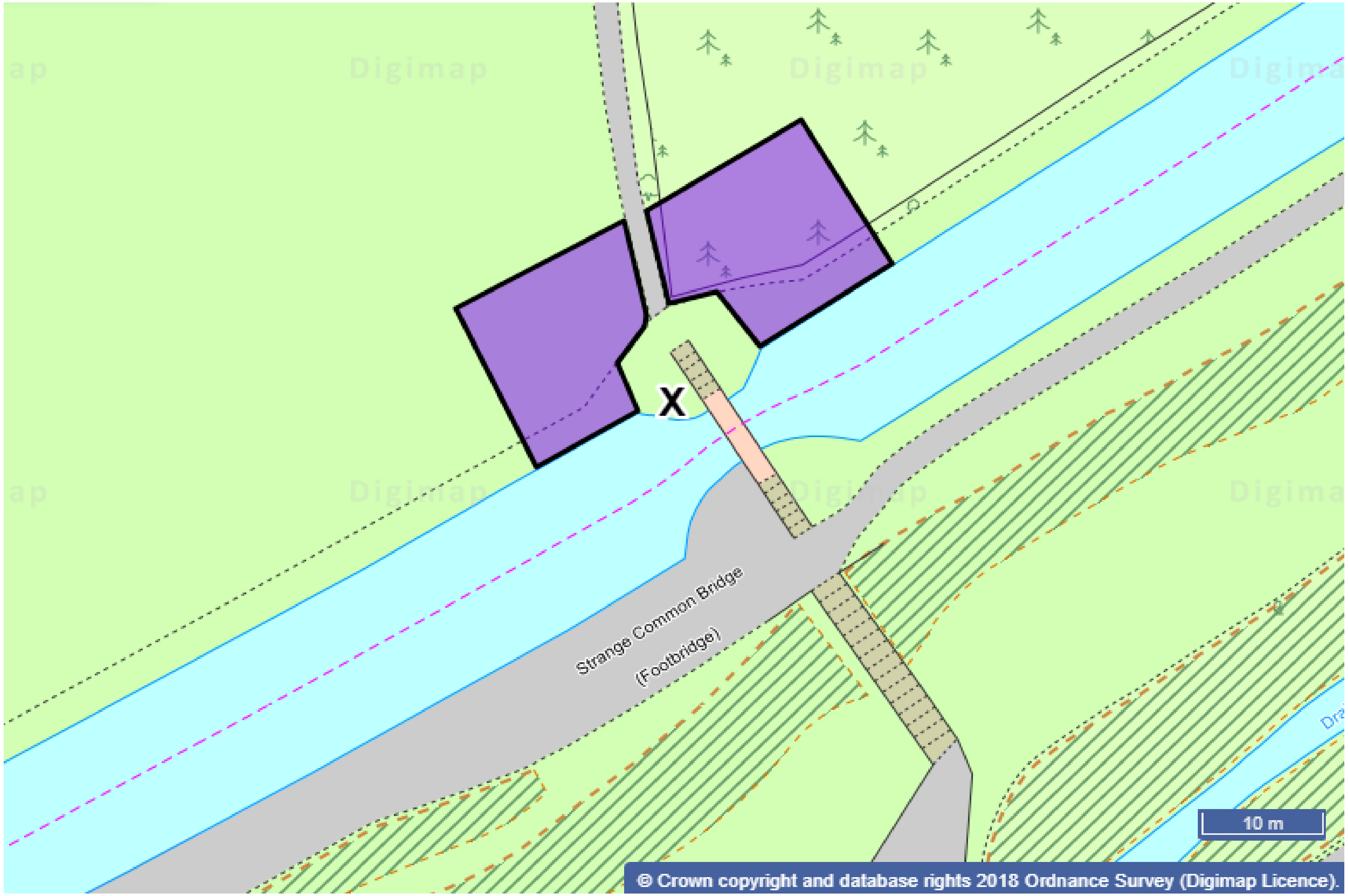
Maps drawn using: www.digimap.edina.ac.uk

Intervention site 1

Comparison site 1A



Comparison site 1B



Appendix M. Participant information sheet

Evaluating the impact of new walking infrastructure and environmental changes along an urban canal on physical activity and wellbeing

Participant Information Sheet

You are being invited to take part in a research study as part of a research project undertaken by a PhD student. The research aims to find out views of the general public about the new walking infrastructure and environmental changes around the Bridgewater canal and Bridgewater Nature Park in Boothstown. In particular, it wants views on the impact of these infrastructural and environmental changes on people's physical activity and wellbeing.

Before you decide to take part it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

Who will conduct the research?

Jack Benton, PhD Health Psychology student.

Supervised by Prof David French, Professor of Health Psychology

School of Health Sciences, Coupland 1 Building, Coupland Street, Oxford Road

Manchester, M13 9PL

What is the aim of the research?

The aim of the current study is to find out the impact of several infrastructural and environmental changes around the Bridgewater canal in Boothstown on peoples' physical activity and wellbeing. These changes include a resurfaced footpath, management of vegetation, installation of benches, new directional signage, new informal play equipment, and changes to Bridgewater Nature Park and Vicars Hall Village Green.

Why have I been chosen?

We are looking to recruit adult volunteers along the Bridgewater canal (at the end of the Moss House Lane Corridor) over the age of 18 years, who are broadly representative of the general population of England, in terms of age and gender.

What would I be asked to do if I took part?

You would be asked to answer some questions about your awareness and opinions of the infrastructural and environmental changes along the canal, as well as the impact of these changes on your physical activity behaviour.

What happens to the data collected?

The data will be used to provide evidence about how infrastructural and environmental changes targeting the use of canals may assist with increasing physical behaviour change and improving wellbeing. This would inform what kinds of changes are most likely to be effective, why and for whom. It will also be used to help the PhD student get their doctorate qualification.

How is confidentiality maintained?

You will not be asked for any information which could identify you as an individual.

What happens if I do not want to take part or if I change my mind?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to provide verbal consent. If you decide to take part you are still free to withdraw at any time without giving a reason and without detriment to yourself.

Will I be paid for participating in the research?

Unfortunately we are not able to provide any payment for your participation in the research.

What is the duration of the research?

The research will take place on a single occasion (today). It will involve completing a questionnaire. It should take less than 10 minutes in total.

Where will the research be conducted?

The research will be conducted at the Bridgewater canal (at the end of the Moss House Lane Corridor) in Boothstown.

Will the outcomes of the research be published?

We are conducting this research as part of a PhD research project. We also expect to publish the results of this study at a future date.

Contact for further information

Should you like any further information, please contact:

Jack Benton

jack.benton@postgrad.manchester.ac.uk

What if something goes wrong?

Should you wish to make a formal complaint about the conduct of the research please contact:

Head of the Research Office,

Christie Building

University of Manchester

Oxford Road

Manchester

M13 9PL

Appendix N. Intercept survey (7 month follow-up version)

Subject ID: ______ **Date/ day:** _____________ **Time period:** _________ **How many in group:** ______

Gender: ☐ Male ☐ Female **Age group:** ☐ Adult ☐ Older adult **Ethnicity:** ☐ White ☐ Non-white

Observed primary activity: ☐ Walking ☐ Dog walking ☐ Jogging/ Running ☐ Cycling ☐ Other (specify)

Q1. Firstly, can I ask do you live in Boothstown?		
☐ Yes (Boothstown)	☐ No (somewhere else)	☐ Prefer not to answer

Q2. Are you using this canal to go anywhere in particular today? ***[Tick all that apply]***		
☐ No destination (just recreation/ exercise)	☐ No destination (dog walking)	☐ Going home from: ……………………………………
☐ Going to: ……………………………………	☐ Escorting child to or from school/ college	☐ Visiting friends/ family
☐ Bridgewater Nature Park	☐ Vicars Hall Village Green	☐ Other ……………………………………

Q3. And do you ever use this canal to go anywhere else other than [INSERT Q2 ANSWER]? ***[Tick all that apply]***		
☐ No destination (just recreation/ exercise)	☐ No destination (dog walking)	☐ Go home from: ……………………………………
☐ Go to: ……………………………………	☐ Escort child to or from school/ college	☐ Visit friends/ family
☐ Bridgewater Nature Park	☐ Vicars Hall Village Green	☐ Other ……………………………………

Q4. Have you been using this canal path for more than 6 months?	
☐ Yes	☐ No

Q5. Have you noticed any recent physical changes to this canal path, and if so what physical changes have you noticed? ***[If yes, tick all that apply]***		
☐ Yes	☐ No [SKIP TO Q7]	
☐ New/ resurfaced footpath	☐ New benches	☐ New/ managed vegetation
☐ New signs/ interpretation boards	☐ Other (please specify) [SKIP TO Q7 IF ONLY OTHER]	

Q6. What do you think of these changes to this canal path?				
☐ Significant improvement	☐ Slight improvement	☐ Neither improved nor worsened	☐ Slightly worse	☐ Significantly worse

Q7. Do you think these changes have encouraged you to visit the canal more often per week?				
☐ Strongly agree	☐ Slightly agree	☐ Neither agree nor disagree	☐ Slightly disagree	☐ Strongly disagree

Q8. Do you think these changes have encouraged you to spend more time using the canal per visit?				
☐ Strongly agree	☐ Slightly agree	☐ Neither agree nor disagree	☐ Slightly disagree	☐ Strongly disagree

Q9. What do you think of these changes to Bridgewater Nature Park?				
☐ Significant improvement	☐ Slight improvement	☐ Neither improved nor worsened	☐ Slightly worse	☐ Significantly worse

Q10. Have you ever visited Vicars Hall Village Green?	
☐ Yes	☐ No [SKIP TO Q13]

Q11. Have you noticed any recent physical changes to Vicars Hall Village Green, and if so what changes have you noticed? ***Iif yes, tick all that apply]***		
☐ Yes	☐ No [SKIP TO Q13]	
☐ New/ resurfaced footpaths	☐ New benches	☐ New/ managed vegetation
☐ Improved access points	☐ Other (please specify) [SKIP TO Q13 IF ONLY OTHER]	

Q12. What do you think of these changes to Vicars Hall Village Green?				
☐ Significant improvement	☐ Slight improvement	☐ Neither improved nor worsened	☐ Slightly worse	☐ Significantly worse

Q13. I'm now going to ask you about your use of this canal path, specifically how often you have used this canal path in the last month and how this compares to this time last year.

In the last month, how often…	**Every day**	**Several times a week**	**Once a week**	**Less than once a week**	**Never**
(a) …have you used this canal path?					
(b) And thinking back to this time last year, how often did you use this canal path?					
(c) I see you are using this side of the canal today, but in the last month, how often have you used the path on the other side of the canal? *[point if necessary]*					
(d) And thinking back to this time last year, how often did you use the path on the other side of the canal?					

Q14. I'm now going to ask you about any physical activity you engage in when using this canal path. Think about all the vigorous activities that you did when using this canal path in the last month. <u>Vigorous physical activities refer to activities that take hard physical effort and make you breathe much harder than normal, such as jogging, running or cycling. Think only about those physical activities that you did for at least 10 minutes at a time.</u> I'm going to ask you about how often you do these vigorous physical activities when using this canal path in the last month, and how this compares to this time last year.

In the last month, how often…	**Every day**	**Several times a week**	**Once a week**	**Less than once a week**	**Never**
(a) …did you do vigorous physical activities when using this canal path?					
(b) And thinking back to this time last year, how often did you do vigorous physical activities when using this canal path?					

Q15. I'm now going to finish off by asking some questions about your physical activity in your everyday life. This includes activities on the canal path but also in other settings such as any activities for recreation, sport, exercise or leisure; at work and at home; or to get from place to place. *[Repeat underlined definition from Q14].* I'm going to ask you about how often you have done vigorous physical activities in your everyday life in the last month, and how this compares to this time last year.

In the last month, how often…	Every day	Several times a week	Once a week	Less than once a week	Never
(a) … did you do vigorous physical activities in your everyday life?					
(b) And thinking back to this time last year, how often did you do vigorous physical activities in your everyday life?					
Finally, think about the time you spent walking in the last month. This includes along the canal, but also at work and at home, walking to travel from place to place, and any other walking that you have done solely for recreation, sport, exercise, or leisure.					
(c) … did you walk for at least 10 minutes at a time?					
(d) And thinking back to this time last year, how often did you walk for at least 10 minutes at a time?					

Q16. Is there anything else you'd like to tell me about your experience of using this canal path?

Appendix O. Intercept survey and results

Subject ID: ______ **Date/ day:** _______________________________ **Time period:** ____________

Intro script: ***"Hello, excuse me, REALLY sorry to bother you, I am from the University of Manchester. We are doing a very short physical activity survey: could you spare 5 minutes to complete one?"***

Q1. Firstly, can I ask do you live in Boothstown?		
☐ Yes (Boothstown)	☐ No (somewhere else)	☐ Prefer not to answer

Q2. When was the first time you started using *this particular path* along the canal? *[Point and emphasise this canal path]*			
☐ < 1 month ago	☐ 1 – 6 months ago	☐ 6 – 12 months ago	☐ > 12 months ago [SKIP TO Q3]
Q2a. Why did you start using this particular path along the canal? ☐ Because of recent changes ☐ Other reason:			

Q3. I'm now going to ask you some questions about your use of this canal in the past month, and how this compares to this time one year ago.

In the past month, how often per week…	**Every day**	**A few times a week**	**Once a week**	**Less than once a week**	**Not used**
(a) …have you used this particular canal path? *[point to this canal path]*					
(b) And do you currently use this particular canal path less than, the same as, or more than this time one year ago? *[point to this canal path]*	***Less often*** **than this time last year**	***About the same*** **as this time last year**	***More often*** **than this time last year**	**Didn't use this canal path this time last year**	
In the past month, how often per week…	**Every day**	**A few times a week**	**Once a week**	**Less than once a week**	**Not used**
(c) … have you used the path on the other side of the canal? *[point to other side]*					
(d) And do you currently use the path on the other side of the canal less than, the same as, or more than this time one year ago? *[point to other side]*	***Less often*** **than this time last year**	***About the same*** **as this time last year**	***More often*** **than this time last year**	**Didn't use this canal path this time last year**	

Intervention questions

Q4. Have you noticed any recent physical changes to this particular canal path and the areas around the canal, and if so, what physical changes have you noticed? *[If yes, tick all that apply]*		
☐ Yes	☐ No [SKIP TO FINAL Q7 and Q8]	
☐ Refurbished footpath	☐ New benches	☐ New/ managed vegetation
☐ New signs/ interpretation boards	☐ Other (please specify)	

Q5. Do you think these changes have encouraged you to visit the canal more often per week?				
☐ Strongly agree	☐ Slightly agree	☐ Neither agree nor disagree	☐ Slightly disagree	☐ Strongly disagree

Q6. Do you think these changes have encouraged you to spend more time using the canal per visit?				
☐ Strongly agree	☐ Slightly agree	☐ Neither agree nor disagree	☐ Slightly disagree	☐ Strongly disagree

Physical activity questions

Q7. For the final two questions, I'm going to ask you about your physical activity levels in your everyday life, not just on the canal path. In the past month, how often did you do vigorous physical activities in your everyday life? <u>Vigorous physical activities refer to activities that take hard physical effort and make you breathe much harder than normal, such as jogging, running or cycling. Think only about those physical activities that you did for at least 10 minutes at a time.</u>				
☐ Every day	☐ Several times a week	☐ Once a week	☐ Less than once a week	☐ Never

Q8. Finally, in the past month, how often did you walk for at least 10 minutes at a time? <u>This includes at work and at home, walking to travel from place to place, and any other walking that you have done solely for recreation, sport, exercise, or leisure.</u>				
☐ Every day	☐ Several times a week	☐ Once a week	☐ Less than once a week	☐ Never

FINISH

(To be completed by researcher immediately after completion of intercept survey)

How many in group: ______

Gender: ☐ Male ☐ Female **Age group:** ☐ Adult ☐ Older adult **Ethnicity:** ☐ White ☐ Non-white

Observed primary activity: ☐ Walking ☐ Dog walking ☐ Jogging/ Running ☐ Cycling ☐ Other (specify)

Table O1. Main intercept survey results

Question	Response options	Frequencies [a] (% of total eligible)
When was the first time you started using this particular path along the canal?	Nov 2017 or earlier (pre-intervention)	20 (37.7%)
	May 2018 or later (post-intervention)	33 (62.3%)
[Of those who started using canal path after the intervention] Why did you start using this particular path along the canal?	Because of recent changes	27 (81.8%)
	Moved house	5 (15.2%)
	Other	1 (3%)
Change in canal usage	Always used intervention path (no change)	20 (37.7%)
	Started using new canal path after the intervention changes, but displaced from elsewhere on canal (displacement)	23 (43.4%)
	Started using new canal path after the intervention changes (new canal users)	7 (13.2%)
	Did not answer	3 (5.7%)
Have you noticed any recent physical changes to this particular canal path and the areas around the canal?	Yes	47 (88.7%)
	No	6 (11.3%)
Do you think these changes have encouraged you to visit the canal more often per week?	Strongly agree	1 (1.9%)
	Slightly agree	2 (3.8%)
	Neither agree nor disagree	20 (37.7%)
	Slightly disagree	13 (24.6%)
	Strongly disagree	12 (22.6%)
	Did not answer	5 (9.4%)
Do you think these changes have encouraged you to spend more time using the canal per visit?	Strongly agree	3 (5.7%)
	Slightly agree	7 (13.2%)
	Neither agree nor disagree	13 (24.5%)
	Slightly disagree	10 (18.9%)
	Strongly disagree	15 (28.3%)
	Did not answer	5 (9.4%)

[a] There was a total of 53 participants

Appendix P. Results of the sensitivity analyses (Natural experimental study 2)

Table P1. Median counts of wellbeing behaviours, with high precipitation observation periods removed [a]

Outcome		Intervention observation periods (*n* = 46)			Comparison observation periods (*n* = 96)			Effect (difference between the change in the two groups)	*p*-value [a]
		Median (IQR)	Total	Change in median (from baseline)	Median (IQR)	Total	Change in median (from baseline)		
Walking (secondary)	Baseline	4 (1.75)	48	-	2 (1)	47	-	-	-
	7 months	12.5 (8.5)	135	8.5	5 (3)	116	3	5.5	.004*
	12 months	11.5 (10)	140	7.5	5 (4.5)	71	3	4.5	.002*
	24 months	12 (10.5)	135	8	4 (4)	64	2	6	< 0.001*
Vigorous (secondary)	Baseline	0 (0.75)	3	-	0 (0)	1	-	-	-
	7 months	0.5 (2)	13	0.5	0 (0)	1	0	0.5	.02*
	12 months	0 (0.75)	7	0	0 (0)	1	0	0	.95
	24 months	1 (2)	13	1	0 (0)	0	0	1	.01*

Sedentary (exploratory)	Baseline	0 (0)	2	-	0 (0)	3	-	-	-
	7 months	1.5 (4)	24	1.5	0 (1)	17	0	1.5	.07
	12 months	2 (1)	31	2	0 (1)	8	0	2	< 0.001*
	24 months	3 (1.5)	24	3	0 (1)	8	0	3	< 0.001*
Connect (exploratory)	Baseline	1.5 (2)	17	-	1 (2)	22	-	-	-
	7 months	4 (5.25)	49	2.5	2 (2)	41	1	1.5	.24
	12 months	3 (5.75)	40	1.5	1 (2)	19	0	1.5	.04*
	24 months	5 (6.5)	49	3.5	0 (2)	18	-1	4.5	.001*
Take Notice (exploratory)	Baseline	0 (0.75)	3	-	0 (0)	5	-	-	-
	7 months	1 (1.75)	11	1	0 (0.75)	6	0	1	.052
	12 months	0 (1)	8	0	0 (1)	7	0	0	.60
	24 months	2 (1)	15	2	0 (0)	3	0	2	< 0.001*

[a] Mann-Whitney *U* tests were carried out to determine if there were significant differences in the change in counts of behaviours from baseline between intervention and comparison groups;
* Statistically significant at $p < 0.05$ (z-test, two-tailed);
IQR interquartile range