Swansea University E-Theses

Investigating the impact of GIS modelled daily exposures to the retail food environment on routinely linked child health data

Mizen, Amy

Investigating the impact of GIS modelled daily exposures to the retail food environment on routinely linked child health data

By

Amy Rhian Mizen, MSc, BSc

Thesis submitted to Swansea University

in fulfilment of the requirements for the Degree of

DOCTOR OF PHILOSOPHY

Swansea University Medical School

Swansea University

Swansea SA2 8PP

United Kingdom

May 2018

Summary

Obesity continues to be a huge public health concern around the globe, and numbers are projected to continue to increase. There is particular concern around the issue of obesity in children because obese children are far more likely become obese adults than children who are a healthy weight. We have so far been ineffective in developing successful public health policies and interventions that report population level reductions in obesity. In order to tackle obesity on a large scale, we need to be creative and develop interventions and policies that drive societal change.

The cause of obesity has been found to be not a linear relationship of cause and effect but a complex and multifaceted system. Geographic Information Systems (GIS) are being used to more fully understand the role of the environment on obesity. There has been a particular focus on exposure to the 'retail food environment' (RFE) and how this may be linked with obesity. Currently, GIS modelled exposures to the RFE along routes to and from school are not adequate to make reliable predictions about exposure. Instead, GPS data are used to obtain accurate exposures. This thesis has developed a GIS method to generate population level exposures to the RFE. In order to advise policies and interventions that will effectively cause societal change, population level research must be undertaken. A novel way that this type of research can be undertaken is through data linkage. This study has calculated exposures to the RFE for school children aged 13-14 years in south Wales and linked these exposures to individual level health data held within the Secure Anonymised Information Linkage Databank (SAIL). These results contribute to the evidence base and shed light on new aspects of the built environment that can be altered to encourage healthy lifestyles.

Declaration

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

AMY MIZEN

May 2018

Statement

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by explicit references. A bibliography is appended.

AMY MIZEN

May 2018

Statement

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan (subject to the law of copyright), and for the title and summary to be made available to outside organisations.

AMY MIZEN

May 2018

Contents

Acknowledgements

First of all, I would like to thank my supervisors Professor Sarah Rodgers, Dr Richard Fry and Professor Ronan Lyons for their invaluable support, guidance and expertise. I have thoroughly enjoyed working with you and I am so thankful for the many opportunities that you have given me.

I would like to thank everyone at SAIL and the Faar Institute at Swansea University who have supported and encouraged me as I have begun my research career. I am so thankful to the team for sharing their knowledge and experience with me. Drinking many cups of tea together has also helped!

A huge thanks goes to my husband Oli, for his support and patience throughout the whole process; particularly in the final few months. Thanks also to my family and friends for their love and support. The endless cheers from them all have been a huge source of encouragement.

Finally, I would like to thank DECIPHer for funding this project and for the opportunity to study a PhD with them. I have appreciated being a part of the DECIPHer community on a professional and personal level. I am very grateful for all of the opportunities I have been given from working with them.

Amy Mizen

Swansea University

December 2017

List of Figures

List of Tables

Abbreviations

Abbreviation	Term
ABP	Address Base Premium
ALF	Anonymous Linking Field
ALPHA	Advice Leading to Public Health Advancement
AONB	Area of Outstanding Natural Beauty
AST	Active School Travel
AT	Active Travel
BIA	Bioelectrical Impedance Analysis
BMA	British Medical Association
BMI	Body Mass Index
CMP	Child Measurement Programme
DECIPHer	The Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement
EH	Environmental Health
Exposure am	Exposure to the RFE on the walk to school in the morning
Exposure pm	Exposure to the RFE on the walk home from school in the afternoon
FOI	Freedom of Information request
FSA	Food Standards Agency
FSM	Free School Meal
GIS	Geographic Information Systems

GP	General Practitioner
GPS	Global Positioning System
HoG	Head of Geography
HSW	Health Survey Wales
IGRP	Information Governance Review Panel
ISO	International Organisation for Standardisation
LA	Local Authority
LAs	Local Authorities
LPA	Local Planning Authority
LSOA	Lower Super Output Area
MLM	Multilevel model
NHS	National Health Service
NRES	National Research Ethics Service
NWIS	NHS Wales Informatics Service
ONS	Office for National Statistics
OS	Ordnance Survey
OSM	OpenStreetMap
PEAR	Physical Environment and Activity Relationships
POI	Points of Interest
PPV	Positive Predicted Value
PSMA	Public Sector Mapping Agreement
QGIS	Quantum GIS
RALF	Residential Anonymous Linking Field
RFE	Retail Food Environment

SAIL	The Secure Anonymised Information Linkage Databank
SNR	Shortest Network Routes
TTP	Trusted Third Party
UPRN	Unique Property Reference Number
WDS	Welsh Demographic Service
WIMD	Welsh Index of Multiple Deprivation
WNE	Weighted Network Exposure
WNR	Weighted Network Routes

Chapter 1

Introduction

Obesity continues to be a huge public health concern around the globe and numbers are projected to continue to increase. Not only does obesity have serious implications to an individual's physical and mental health, but obesity is costing the UK government billions of pounds [1]. Governments and organisations around the world state that we are reaching breaking point in dealing with the challenge [2]. There is particular concern around the issue of obesity in children because they are a vulnerable demographic. Obese children are far more likely become obese adults than children who are a healthy weight because of habits and attitudes that are established at an early age [3, 4].

We have so far been ineffective in developing successful public health policies and interventions that report population level reductions in obesity. Interventions are often small scale and focussed on changing individual-level habits: such as cooking classes, "get-fit" camps, and family-focus groups. Many studies that have employed this approach have reported participant weight decreasing and improved attitudes to health [5]. However, these studies are resource and time intensive and it is not possible to implement these interventions on a population level.

Furthermore, it is also imperative that interventions and policies do not contribute to rising health inequalities. People with a lower socio-economic status are more likely to be overweight or obese. In southern Europe, education-related inequalities in obesity are four times greater for the least educated [6]. In the UK, 28.5% of children living in the most deprived areas are overweight or obese [7]. In order to remedy this inequality, the UK government is encouraging Local Authorities (LAs) to implement policy to tackle health inequalities.

In order to tackle obesity on a large scale, we need to be creative and develop interventions and policies that drive societal change. Governments are looking to public health researchers to better understand obesity and its pathways. This is in order that interventions and policies can be developed and applied to the complex, adaptive

systems within which most public health interventions and policies have to operate [8]. An example of this is the sugar tax that the UK Government employed in April 2018. This tax is applied to food manufacturers and the more sugar that a product contains, the higher the tax. Some manufacturers have already reduced the sugar content in their products. For products that become more expensive for consumers to purchase, it is hoped that this will discourage consumption.

The cause of obesity is not a linear relationship of cause and effect but a complex and multifaceted system. The Foresight Report was commissioned by the UK government to bring together evidence and to identify the broad range of factors that influence obesity [9]. The report highlighted the importance of the environment on obesity and policy makers are beginning to understand that modifying the built environment may be a way to help people to follow lifestyles less conducive to obesogenic behaviours. This may help to reduce, or at least curb, obesity rates at a population scale. Researchers are looking to develop their understanding of which aspects of the built environment influence Body Mass Index (BMI), and attitudes to food or physical activity. Rutter et al. [8] recognised that many of the challenges within public health remain resistant to policies and interventions. Rutter et al. attributed this to evidence being focussed on linear cause and effect models, rather than understanding complex systems [8]. The report by Rutter et al. goes on to suggest that research and policy need to consider processes and outcomes within a system that can drive change and move away from linear, causal models. This means identifying systems and facilitating change in these systems in favourable ways. With regard to obesity research, systems such as food, employment, transport, economic and other systems influence energy intake and expenditure [8].

Geographic Information Systems (GIS) are being used to more fully understand the role of the environment on obesity. Researchers use GIS to analyse aspects of the environment, such as access to unhealthy food or opportunities for physical activity, to evaluate whether these factors are linked with obesity. However, GIS methods used in these studies are heterogeneous and the results equivocal. It is therefore difficult to interpret the influence of the built environment on obesity. There has been a particular focus on exposure to the 'retail food environment' (RFE) and how this may be linked with obesity. GIS measures of exposure to the RFE widely differ, with studies using different data sources, defining the RFE in different ways, and using heterogeneous

methods to measure access or exposure to the RFE [10]. Currently, GIS modelled exposures to the RFE along routes to and from school are not adequate to make reliable predictions about exposure. Studies use shortest network routes as a proxy for the route that a child travels to and from school. However, this methodology has been found to be a reliable proxy for the distance a child travels, but not for the environment they are exposed to [11]. Instead, GPS data are used to obtain accurate exposures. However, collecting GPS data can be resource and time intensive and sample size is restricted. This thesis seeks to develop current GIS methods in order for population level exposures to the RFE to be generated.

In order to advise policies and interventions that will effectively cause societal change, population level research must be undertaken. A novel way that this type of research can be undertaken is through data linkage. The process of data linkage involves combining datasets that already exist and are routinely collected, using a common identifier. This gives researchers access to extensive population data which makes it an efficient and cost effective way to undertake large-scale studies.

This study has calculated exposures to the RFE for school children aged 13-14 years in south Wales and linked these exposures to individual level health data held within the Secure Anonymised Information Linkage Databank (SAIL). The exposures were calculated using LA food outlet data and an open source, weighted, road network. The aims, objectives and main research questions of the thesis are stated below.

Aims:

1. Understand the existing evidence base and identify gaps in the literature;
2. Collect food outlet data from Welsh Local Authorities;
3. Evaluate LA data quality;
4. Evaluate quality of alternative food outlet data sources;
5. Evaluate whether OpenStreetMap is an appropriate alternative to an authoritative road dataset;
6. Assess the impact of including footpath data in modelled walking routes;
7. Establish a method to collect route data on walking routes to school;
8. Collect a robust and representative dataset of walking route data;
9. Generate a GIS model that can accurately predict children's walking routes and their associated exposures to the retail food environment;

10. Investigate whether higher exposures to the retail food environment in children's day to day lives is associated with a higher BMI.

Objectives:

1. Undertake a literature review to understand what research has already been undertaken in this field and what gaps in the literature exist:
    a. Obesity and its causes;
    b. GIS and its use in Obesity Research;
    c. Data Linkage.
2. Request and obtain food outlet data from 22 Welsh LAs ;
3. Document experience of obtaining data from LAs;
4. Assess the spatial quality of the LA longitudinal food outlet dataset using an internationally recognised spatial data quality standard assessment;
5. Assess the spatial quality of two alternative food outlet datasets commonly used in obesity studies, using an internationally recognised spatial data quality standard assessment;
6. Evaluate the most appropriate dataset to use as food outlet location data source in exposure modelling;
7. Compare OSM coverage in Wales with an authoritative dataset to see whether it is a reliable data source to use in exposure modelling;
8. Assess the influence of including footpath data in the road network in exposure modelling;
9. Run a pilot data collection to investigate whether mapping survey's produce reliable and detailed data of children's routes to school;
10. Collect a large, detailed sample of children's walking routes to and from school;
11. Prepare this data so that environmental characteristics along routes to and from school can be investigated;
12. Incorporate the environmental characteristics that differ between the shortest network routes and the GPS routes in to GIS model by weighting the network;
13. Generate routes for all individuals in the GPS dataset using the weighted network model and compare the difference between the route shape and associated exposures for the two route types;

14. Link the GIS generated, weighted network routes with health data in an anonymised safe haven;

15. Statistically analyse the linked dataset to understand relationships between the RFE and BMI.

Research Questions

- Is LA data the most appropriate data source for food outlet data?
- Is OSM a reliable alternative to ITN?
- Do environmental characteristics of walking route differ between modelled walking routes and actual routes travelled by children?
- Can a weighted network be used to accurately model children's walking routes and their associated exposures?
- Are higher exposures to the retail food environment in children's day to day lives is associated with a higher BMI?
- Do differences between environmental exposures and BMI vary between deprivation and affluent individuals?
- Are there differences in BMI between children that have the potential to walk to school and children that do not?

This thesis acknowledges the importance of considering how processes and outcomes drive change within a system. This project investigated exposure to the RFE and makes recommendations to make positive changes to the built environment which have the potential to drive change. These results contribute to the evidence base and shed light on new aspects of the built environment that can be altered to encourage healthy lifestyles.

Chapter 2

Review of Literature

This chapter is divided into three principle sections and will give an overview of:

a) obesity and its causes;

b) Geographic Information Systems (GIS) and how it has been used in obesity research;

c) data linkage and how it has been applied in health and obesity research.

2.1 Obesity and its causes

2.1.1 What is obesity?

The global obesity epidemic is a public health priority in countries around the world [12, 13]. It has been acknowledged as one of the most serious health problems of the 21st century [14]. Obesity is a serious problem that is prevalent in huge proportions across the globe [15]. The term obese describes a person who has accumulated a large amount of body fat and this may have a detrimental effect on their health [16]. The most common measure of obesity is the Body Mass Index (BMI). It has been used in clinical studies, public health research and community based programs [17]. BMI has been criticised as an unreliable measure of obesity because the measure does not differentiate between muscle, bone and other lean body mass [18]. In addition, BMI and other indicators of obesity, such as bioelectrical impedance analysis (BIA) and waist circumference, have different impacts on outcomes [19, 20] and longitudinal obesity trends [21–23]. There are also concerns with the standardisation of measuring BMI within studies and between studies. Furthermore, many large scale studies use self-reported weights and heights as it is possible to collect large cohorts of measurements. However, a systematic review comparing objective and subjective measures of height and weight concluded that participants under-report their weight and over-report their height [24]. The limitations of using BMI in studies have been discussed at length in the literature [25], and researchers are looking to develop a more accurate measure that can be routinely collected on the scale that BMI has been

collected [23]. In contrast, many researchers maintain that it is an appropriate measure of fatness, particularly for large-scale studies. Smaller, clinical studies favour using additional methods to measure fatness in conjunction with BMI [26]. And broader population studies have found that BMI remains the most robust measure of body composition that can be collected on a large scale [7, 27–31].

2.1.2 Causes of obesity

Obesity is caused and influenced by many synergistic factors [32–34]. Physiological, economic, social and environmental factors all contribute to promoting obesogenic lifestyles, which have resulted in the increased prevalence of obesity across the world. It can be difficult to separate out these factors as they are influenced by and have an influence on one another.

The Foresight report [9] published a complex graphic to capture the complexity of obesity which included 108 variables. The map was praised for contributing a greater understanding of the complexity of obesity and the many challenges that are faced in addressing obesity [33]. The Foresight report [9] highlighted that the prevalence of obesity in the UK is due to sedentary lifestyles and easy access to energy dense foods. The drivers of these challenges span the social, economic and environmental domains.
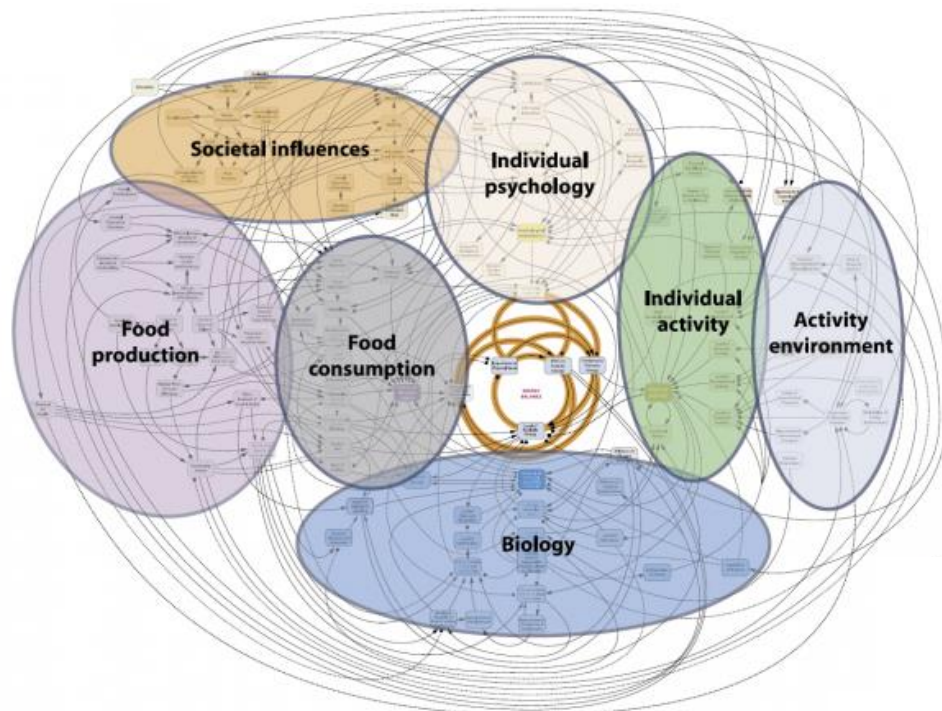
Figure 2.1. Foresight map of causes of obesity. Main domains of drivers of obesity overlain (from https://foresightprojects.blog.gov.uk/2017/10/04/dusting-off-foresights-obesity-report/)

2.1.2.1 Biology

Obesity works at an individual level and therefore some people are more vulnerable to becoming overweight and obese than others [9]. A number of genes have been discovered that make individuals predisposed to putting on weight or becoming obese [35] and medications have also been found to cause weight gain as a side effect [36]. In addition, chemicals present in processed foods have also been found to disrupt the metabolism of fat and sugar [37], resulting in weight gain.

2.1.2.2 Societal influences

Social and economic influences in western culture have further encouraged obesogenic behaviour. Increased portion sizes [38, 39], more food consumed away from home [40], inactivity in jobs and travel [41], the relatively high cost of healthy foods, and the ready availability of energy dense foods are believed to have resulted in a sharp increase in the number of obese people over the past fifty years [42, 43]. Furthermore, the cheap cost of energy dense food has meant that it is easy to purchase unhealthy food for all social groups. For the poorest in developed societies, buying cheap, energy dense food is an attainable way to ensure a filling meal is had. Studies have also found that with individuals working longer hours and more women working, there is a greater reliability on convenience foods that are often high in sugar and fat [44, 45]. Sedentary lifestyles have also been encouraged by car ownership, desk jobs, inactive leisure times and machines that make it easier to carry out domestic practices, such as dishwashers and washing machines. For children, televisions, games consoles and the rise of social media have been attributed to reducing physical activity in children [46, 47]. Many children currently do not reach recommended daily physical activity targets.

Child dietary behaviour has been suggested as being strongly influenced by factors at home [48]. This is because children have less autonomy in food choice than adults. Parent's intake of fruit and vegetables is associated with child's intake, and foods available at school have a large impact on food choices [49, 50]. Furthermore, aggressive marketing of unhealthy food are also believed to have contributed to a preference of unhealthy foods [51]. Children are receptive to marketing efforts and a study in Boston found that for every increase of 1 hour of television viewing was

associated with an increased calorie consumption of 167kcal per day [52]. This study also found that children who watched more television ate more of the food commonly advertised on the television [53, 54]. These tend to be energy dense, processed foods. The above factors combined with a poor nutritional education have significantly contributed to our changed eating habits.

2.1.2.3 Environmental causes

The built environment is increasingly being acknowledged as having a significant influence on obesity levels [55]. The built environment that people live in and interact with in their daily lives can be conducive to healthy or unhealthy lifestyles. Positive influences of the built environment on population health include: good access to healthy food, greenspaces and physical activity centres (e.g. gyms and leisuire centres); highly mixed land use type; infrastructure that supports active travel; and buildings designed to reduce sedentary behaviour [56]. Conversely, negative influences from the built environment on population health have been attributed to the increasing prevalence of obesity. The quality of the built environment has been found to correlate with the socio-economic status of an area, poor quality built environments being associated with more deprived areas [57]. The prevalence of obesity in more deprived areas has been encouraged by the low cost of the energy dense foods for people who have limited budgets [58]. Less deprived areas are characterised by a larger choice of foods with healthy foods often priced at more competitive rates [59–61].

The term 'obesogenic environment' has been coined to describe an area of the built environment where there is an abundance of unhealthy food sources, poor availability of fresh and healthy foods, and lack of access to green spaces and walking opportunities [14, 42]. This definition is believed to capture a more complete picture of the environment that affects our eating habits and level of physical activity. Obesogenic environment research brings together evidence that links aspects of the built environment that are believed to influence the prevalence of obesity; access to food and the opportunity to be physically active [62, 63].

Access to food

The literature has defined access to food using different terms. 'Food dessert' and 'retail food environment' are the two most common terms that have been coined as a way to measure the influence of the access to food on obesity.

The term 'food desert' was coined to define areas that lack in healthy food provision [64, 65]. Food deserts tend to be characterised by a lack of supermarkets or shops selling fresh fruits and vegetables. Studies that investigate 'food deserts' use supermarkets and larger grocery stores as a proxy for access to healthy food. Findings on the relationship between 'food deserts' and obesity have been mixed. In the USA, supermarkets have moved out of deprived areas and therefore 'food deserts' have been correlated with deprived regions. In developing countries where food insecurity is a widespread issue, the term 'food desert' is being used to define areas where it is difficult to obtain affordable, nutritious and safe food [66]. In contexts where access to healthy food is limited and food resources are stretched, 'food deserts' are an important method of defining the environment [67]. However, identifying areas that have poor access to healthy food may not be the most appropriate way to understand the role of access to food on obesity in the UK [65]. This may not be the most appropriate way to define access to food as supermarkets sell unhealthy food at affordable prices too [68]. Shier et al. [43] proposed that any food outlet type should be considered in assessing accessibility to food. Regardless of the type of food outlet, Shier et al. [43] found an increased density of food outlets was associated with a higher Body Mass Index (BMI) and concluded that it is the availability of all food that is of concern.

The 'Retail Food Environment' (RFE) is thought to encompass a more accurate picture than 'food deserts' of how people are influenced by the availability of unhealthy or lack of availability of healthy foods. The RFE is characterised by the opportunity to buy food and this includes the price and variety of food [69, 70].

A systematic review undertaken in 2015 [71] found limited evidence of an association between the RFE and obesity. There were 71 studies identified which contained 65 cohorts and overall the study quality was low. There was a direct association between fast food availability and obesity in lower income children. However, the paper states

that this finding should be interpreted with caution because of the overall poor quality of studies. Obesity promoting environments have been found to correlate with the socio-economic status of an area; obesogenic environments are more strongly correlated with deprived areas [57]. Many studies show that diets predominantly made up of fruit, vegetables, wholegrains and lean meats are associated with higher costs than energy dense, nutrient poor foods [72–83]. The over availability of unhealthy food is believed to be promoting an unhealthy lifestyle and an over consumption of energy dense foods. An initiative commissioned by Tower Hamlets Council, London in 2011 looked to develop a management framework for managing the number and location of hot-food takeaways in the county. The report concluded that there is a strong case to develop a number of planning policies that will improve the quality of the food environment in Tower Hamlets.

The influence of the built environment on childhood overweight or obesity has not been as widely investigated as adult obesity [84], but local planning and education authorities are particularly concerned with the prevalence of obesogenic environments surrounding schools. This is because they believe obesogenic environments are a threat to children's health [85]. However, the literature is inconclusive as to whether children's exposure to the RFE is linked to childhood obesity [86, 87]. American studies have the most prevalent inconsistencies amongst the literature. Some studies support the view that students with fast-food restaurants near their schools are more likely to be overweight or obese [31, 88, 89]. Where others have found no correlation between the proximity of fast food outlets to schools and childhood obesity [60, 90, 91]; or even with diet [92]. Beauluc et al. identified an association in deprived areas but none in more affluent areas [58]. With regard to UK research, there is also uncertainty about the extent of the effect of proximity to unhealthy foods and obesity. Currently, council planners believe there is harm caused from an overabundance of food outlets in an area and deny planning applications. However, appeals against blocked planning applications are often successful due to a lack of evidence of the harmful effects. Local councils currently do not yet have the evidence they need to prevent additional fast food outlets from opening. In order to help planners and councillors, it is important to show the cause and effect of access to unhealthy food along children's routes to school.

Opportunity to be physically active

Features of the built environment not only influence people's diet and relationship with food, but also how physically active people have the opportunity to be. Features of the built environment such as: greenspace; building design; location of workplaces and schools; pedestrian and bicycle facilities; land-use mix; and dedicated facilities for recreation, such as sports centres, all influence physical activity levels [93]. The presence or absence of these features can have a positive or detrimental effect on physical activity levels. In the same way that poorer quality food environments tend to be located in more deprived areas, features of the built environment that support physical activity are often absent in more deprived communities [94]. Although the built environment can provide us with opportunities to be physically active and make choices that are less conducive to obesogenic behaviours, the current condition of the built environment has been attributed to contributing to the rise in obesity [95].

Good access to greenspace is thought to be a valuable resource for increasing physical activity levels and reducing obesity and overweight [96]. The literature suggests that better access to greenspace improves general population health [97, 98]. Access to greenspace has been found to decrease with deprivation [94] and vary when analysed with respect to ethnic and religious groups [99]. In contrast to this, Jones et al [100] reported that accessibility to greenspace was better in more deprived areas but those residents had more negative perceptions and were therefore less likely to us the greenspaces accessible to them. A systematic review on greenspace and obesity by Lachowycz and Jones [101] found 60 papers that investigated the impact of access to greenspace and obesity. A third of studies that were included in this systematic review found no relationship and two studies found a negative relationship. The paper recommended that "Developing a more solid theoretical socio-ecological framework which considers the various correlates and interactions between different types of greenspace and health would help both formulation and interpretation of the body of research" [101]. A paper written by Cummins and Fagg in 2012 [102] supported this view in calling for better evidence on the utility of greenspace before greenspace interventions are developed. Following the publication of this paper, a UK study found that living in a greener area was associated with healthy weight outcomes [103].

The influence of the built environment on physical activity levels also extends to building design and the location of work and school premises [93]. Studies have shown that obesity is also related to physically inactive job and school environments that the majority of people now experience in developed nations [104]. Policies and interventions have been developed to reduce overall sedentary behaviours [105] and schools in particular have been identified as being a place to increase young people's physical activity levels [106]. Successful methods of increasing children's physical activity levels include focusing on the Physical Education curriculum and including regular activity breaks and working with families [107]. Furthermore, schemes such as the Active by Design programme, supported by Design Council in the UK seeks to advise policy makers, town planners, transport planners, architects and employers on how building design can increase physical activity and help people to make healthier choices [108]. For example, identifying places where schemes can be implemented or designing staircases that are more attractive to look at and use.

Whether people walk to school or work may depend on the built environment [40]. Places that have infrastructure that support active travel such as cycle lanes and pedestrian facilities have higher rates of active travel, compared with areas that do not have these supportive facilities [109, 110]. However, overall, the number of children actively commuting to school has decreased in developing countries in the past 40 years. Many different factors have contributed to the decline of actively travelling to school. Increased car ownership has seen driving become the most popular way to travel and infrastructure being designed with the adult car driver in mind. This infrastructure design does not encourage child active travel (AT) and the prevalence of automobiles on the road has contributed to rising parental concerns for their children's safety along their route to school. Parents and carers concerned for their child's safety tend to drive them to school. Also, a greater number of working mums has contributed to reduced AT to school [111]. This is because it is more convenient for parents to drop off their children at school on their way to work. A further contributing factor is that children do not necessarily go to the closest school anymore. Moreover, in many cases, children live too far away from their school to walk or cycle.

School AT rates are relatively low in developed countries, and have been proposed as a way to significantly increase the amount of physical activity that children undertake

daily. Increasing active transport rates in school children has been suggested as an effective public health intervention in reducing overweight and obesity in children. Sirard et al. [112] suggested that increasing the prevalence of active transport to school helps to maintain a healthy lifestyle and decrease the prevalence of obesity. Furthermore, children who AT tend to be more active in general [113]. For children who do not AT, it has been found that they tend not to compensate for this lack of physical activity elsewhere in their day. Active Travel to school has many benefits in addition to increasing individual physical activity rates. AT has been positively associated with social interactions [114] and reducing localised congestion [115]. AT has been promoted and encouraged through infrastructure interventions. The idea being that making AT more accessible will increase physical activity and reduce obesity. However, although anecdotal evidence is that these schemes are successful, these interventions have not been robustly evaluated [116]. Overall, there has been a large decrease in the number of children actively commuting to school, however, it should be noted that rates of AT vary by country. Countries with good infrastructure that encourages AT, such as the Netherlands and the Scandinavian nations, have higher rates of AT compared with countries that have poorer infrastructure such as the UK. A further consideration in understanding AT rates is that transport surveys often do not report walking or cycling that are linked to other modes i.e. a combination of travel modes so the rate of AT is likely to be underestimated. AT to and from school is a time when older children will usually be unaccompanied. Care will need to be taken when reporting results on exposure to unhealthy food as so not to undermine the encouragement of walking to school, that is, walking to school means that you have a greater opportunity to buy unhealthy food.

A diverse mix of land-use, in their home neighbourhood, also encourages people to be more physically active. Studies have shown that locating residential areas near to shops, offices, schools and other services is related to increased walking and physical activity levels [117–119]. Living near to amenities encourages people to walk to them, rather than to choose a sedentary travel mode. Furthermore, communities that are well designed to encourage the residents to be physically active, tend to provide a visibly interesting built environment which further promotes walkability [120, 121].

Many features of the built environment support physical activity. However, features such as crime and traffic may contribute to residents making choices less conducive to being physically active [122]. In areas where there are higher rates of crime people can feel unsafe and are therefore less willing to walk in their neighbourhood and allow their children to play outside. One study suggested that the greatest increase in physical activity would be seen in a population if everyone felt safe during the night [123]. For children, the prevalence of violent crime is a significant environmental barrier for being physically active outdoors [124] and parents feel that it is easier to protect children from danger by doing indoor activities [125]. High volumes of traffic also influence physical activity levels in the same way. Areas with high volumes of traffic tend to be less aesthetically pleasing and are not safe environments for children to play in and/or be physically active in [126]. Implementing reducing speed limits, traffic-calming devices, cycle lanes and wider pavements have been found to increase physical activity.

Features of the built environment can provide opportunities to be physically active and support lifestyle choices that are good for people's health. Conversely, there are some features of the built environment that can be barriers to people being physically active in their day-to-day lives. The causes of obesity are complex and multifactorial. And the role of the environment on obesity is an important facet to understand. No one mechanism will provide a solution to the obesity epidemic, but making positive modifications to the built environment does provide the opportunity to help people to make decisions more conducive to a healthy lifestyle on a large scale. For example, improvements made to pedestrian facilities have been shown to increase rates of active travel [127, 128]. The studies that have observed increases in active travel, following an intervention or change in policy, have not evaluated the effect that this has on reducing obesity. However, the literature unequivocally states that increasing the amount of physical activity that people do and providing environments that support healthy food choices can help contribute to reducing obesity on a large scale.

2.1.3 Implication of obesity
The causes of obesity are complicated and there are many implications of being obese that effect the individual and society. The physical implications include high blood

pressure, coronary and respiratory problems, type 2 diabetes, types of cancer and a number of other serious physical health implications [51, 129, 130]. Obesity can also have detrimental effects on a person's well-being, quality of life and their ability to earn [9]. There is particular concern for the issue of childhood obesity because of their vulnerability as a demographic. Reilly et al. [131] dismissed the idea that childhood obesity is largely a cosmetic problem from conducting a systematic review of the health impacts of childhood obesity. This systematic review found that childhood obesity has significantly adverse physical, psychological and social impacts on children [132–137]. Furthermore, many studies support the idea that obese children are at much higher risk of becoming obese adults than children and adolescents who have a healthy weight [3, 4]. It has been reported that 70% of obese children grow up to be obese adults [32] and research has found that the first 9 and 24 months of a child's life are important landmarks in predicting whether a child will be overweight at age 8 [138]. The prevalence of obesity is also a burden to health systems and economies around the world [139, 140]. A systematic review by Withrow concluded that treating obese patients accounted for up to 2.8% of a country's total health costs. Furthermore, medical costs for obese individuals are over 30% higher than people with a healthy weight [141]. In the UK, the cost of obesity to the National Health Service (NHS) is around £5 billion per year and the cost to society and the economy has been predicted to rise to £49.9 billion per year by 2050 [9]. The Foresight report that calculated these figures was written by the UK Government Office for Science. This report predicted that the economic cost of obesity will not go down in the foreseeable future but it advised that curbing the rate of increase would be a realistic goal. The report only considered the direct economic cost of rising obesity levels. The report does not consider the wider impact of social consequences and it is proposed that economic productivity could be affected.

2.1.4 Combatting obesity

Obesity rates are continuing to rise and show no signs of waning [142]. To date, there have been no reported population level reductions in obesity levels. The complexity of the issue suggests that complex solutions are required. The Foresight report concludes that any effective interventions will need to be long term, and consider the context (i.e. social, economic, environmental and political). Such interventions can be

termed complex interventions. The UK's Medical Research Council (MRC) describes a complex intervention as an intervention that contains several interacting components [143]. The MRC guidelines suggest that complex interventions may be aimed at different levels: at individuals, at community level and at whole populations [143]. Designing interventions to target different levels is an important consideration in combatting obesity. This is because research has suggested that individual level changes may not result in population level changes [144, 145]. Furthermore, changes in practice, such as the adoption of guidelines or policy changes, are not necessarily mirrored by behaviour changes at individual level [146].

A recent paper by Rutter et al. [8] acknowledged that many of the biggest public health challenges are resistant to policy and intervention, including obesity. The paper called for public health research to move from simple linear models of cause and effect to a focus on complex systems [147]. Rutter et al. suggests that public health researchers should conceptualise poor health and health inequalities as:

> *"outcomes of a multitude of interdependent elements within a connected whole. These elements affect each other in sometimes subtle ways, with changes potentially reverberating throughout the system."*

In the context of obesity research, the prevalence of obesity is a result of multiple systems (food, employment, transport and economic etc.) that influence energy intake and expenditure [8].

2.2 GIS and Its Use in Obesity Research

This section introduces Geographic Information Systems (GIS) and how it has been used in health research to date. It also provides a critical evaluation of research methods and signposts the way to improve on this previous work.

2.2.1 Geographic Information Systems

John Snow famously solved the mystery of the London cholera outbreak in 1854 by demonstrating the association between cholera-related deaths and geographical

location. Dr Snow plotted disease cases and water pumps on maps and drew concentric circles around the points to determine where the highest density of cases was occurring. From his findings, the Broad Street pump was removed and this led to an end of new cases in the area, proving that drinking water from the pump was causing many people to become ill [148]. This technique was adopted by physicians and medical geographers and proved to be very useful in understanding how disease and spread was associated with space.

Epidemiology grew as a discipline but researchers were limited by technology and data collection. However, the large scale migration of people, exposure to previously unknown diseases and the opening of medical schools and journals in the late 19th and early 20th centuries transformed medical geography in to an analytical science. Technological advancements allowed the research area to grow and governments began providing funding for this type of research [148].

By the mid 20th century, the field of medical geography had been influenced by several important thinkers such as August Hirsch, Jacques May and Tom Spies [149]. May was fascinated by the relationship between geography and infectious disease and he compiled numerous maps of disease distribution. He is most famous for recognising disease was the interaction between pathological and geographical factors [148, 150].

As computer technology advanced in the late 1960s, maps could be generated more quickly. As a result, disease diffusion could be mapped more thoroughly and analysed using more sophisticated methods.

GIS was developed in the 1960s and has evolved from its primitive stages, of being a means of digitally displaying data, to now facilitating robust and sophisticated spatial analysis. Not only has medical and healthcare research benefitted from GIS, but environmental research and management, infrastructure planning, advertising, insurance and governmental organisations are all industries that have benefitted from the development of spatial analysis.

For medical geographers and epidemiologists, GIS has furthered our understanding of both differences and similarities between populations [148]. GIS has allowed epidemiological studies to identify the clustering, distribution and spread of diseases on a much larger scale. This has optimised the planning of interventions and the monitoring of outcomes and success rates. GIS has been fundamental in successfully

monitoring and controlling onchocerciasis in Guatemala, trypanosomiasis in Africa, and malaria in Israel and Mexico [148]. In the UK, GIS has been used in epidemiological studies such as investigating alcohol-related harm to population health [151, 152], the prevalence of obesity [56, 153], and geographical variations in health inequalities.

In addition to epidemiological studies, GIS is also being used in health research to analyse complex questions in public health research. For example, identifying and evaluating risk factors for cancer [154–156], environmental health risks in Europe [157], evaluating the impacts of climate change on human health [158] and the impact of traffic-related air pollution on health [159]. GIS is also used in community health profiling, which is the mapping of multiple variables of a population's health. Profiles usually include health outcomes as well as factors that are known to influence health. Health profiling links all of the included factors with the spatial locations of infrastructure such as restaurants, supermarkets, roads and hospitals [160]. Accurate linking allows for the general relationships between health and the built environment to be identified. Community health profiling has been used to help planners and policy makers with decision making [161, 162] and it is also a useful tool in conveying health messages to the public.

2.2.2 GIS in obesity research

GIS is being used in public health research to evaluate the influence of the built environment on obesity levels. As aforementioned, the influence of the environment on obesity is of increasing interest to researchers because of the opportunity to modify the built environment to encourage healthy lifestyles and make the built environment less conducive to obesogenic behaviours. Furthermore, a better understanding of the influence of the environment on obesity will aid policy makers and practitioners to develop successful population level policies and interventions.

The literature is equivocal on whether the built environment is influencing obesity rates. Accessibility modelling is the basis of these studies but the methods employed to measure accessibility are diverse. Studies have investigated access to entities within food deserts and obesogenic environments. These entities are usually things like fast food outlets, supermarkets and greenspace.

2.2.3 GIS in accessibility research

Accessibility can be defined from multiple perspectives but geographic accessibility is defined by: 1) the distance people live from services, 2) how good public transport links are to the health services and 3) how long it takes to travel to such services [163]. In a broader research context, modelling accessibility has been used by public and private sectors to inform and investigate public access to relevant services or infrastructure. Business planners are able to find the most suitable location for new sites by locating the optimum location for customers, workforce, accessibility and infrastructure. Policy makers and public health researchers are able to evaluate public access to health care services, greenspace and public amenities. The results of these studies are important as they have been used to inform the best location of new services where it can serve the optimum population.

In addition to quantitative modelling of accessibility, survey data have been used to evaluate people's *perception* of their access to services. This may be just as important as geographical access as how a person perceives accessibility to a service is correlated with how often or how much they use a service [126, 164]. Such data has been used by studies to measure accessibility for different travel modes such as by bus, by car and by foot. This is an important consideration when measuring accessibility because not only does an individual's primary mode of transport vary, the main mode of transport used to access different facilities will be different. For example, accessibility by foot would be an important consideration for the modelling of accessibility to a high street supermarket. However, accessibility to a hospital or large shopping centre, the primary mode of transport would be motorised. Road datasets are available where the road type is defined, as are speed limits and one way roads and dead ends. This means that accessibility models can represent motorised transport systems very well. When modelling routes that could be taken on foot, the road network should contain footpaths and thoroughfares that are used by walkers. This is so that the accessibility model can represent the true road network in the most accurate way that is available to researchers.

In public health research, accessibility methods are used to measure access or exposure to entities within the environment that can be beneficial or detrimental to our health – both mentally and physically [165]. Studies investigating the influence of the

environment on obesity have focussed on exposure to supermarkets, fast food outlets, alcohol outlets; and access to greenspace and physical activity facilities.

The RFE has received increased attention from researchers and policy makers as it provides the opportunity to make positive changes to the built environment to encourage healthy lifestyles. The literature contains conflicting evidence on the influence of the environment on obesity because of the use and reporting of heterogeneous methods used to measure exposure to the RFE [69, 71, 166, 167]. Wilkins et al. [70] identified five dimensions of methodological diversity:

1. food outlet data;
2. extracting food outlet data;
3. defining food outlet constructs;
4. geocoding methods;
5. access metrics.

These five dimensions make up the Geo-FERN reporting checklist which is proposed to support how researchers report and interpret studies on the RFE using GIS [70]. Discussion of the five dimensions are below and make reference to the checklist.

2.2.4 Food Outlet Data

Studies investigating exposure to the RFE most often use secondary food outlet datasets. This secondary data are either administrative data or commercial datasets. Administrative datasets are collected by local authorities (LA) or governments [168, 169] and are free to obtain. Commercial datasets are produced by companies such as the Yellow Pages [170] and InfoUSA [171] and these data cost to obtain. Newer sources of secondary food outlet data include open source datasets such as Ordnance Survey's (OS) Points of Interest (PoI) [56] and Google Street View[TM] [172].

Administrative food outlet data have been in many studies investigating the influence of the exposure to the RFE on weight status [173–180]. Studies in the UK have used administrative data obtained from LAs [181–183]. This data has been documented as being difficult and time consuming to obtain. This is because LAs routinely collect data on food outlets for their own licensing and monitoring purposes but many LAs do not consider that this data can be used for secondary purposes.

In LAs, food outlet data are collected by two different departments; Environmental Health (EH) and Local Planning Authorities (LPA). These departments have different aims and agendas and therefore do not necessarily collect the same information on food outlets. Local Authority EH work with the Food Standards Agency (FSA) to enforce food safety and food hygiene across the UK. The FSA is an independent public body that is responsible for food safety and food hygiene across the UK. The FSA was born out of the Food Standards Act (1999) [184], which looks to ensure high quality food throughout the food chain. The data collected helps LAs to hold food outlets accountable for providing a quality service. LAs send data to the FSA through the Local Authority Enforcement Data (LAEMS) and Food Hygiene Rating Scheme (FHRS) systems. LAEMS is a web based system run by the FSA. Although LAs are not required by law to contribute to LAEMS, LAs are encouraged by the FSA to participate. The FHRS system runs in all areas of Wales and participation was set out in law (Food Hygiene rating Act, 2013) [185]. LAs are responsible for uploading the data they have collected whilst inspecting businesses to the FHRS.

LA datasets are not required by law to have internal quality checks. However, the food outlet data collected by EH is field validated as the premises are visited by officers from the food safety teams. Participation in the FHRS and LAEMS improves consistency in data collection and also data management, as data must be uploaded in a predefined format [186]. This encourages LAs to have stringent and comprehensive data collection and data management protocols in place. Furthermore, the data should be up to date as businesses are required, by law, to register themselves with their local council [187]. This applies when an outlet is opening for trading and also closing down. Although many studies have stated that LA data is highly accurate, these studies have also stated that obtaining data from LAs is time consuming and complex [182, 188–190]. However, the detail of the challenges faced in obtaining food outlet data from LAs has not been discussed at length [190]. This study found that obtaining data from LAs was time consuming, and that data custodians were not always forthcoming with sharing data.

There is currently no formal guidance for researchers on how to obtain data from LAs, and LAs are not always forthcoming in adhering to data requests [182, 188, 190]. Therefore, more recently, alternative secondary data sources such as OS PoI are becoming more popular [152, 191, 192]. Relatively little is known about the accuracy

of these open source data [188] but these data sources are more convenient to obtain than administrative data. However, the methodological inaccuracies are often not documented in the literature. Wang et al [193] suggested that caution should be taken when using secondary data and that field validation should be undertaken to assess the validity of the secondary data. One study recommended that POI data was a 'viable' alternative to primary source Local Authority data [188]. The results of the sensitivity analysis between POI and LA data were conservative. With most POI outlet types being only moderately associated with LA data [188]. The sensitivity cut-offs here were reported as: 'poor' <30%; 'fair' 31-50%; 'moderate' 51-70%; 'good' 71-90%; 'excellent' >91%. These cut-offs were used as they have been used elsewhere in the literature [194]. However, the origin of these cut-offs are from a systematic review and meta-analysis that explores agreement in quality-of-life assessments between patients and physicians [195]. More research needs to be undertaken to understand the influence of the quality of different datasets on analysis [196]. A small number of studies have shown that different data sources can lead to different conclusions regarding associations between exposure to the RFE and area-level demographics [197]. Furthermore, Hobbs et al. [10] reported differences in the size of associations between exposure to the RFE and weight status when comparing OS PoI and administrative data. However, 11/12 of the tested associations were non-significant and therefore the overall conclusions of the study were not affected. This study used large area geographies as definitions of neighbourhood and it has been proposed that these geographies are not necessarily representative of what people identify as their neighbourhood [198].

Many studies have used traditional measures of quality [196, 199, 200] which use an authoritative dataset as the gold standard. The widely cited parameters for the quality of geographic data are defined by the International Organisation for Standardisation (ISO) [200]. These are: completeness [201, 202]; logical consistency [203]; positional accuracy [204–206]; temporal accuracy [207]; and thematic accuracy [208, 209].

The food outlet datasets not being complete is of concern to researchers. Most data sources do not include mobile food outlets or market stalls [188, 190]. Market stalls have been reported as being an important source of food for some communities [210] and mobile food outlets may be important for children. If the datasets do contain these outlets, they are usually registered to the owner's home address and the locations

where mobile outlets trade are usually not recorded either. Not having a record of these outlets could underestimate food access for some subsets of the population [70].

Temporal mismatch between the food outlet and the health outcomes is also a limitation of using secondary datasets as it is not always possible to know exactly when the data was collected [70].

2.2.5 Extracting Food Outlets

The different ways in which outlets are categorised within a dataset has led to homogeneous methods for extracting food outlet data from data sources. Usually researchers obtain a dataset and have to process the data to extract food outlets that are being investigated. Several different methods of extracting food outlet data from datasets were identified by Wilkins et al. [70]. These were extracting outlets; a) using proprietary classification codes [43, 211–213] and; b) searching attributes in the data [214, 215]. Using propriety classification codes means that the researcher must select which classifications to include and exclude. Some classification codes, such as those in OS PoI and some administrative datasets, are highly detailed and permit researcher to even classify by takeaway type (e.g. fish and chip shop, Indian takeaway etc). However, other classifications such as the Food Standards Agency are far broader which does not permit researchers to subsample the outlets. In datasets where detailed classifications are not provided, it can be challenging to extract outlets that may or may not sell food [216]. Studies have adapted the Food Marketing Institute definitions to create eleven categories of food outlets, including bakeries and pizza shops [217]. Different methods of classifying food outlets has likely contributed to the differing conclusions being drawn. It should be noted that classification systems have tended to be developed based on the data sources available to them [189]. Studies have used local council lists [218], commercial directories [219], government agency lists [220] and field-based classifications. Lake et al. [189] developed a comprehensive food environment classification tool that has been field validated. Burgoine et al. [219] created a three point classification system of the longitudinal environment and White et al. [221] developed 19 categories for shops selling food. Some American studies used the classification system developed by Wang et al. [220]. This classification system has been adapted from the North American Industry Classification System and

the Food Marketing Institute definitions. The second method was implemented by Cummins et al. [222] to investigate the association between the presence of McDonald's outlets and neighbourhood deprivation. The literature however, now suggests that all food outlets should be included in studies of the RFE and not just specific food outlet type or chains [43].

Wilkins et al. [70] reported that many studies provide no documentation of how they selected the food outlets from the main dataset. This means that it can be challenging to interpret the results in the context of what aspect of the RFE has been investigated. On the other hand, there are a number of published studies that document the extraction of food outlets in a comprehensive way that can be reproduced [43, 211].

2.2.6 Defining Food Outlet Constructs

Defining food outlet constructs refers to what outlet types are included in the studies investigating the RFE. Different constructs have been discussed in earlier sections and these include food deserts and aspects of the RFE. Food desert studies have used supermarkets to represent sources of healthy food [90, 129, 221]. However, as aforementioned, in the UK, identifying food deserts has been found to be unsuitable for characterising the RFE. This is because supermarkets offer a wide range of unhealthy foods in addition to healthy food. In particular, smaller supermarkets have been found to stock a greater proportion of unhealthy food than healthy food [216]. Other studies have investigated other aspects of the RFE such as exposure to "fast food outlets" [175, 223, 224]. The literature has diverse food outlet constructs and this makes study comparison very difficult. The basis of this research is to inform policy and the homogeneity of the evidence base makes it very difficult for policy makers to interpret findings. Wilkins et al. [70] identified that there is currently little evidence to guide construct definitions and suggested that researchers should use established definitions to allow for inter-study comparison and improve the reporting of the food outlet constructs. Policy makers and local government should also be considered as LPAs classify outlets in different ways to the food outlet classifications that have been discussed. Current practice is that LPAs are unable to distinguish between healthy and unhealthy hot food outlets (classified as A5 premises). Furthermore, from a LPA perspective, the takeaway chains, such as McDonald's and KFC are registered as

restaurants (A3 premises) [225]. This is a particularly important perspective to acknowledge if researchers are going to provide evidence to policy makers that can help to make positive changes to the built environment.

2.2.7 Geocoding

Once the food outlets have been extracted from the source data, some datasets need to be operationalised before they can be used in GIS software. This means converting address data into coordinates or other geographical identifiers such as Unique Property Reference Numbers (UPRNs). These are unique identifiers for every man-made entity in the UK, developed by OS [226]. The process of converting address data to spatial locators using algorithms is geocoding. This allows for food outlets to be mapped and spatial analysis to be undertaken.

Geocoding has been reported as a complex process and Wilkins et al. [70] proposes that this is why many studies do not document their geocoding process in GIS-RFE studies. The spatial resolution of the geocoding can be variable in these studies with studies geocoding to a) areal units e.g. postcode/ZIP code units, b) street segments, c) land parcels, and d) address points. The choice of which spatial unit is used to geocode the outlet data will influence the spatial accuracy of the geocoded outlet. However, there is generally a trade-off between high match rates and spatial accuracy [182]. The software used and the reference data chosen will also influence the quality of the geocoding [227]. Different sources of reference data have been found to produce different accuracies of match rates [228].

2.2.8 Access Metrics

2.2.8.1 Static and dynamic exposure zones

Once the data has been operationalised (e.g the data has coordinates or UPRNs attached to the food outlets), researchers can compute the exposure metrics. Many studies have defined a single exposure environment such as exposure to the RFE in the home neighbourhood, school or work premises. For example, there has been much interest in exposure to fast food within 400m of school premises and whether this influences fast food consumption and BMI. Researchers measure exposure to the

desired service (i.e. food outlets or greenspace) with counts [229–231] or density scores [152, 232].

Other studies have evaluated exposure to entities in "dynamic" spaces such as the commute to work or the journey to and from school [169, 181]. Harrison [11] highlighted the importance of capturing a dynamic exposure environment that children experience on their journey to school. People move around and interact with different environments throughout their day e.g. home, commute, work, school, hobbies. In order to better understand causes of childhood obesity, there is a need to investigate the environment that children experience throughout their day, not just one area of exposure that children experience [233].

2.2.8.2 Euclidean buffers
Euclidean buffers around homes, workplaces and schools to measure exposure are widely reported in the literature. Researchers calculate exposure within these buffers and calculate whether higher exposures are associated with higher BMIs [169]. The home environment [178] is a popular environment to define and investigate whether or not this is linked to obesity. Other studies have focussed on the surrounding area of a school or workplace [234, 235]. However, Euclidean buffers have been reported as misrepresenting the accessibility of an area [163, 236, 237]. Euclidean buffers do not take into account the underlying topographic nature of the built environment and the topology of road networks i.e. how people can access a space.

2.2.8.3 Network buffers
Network buffers [178, 235] and cluster analysis [234] have also been explored as a way to evaluate the impact of exposure on obesity. Although this is seen as a more methodologically complex approach, network buffers are static environments that do not capture the diversity of environments that individuals may be exposed to in their daily lives.

2.2.8.4 Dynamic activity spaces
Studies that measure dynamic activity spaces such as the work or school commute use a variety of methods. Studies have used Euclidean distances, network analysis and GPS data to investigate exposure to the RFE within dynamic activity spaces. This

contributes to the cumulative exposure that people experience in their day to day lives [181].

Studies have undertaken network analysis along a road network to model journeys made to work or school. The network analysis undertaken is shortest network analysis but this has been reported as an unreliable proxy for the environment that a child is exposed to along their route to school. This may be because shortest network routes (SNR) are not sufficient to capture route choice of children that walk to and from school. Furthermore, the road networks that have been used in previous studies may not always be complete as pedestrian footpaths are not always available.

GIS generated exposure environments along routes to and from school have been compared with GPS exposure environments. GIS generated network routes were reported as being a suitable proxy for the *distance* that a child walks, but not for the environment they are exposed to. The GIS routes taken by children have been found to be weakly correlated with their GPS routes [11]. However, GIS route modelling in past investigations have only used shortest network analysis along a road network and route choice has been found to be more complicated than this [47, 114, 238]. This suggests that children do traverse along short paths but not necessarily along main roads. There is the potential to improve GIS network modelling by increasing the complexity of GIS modelling. GIS modelled routes offers the potential to routinely generate routes to model population level exposures. This could be a powerful tool in informing public health interventions.

To model commuting journeys or routes to school, there are several different road networks that are available to researchers to use. Many studies use authoritative datasets, usually created and provided by a national mapping agency. However, in recent years, the growth of OpenStreetMap (OSM) has given researchers an attractive alternative.

With advancement of the internet and GPS, the volume, availability and nature of geographic information has been subject to change over the past decade. This has allowed for members of the general public to engage in collecting and uploading geographic information on to the internet [239]. Data collected in this way is known as Volunteered Geographic Information (VGI) [240] which is also a version of crowdsourcing [241]. VGI has generated vast volumes of geographic data because

anyone can select a region on the Earth's surface and attach georeferenced information to the regional area [240]. OSM is one of the most popular examples of VGI [242, 243]. OSM was set up in 2004 with the main goal of creating a free and editable world map [244]. The project has been highly successful in realising its goal and encouraging participation. As of January 2016, 2.4 million people were registered as users [245]; a number which has doubled in the past 3 years.

OSM has presented many opportunities to researchers and users of the data as it is free, timely and includes data types that are not necessarily captured in authoritative datasets [242]. There are, however, some concerns over the quality of VGI as there is no one accepted methodology for evaluating VGI. Those studies evaluating OSM have used a breadth of approaches and analysis techniques. Many studies have used traditional measures of quality [199, 200, 246] which use an authoritative dataset as the gold standard. The widely cited parameters for the quality of geographic data are defined by the International Organisation for Standardisation (ISO) [11]. These are: completeness [201, 202]; logical consistency [203]; positional accuracy [204–206]; temporal accuracy [207]; and thematic accuracy [208, 209].

Although the ISO defined quality indicators have been used in many studies evaluating the quality of VGI, it has been proposed that traditional quality measures are not suitable for evaluating the quality of VGI [248]. This may be because high quality authoritative datasets are not available for comparison but also because of how and by whom VGI data is collected. Different types of OSM contributors were identified by Neis and Zipf [249] and found that only 5% of OSM members contributed to the project considerably; with 38% of members making at least one edit. In addition to this, Arsanjani et al. [250] compared the volume of contributor's data with the quality of the data. Their research found that a small number or contributors share the most accurate data but many members contribute data small amounts of data at the minimal quality. The novel nature of VGI data collection and its associated uncertainties means that quality checks are important. Barron [251] summarised studies that have evaluated the quality of VGI without an authoritative reference datasets which have been developed as an alternative. Such methodologies included investigating the data's history [252] and lineage [253]; evaluating the quality of OSM features by analyzing objects with more than 15 different versions [254]; analysing the relationship between tag assignment and number of contributors [255]; and a model

to estimate the uncertainty of geometric measurements of vector objects has been introduced [256]. Furthermore, the iOSMAnalyser was developed by Barron et. al. [251] and uses over 25 methods and quality indicators to calculate a quality measure that is based only on data history.

A challenge of evaluating the quality of VGI datasets with traditional methods is that VGI datasets are now, in many parts of the world, more complete and accurate than authoritative datasets. This has been shown to be true particularly in crisis and disaster management [257]. However a systematic study in 2015 of VGI evaluative methods stated that "research is still far from providing concrete answers and methods regarding the evaluation of VGI quality" [258]. Moreover, Barron et. al. stated that "absolute statements on data quality are only possible with a high quality reference dataset as a basis for comparison" [251] .

A systematic study carried out by Sehra et al. [259] concluded that OSM is "quite developed and mature". This review proposed that a development in evaluating OSM will be to investigate its use in routing. In UK studies, Ordnance Survey's Integrated Transport Network is widely used in accessibility studies. The ITN is a trusted, authoritative dataset as it is produced by trained specialists. There are, however, expensive licensing fees attached to using the dataset and it has been proposed that OSM data is very good in the UK [204, 208, 260]. Furthermore, studies from around the world have found the same [207, 261]. On the other hand, some studies have drawn more tentative conclusions about the quality of OSM datasets [202, 203, 262–264]. This study will evaluate the quality of OSM roads and footpaths in Wales to see whether it is an appropriate alternative in route modelling to authoritative datasets.

2.2.8.5 GPS
An alternative method of investigating exposure environments is to use GPS data. It is now accepted that GPS data is the most accurate way to record data on where people travel in their day to day lives; more detailed and accurate than surveys and current GIS methods [265]. The literature reports that GPS devices are the best way to record exposures because they have become reliable and accurate to a high spatial resolution [266]. Therefore, GPS studies are a reliable way of evaluating an individual's exposure environment by capturing where a person travels in their day [267, 268]. Where a person may change location, these environments are known as activity spaces. They

can be defined as "the local areas within which people move or travel in the course of their daily activities" [269]. Modelling activity spaces for populations and not just static measures is a relatively unexplored subject in obesity research. It is, however, an important avenue to explore as people are mobile, and fixed measures do not adequately represent individual level daily exposures [270]. There are a few studies that have explored children's activity spaces [268, 271, 272]. Some studies have used GPS data to record the route that a child takes to school. The studies have looked specifically at children's school commute; investigating physical activity levels and exposure to the 'retail food environment' (RFE) along these routes [11, 47, 56, 273–275].

As with any methodology, there are associated limitations. When using GPS devices, data loss can occur due to signal drop-out [266] and participants may forget to wear or recharge their devices which can result in missing data [276]. Collecting GPS data is resource intensive and therefore cohort sizes tend not to exceed several hundred participants [274, 277]. However, collecting GPS data can be time consuming and costly [265]. Procuring the GPS loggers, ensuring that participants charge the devices and obtaining the loggers from participants once the study finishes are all challenges of collecting GPS data [278]. Furthermore, several studies suggest that when participants wear GPS loggers, their behaviour is influenced by an awareness of the device [279, 280].

Cleaning GPS data requires technical expertise, and therefore there can be difficulty when interpreting the data [281]. Studies have been criticised for not differentiating between travel modes when calculating environmental exposures from GPS routes. For example, researchers may only be interested in foot and cycle journeys as passengers in motorised vehicles will potentially have lesser opportunity to buy food on their journey. Cetateanu et al. [281] stated that this problem has largely been ignored.

Several studies attempted to extract travel mode using crude criteria such as travel speed [275, 282] to differentiate between travel modes. Harrison et al. [11] manually cleaned the GPS data to extract travel mode. Other studies have attempted to develop more robust cleaning algorithms in order to be able to extract meaningful information from the data, such as travel mode.

A number of studies have produced algorithms for automated cleaning of GPS data [283–287]. However, there are no uniform standard amongst studies but they all tend to split the raw GPS data into smaller partitions which represent single journeys [281]. Following this, analysis can be undertaken on these smaller segments. There are three types of cleaning algorithms that have been developed. These are criteria based methods, supervised algorithms and unsupervised algorithms. Criteria-base methods are the least technically demanding of these approaches but they have been successfully implemented other studies [288, 289]. This approach has been based on arbitrary criteria and criteria supported in the literature (e.g. average walking speeds). Using arbitrary criteria has been criticised however, as it can introduce bias in to the algorithm [281].

The supervised and unsupervised methods are based on machine learning principles. Supervised algorithms use manually classified data in order to make inferences about unknown data. However, these inferences could be incorrect if an inappropriate training dataset is used [70]. Unsupervised algorithms do not require training data and therefore overcome this challenge of supervised cleaning methods. Arguably the most sophisticated method of GPS cleaning was published by Cetateanu et al. [281] whereby the researchers had developed an unsupervised algorithm to clean GPS data and distinguish between travel modes. This was undertaken on the Personal and Environmental Associations with Children's Health (PEACH) dataset containing the GPS locations of a sample of children in Bristol, UK [277].

2.2.9 Conclusions

The quality of food available to residents, the opportunity to be active and the affordability of healthy foods all make up the environment that people experience and are exposed to in their day to day lives. GIS is an important tool in evaluating the effect of the effect of the environment on obesity.

Burgoine et al. [153] highlighted the need for a methodologically consistent evidence base that will allow for inter-study comparisons to be undertaken [32]. Until a higher quality evidence base is developed, it is not possible to draw conclusions to support policies aimed at regulating food environments [14, 65, 290]. There are currently no published longitudinal studies that investigate environmental exposures and obesity in

the UK. It should also be noted that in addition to different methodologies being used, the evidence base is from many different countries where social and economic factors may effect behaviour. Although there has been an increase in research efforts to evaluate the effect that the environment has on obesity, methodologies and datasets have varied between studies. This has resulted in conflicting evidence which limits the conclusions that researchers are able to make about how the environment effects obesity.

In summary, it is important to further evaluate whether the environment that children are exposed to on their route to school has a detrimental impact on their health. If children are being encouraged to AT to school, it follows that the environment they are exposed to should not be detrimental to living a healthy lifestyle. For the children who already AT, it is essential to ensure that their route to school will not be detrimental to their health.

The use of GIS in public health research is somewhat limited in its applications compared with more sophisticated applications in more traditionally geographic studies such as environmental management and climate change research. As appreciation for the place of GIS in public health research grows, there is a call for more sophisticated techniques to be used that are applied in other disciplines. This would improve understanding of how populations vary spatially and how geographical variations influence population behaviour and health.

Creating high resolution spatial patterns of the exposure environment that children are subjected to will help researchers to better understand the environment that children are interacting with. It will be just as important for researchers to find out whether exposures to the RFE along children's walking routes are associated with BMI measurements. Understanding what will help reduce obesity at a population level is just as important as discovering individual level drivers.

2.3 Data Linkage

2.3.1 What is Data Linkage?
Over the past 40 years, vast amounts of individual level, routinely collected data have been recorded. This data is usually collected primarily for health or administrative

purposes [291], but can be used for secondary research purposes. These data have been identified as a "goldmine" of information that can help to inform healthcare provision and services by linking together datasets to build up a picture of people's lives [292, 293]. For example, in the UK, different bodies record many different sources of medical information such as General Practitioner (GP) (primary care physician) and hospital records, administrative data collected by the public sector, Biobank data and clinical trials data. Researchers can join or "link" together these datasets, providing they have a common identifier, such as; name, address, gender, date of birth or a specific ID number [291]. This joining of different datasets is known as data linkage. Linked data is particularly used in epidemiological studies, health research, planning and evaluating health care and informing clinical decision-making [293, 294]. Data linkage is a powerful, population-level resource and a cost-effective alternative to cohort studies [295, 296]. This has led to health data linkage research becoming a rapidly growing research field and it is acknowledged as having the potential to revolutionise health research [293].

2.3.2 Concerns of Data Linkage

There are, however, concerns that identifiable data is available without any consent [297]. EU and UK Legislation does protect the availability of private information but public trust is relied upon to permit the access to identifiable data. Scandinavian countries have a culture where it is expected that information gathered with public funds should be used to generate research that benefits the population [297]. The populations there understand that researchers look after their data and that it is managed by strict protocols [298]. However, with recent major data breeches in commercial companies, there are growing concerns that personal information is not safe in custodian's hands. There are worries of using identifiable data, lack of transparency and access by commercial organisations [299, 300]. It is imperative that individual identities are protected in research projects; but in some cases, the restrictions on data accessibility have been suggested to be severely limiting to researchers and policy makers. Furthermore, the use of tax payer's money to complete expensive data collections would suggest that datasets should be made available to researchers so that society can benefit [37]. UK linked data research projects are often

funded by public budgets so it is expected that research should benefit society and the economy [301].

There is a small risk that patient data could be re-identified [302]. There have been cases where information has been disclosed from Google street view and papers publishing maps that include identifiable, sensitive information; such as routes to school [303]. It is important that the dissemination of results is considered to safeguard against releasing identifiable data. Exeter et al. [297] set out a framework which advises how to effectively distribute results without disclosing individuals' identities. Particularly when generating spatial indices where a location linked to an individual could be disclosed. When visualising spatial results, consideration should be given to maps and images that they are not disclosive either.

Research has shown that a map containing point data can disclose original addresses that the point data is associated with; even when there is minimal other spatial data included on the map [304]. Identity may also be disclosed when multiple visualisations of the same masked dataset are created [305]. Studies have also found that when data is visualised on a map, preserving anonymity can distort results [306]. Guidelines do exist for the reporting of sensitive and identifiable spatial data. However, it is a researcher's responsibility to protect individuals' identities while still preserving the pattern of the results. Different methods have been developed to preserve identity when visualising study results on a map. The process of disguising identity but endeavouring to maintain spatial patterns is known as geomasking. A review by Zandbergen [307] evaluated the five geomasking methods that have been developed. These were random direction and fixed radius, Random Perturbation within a Circle, Donut method [308, 309], Gaussian Displacement and Bimodal Gaussian Displacement. Despite these novel methods being developed, there is no one accepted approach to geomasking. There are even other alternative approaches that have been used in studies to preserve anonymity and present meaningful results. These include generating synthetic data [310], using software agents [306] or using flexible aggregation methods [311]. This study will consider the issue of preserving anonymity with high regard whilst creating meaningful representations and visualisations of the results.

2.3.3 What Can Be Done to Alleviate Risk?

UK legislation protects the misuse of identifiable data and restricts access with the Data Protection Act 1998. There are exemptions that allow certain data to be used without consent and this includes the monitoring of cancer and infectious disease. EU data protection legislation is currently under review and there is the possibility that consent requirements set out by the EU may become stricter and therefore incompatible with studies that are currently being undertaken. In support of this EU legislation, there are also calls in the UK have stricter controls when using identifiable data for secondary purposes. Data must be managed in a strict and sensitive way by trained staff so to minimise risk. In Australia, data scientists working with sensitive data, work hard to communicate the importance of their work and to involve the public in the direction of their research [312].

Legal processes look to protect data from being misused and ensure protection of individuals' identity. Data linkage centres adhere to legal requirements and work under stringent conditions in order to ensure that data are stored safely and protected. In addition, data linkage centres commit to only undertake research that is of public benefit [313].

Health researchers in the UK must balance generating meaningful conclusions with data privacy concerns [314]. Much attention has already been given to this issue but there remains a way to go before a satisfying balance is reached [297, 315–318]. Data linkage systems have developed a number of different approaches to protect identity. Some data linkage databanks may hold identifiable data but other databank remove any data that may be disclosive [291]. However, prior to working with linked data, researchers remove identifiers before analysis is undertaken. In some systems, identifiers are pseudonymised, whereby identifiers are replaced with a code. Data banks such as SAIL anonymise data by removing all personal identifiers or many health researchers generate their analyses using areal aggregation units which minimises the ethical pressures that researchers face regarding the maintenance of individuals' anonymity [317, 319].

Whatever the context of using health data, security and care should be taken very seriously by data custodians. This is because the uses of health data are powerful and not utilising health data is a serious problem. The benefits and limitations of analysing

and linking health data should be kept in perspective whilst there is a move towards responsible reuse of data [314].

2.3.4 Data Linkage Centres

Lyons et al. stated that linked data has the potential to revolutionise health research and highlighted the value and need for more longitudinal data [293]. In the UK, the National Health Service (NHS) uses linked data to provide consistent care for patients. The data is also used to inform policy and planning, to calculate treatment costs and for audits [302]. Outside of the NHS, health data linkage is used in public health monitoring and management, academic research and commercial organisations [302]. Linked data is used in research all around the world by organisations that exploit the rich data sources to undertake sophisticated analyses, providing unique insight to individual and population level health. There are a number of established data linkage systems around the world. Examples are the Oxford record linkage study, Population Health Research Network in Australia [320, 321] and Secure Anonymised Information Linkage (SAIL) Databank in Swansea, Wales [292, 293].

The Consumer Data Research Centre (CDRC), in the UK, link big data to investigate influences on obesity. The CDRC also set up the Obesity Strategic Network with the goal of inform positive change that is evidenced by population level research [322]. Linked health data have also been used as the basis for many other overweight and obesity studies. For example the Norwegian Institute of Public Health have investigated social influences on obesity[138, 323, 324]. To date, no research has been undertaken linking commuting routes with health data.

2.3.5 Linking Environment and Health Data

There are a small number of studies that have linked routinely collected health data with environmental indices to evaluate the impact of aspects of the environment on health [325] . Villanueva et al. [326] linked routinely collected health data, for 0-64 year olds, with GIS generated built environment variables. Rodgers et al. [327] devised a methodology to effectively and anonymously link 1.3 million addresses to the local environment. A novel approach was taken by Fone et al. [152, 182] whereby routinely collected health data was linked with GIS generated change in alcohol outlet density.

Individual level data is important because the environment and social influences affect individual health and well-being [37]. It is important that social and health data is georeferenced because our health is affected by where we live. Health outcomes have been noted to be directly related to environmental influences, including water and air content, exposure to hazardous materials, occupation and tobacco smoke [37]. By linking health data and giving it spatial relevance, powerful analyses can be undertaken which have the potential to provide new insights into the spread and cause of disease [37]. The potential for using individual data is not being exploited because of very valid concerns where sensitive data could be used for detrimental purposes. Recommendations have recently been made on the importance and power of using individual level health data in research, whilst safe guarding personal information and protecting identities [314].

2.3.6 The Secure Anonymised Information Linkage Database

2.3.6.1 What is SAIL?

The SAIL Databank at Swansea University is a world class data linkage system that brings together billions of anonymised, person-based records. This allows trusted researchers to link together health, administrative and environment data in a secure environment [292, 328]. SAIL contains routinely collected health and administrative datasets about the population of Wales [329]. These datasets include outpatient and primary care data, GP records and a national child health dataset. The detailed data stored in SAIL allows researchers to answer important societal questions that may otherwise not be possible to answer [330].

To provide a safe haven for sensitive data, before data arrives in SAIL, identifiable details are anonymised by a trusted third party (TTP), NHS Wales Informatics Service (NWIS) [331]. The process of anonymisation is called the split-file process [331, 332]. This process describes how identifiable details are removed from the dataset and replaced with an Anonymous Linking Field (ALF) or a Residential Anonymous Linking Field (RALF). An ALF is assigned if the data is individual level records or a RALF assigned if the data is address level data.

2.3.6.2 Research Ethics and Information Governance

This study involved the analysis of retrospective and routinely collected anonymised health data, it was not possible to obtain individual consent from each individual. The anonymisation process within the SAIL databank means that consent does not have to be given by each individual.

Data is anonymised in the SAIL databank by a trusted third party, NWIS. NWIS encrypt NHS numbers and address identities without accessing any clinical content. Following encryption by NWIS, a second stage of encryption occurs at Swansea University. This means that I was not able to identify individuals or addresses. The SAIL analysts prepared my data by working to ONS guidance to make sure that no individuals could not be identified through small numbers.

Datasets are split into demographic data and clinical data. A system linking field is used to ensure that data can be re-joined at a later time. The demographic data contains person-based variables such as name, gender, date of births and postcode. The clinical data are sent directly to SAIL and the demographic data are sent to NWIS. The demographic data are pseudonumised by NWIS and are allocated an Anonymous Linking Field (ALF) or a Residential Anonymous Linking Fiels (RALF). An ALF is a unique 10 digit number assigned to each individual in a dataset. This data is then sent to SAIL where it can be linked to the clinical data using the system linking field [292]. The split file method is used to overcome any confidentiality and disclosure issues that are related to storing and managing sensitive data.

Chapter 3

Reflections from Collecting Local Authority data

3.1 Introduction

This chapter documents a methodology designed to collect food outlet data from Local Authorities (LAs) in Wales, United Kingdom and reflections on the experience. LA data has been reported as the most accurate data to represent the retail food environment [217, 333]. However, collecting LA data has been documented as being a lengthy and sometimes complicated process [182, 188] .

Collecting data from many different LAs results in data collection across a range of contexts. Secondly, the causes of problems targeted by the data collection method may differ from one LA to another. Furthermore, when collecting data from LAs, Burgoine concluded "assumptions that are often made at the outset of a research project, and that will, more often than not, impede upon the research design, or extend the duration of the research in the long run." I wanted to minimise the assumptions that I made in the design of my data collection method by founding the design in theory. I designed my data collection methodology based on the lessons learned from previous studies [182, 190].

In the United Kingdom (UK), LAs are the administrative body in local government. LAs play a central role in governing the UK as they provide local leadership and services for the communities they serve [334]. Services run by LAs include social care, education, housing and planning, environmental protection, waste management, roads and transport, cultural and leisure. LAs are required by law to collect data on the services they provide. The data that central government require LAs to collect are recorded in 'The Single Data List' [335]. These data are used to monitor service provision and use, advise policies and inform decisions made by LAs and central

government. LAs therefore collect and store vast quantities of data on a diverse range of subjects. With regard to food outlet data, two different departments collect data on food outlets within LAs; Environmental Health (EH) and Local Planning Authorities (LPA). These departments have different aims and agendas and therefore do not necessarily collect the same information on food outlets.

Local Authority EH work with the Food Standards Agency (FSA) to enforce food safety and food hygiene across the UK. The FSA is an independent public body that is responsible for food safety and food hygiene across the UK. The FSA was born out of the Food Standards Act (1999) [184], which looks to ensure high quality food throughout the food chain. The data collected helps LAs to hold food outlets accountable for providing a quality service. LAs send data to the FSA through the Local Authority Enforcement Data (LAEMS) and Food Hygiene Rating Scheme (FHRS) systems. LAEMS is a web based system run by the FSA. Although LAs are not required by law to contribute to LAEMS, LAs are encouraged by the FSA to participate. The FHRS system runs in all areas of Wales and participation was set out in law (Food Hygiene rating Act, 2013) [185]. LAs are responsible for uploading the data they have collected whilst inspecting businesses to the FHRS.

LA datasets are not required by law to have internal quality checks. However, the food outlet data collected by EH is field validated as the premises are visited by officers from the food safety teams. Participation in the FHRS and LAEMS improves consistency in data collection and also data management, as data must be uploaded in a predefined format [186]. This encourages LAs to have stringent and comprehensive data collection and data management protocols in place. Furthermore, the data should be up to date as businesses are required, by law, to register themselves with their local council [187]. This applies when an outlet is opening for trading and also closing down. Although many studies have stated that LA data is highly accurate, these studies have also stated that obtaining data from LAs is time consuming and complex [182, 188–190]. However, the detail of the challenges faced in obtaining food outlet data from LAs has not been discussed at length [190]. This study found that obtaining data from LAs was time consuming, and that data custodians were not always forthcoming with sharing data.

There is currently no formal guidance for researchers on how to obtain data from LAs, and LAs are not always forthcoming in adhering to data requests [182, 188, 190]. This chapter has contributed to attenuating this issue by reflecting on my methodology designed to collect food outlet data from Welsh LAs.

3.2 Methodology

A logic model was developed to provide a clear guide on the data request method [336]. A review of relevant literature was undertaken so that the context of how LA data has been gathered and used could be understood (Chapter 2). Developing the logic model helped me to identify causal assumptions and key uncertainties. The assumption that has most often been made by previous studies is that LAs will respond to a general enquiry by email. From reading the literature, I knew that this could not be assumed. Previous studies also suggested that data collection would be a lengthy process. I therefore based the logic model on recommendations from the CHALICE study [151, 152, 182] and Burgoine et al [190] (Figure 3.1). This meant that initial contact was made with a senior figure in Public Protection Wales (PPW) and significant time was allocated to data collection.

Data was collected from LAs from December 2014 to January 2016 (summarised in Figure 3.2). Firstly, the Director of PPW was contacted for a list of senior EH figures for each LA in Wales. I chose to approach EH directors through the Director of PPW as I thought that EH directors would be more likely to take note of my request, give quicker responses and potentially more complete datasets if the request had the permission and support of the Director of PPW. An email request was sent in December 2014 to each EH contact in every Welsh LA. The email gave a brief explanation of my research project and requested access to the data their department held on food outlets (Appendix 1.).

Attributes for all premises that sold any type of food were requested from all Welsh LAs. The 7 attributes requested are documented in Appendix 1. and included Unique Property Reference Numbers (UPRNs), address level XY coordinates and outlet classification. These specific attributes were requested in order to create a standardised

dataset for Wales. The aim of this project was to investigate exposure to the 'retail food environment' and therefore all food outlets were requested. This was because each LA was expected to use a different food outlet type classification system [182]. Obtaining all food outlets meant that categories of different food outlets (e.g. unhealthy food outlets) could later be extracted according to a standardised classification system [153].

After one month the email was resent to the contacts who had not replied to the initial email sent December 2014. If no response was received from this email, EH general enquires were contacted by email. For EH departments that were unable to/did not respond to the request within 1 month, the Local Planning Authority (LPA) was contacted. The same email request was submitted to the LPA as was sent to EH and the original senior contact (Appendix 1.).

After multiple requests over a period of 12 months, three LAs had failed to respond to any requests. In response to this, Freedom of Information (FOI) forms were submitted. FOI requests are in place to give the public access to information held by LAs. LAs are legally obliged to respond to the request within 20 working days. Although LAs are not obligated to provide the data that has been requested, for LAs that had not responded to any other emails, the FOI meant the LA had to at least engage with the research team.

Following a timely discussion with the Policy Officer Regulatory Services at the Welsh Local Government Association, the policy officer suggested that LAs may require an incentive to encourage the sharing of data; advising that it would be worth him sending around the original email edited to include incentivising text. The email now explained the outcomes of the evidence and usefulness of the research; such as providing public health practitioners and policy makers with evidence to limit the growth of unhealthy food outlets. The policy officer also confirmed the disjointed communication between licensing and planning. As a result of the policy officer's help two more LAs provided data.

Once the data were received, duplicates were removed, and each LA dataset was organised longitudinally, based on two assumptions: 1. When there was no record of the premises closing, the outlet was assumed to still be trading; 2. When there was no entry for the premises opening, the outlet was presumed open since the database began.

The quality of the data that was collected has been evaluated based on a quantitative spatial data quality analysis that is recommended in the literature [196, 246]. This evaluation is documented in Chapter 4.

Situation	Inputs	Activities	Outputs

Challenging to obtain LA data for spatial studies

No longitudinal food outlet dataset for Wales

No formal guidance on how to obtain data from LAs

Researcher

Email Contacts

Literature-informed data collection strategy

Define data required

↓

Email data request to WCHE contacts

↓

Email sent to EH

↓

Email sent to planning

↓

FOI submitted

Longitudinal food outlet dataset for Wales

Figure 3.1 Logic model for the methodology of collecting food outlet data from Welsh Local Authorities
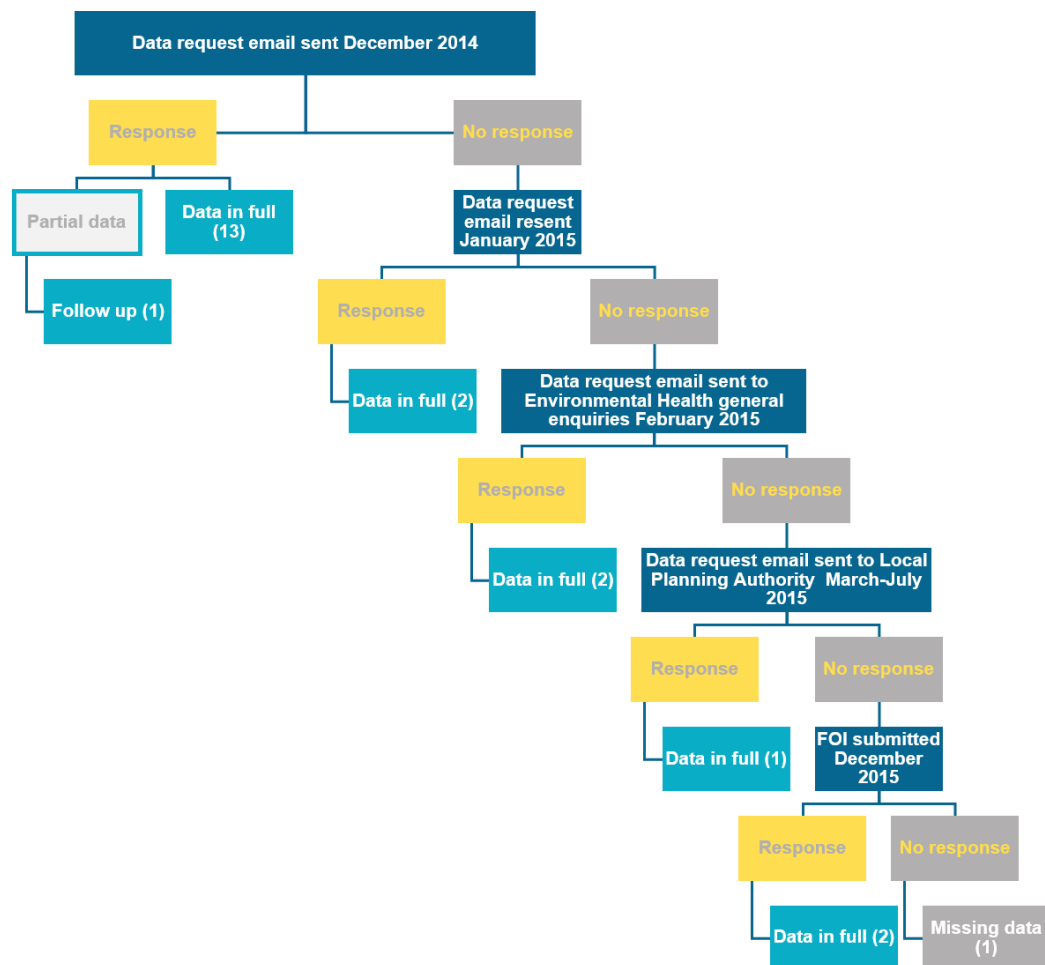
Figure 3.2. Flow diagram of the data collection process. Number of Local Authorities that responded at each stage is reported (n)

3.3 Results

From the 22 Welsh LAs, 21 provided food outlet data. One LA was unable to provide any data. Figure 3.3 shows the attributes that the LAs were able to provide. Four LAs were able to provide all attributes that were requested. Trading times was the most poorly provided attribute where only three LAs provided this. All LAs provided the addresses for the outlets and nineteen LAs provided outlet-type categorisation. This categorisation was usually a LA-defined categorisation which differed between LAs. Five LAs provided the FSA classification scheme for the outlets. The LAs that provided their own classifications were more detailed than the FSA classification. The LA classification schemes differentiated between cafes and restaurants and often

takeaways were defined separately, with some LAs even defining the type of take away the outlet was. Twenty LAs provided data suitable for creating a spatio-temporal dataset.

The length of time it took to obtain the data from LAs varied from two days to eighteen months. Over 70% of LAs provided data by the original senior contact forwarding the request to a colleague, who then sent the data to the research team. Before data were received from the twenty-one LAs, there were lengthy delays in some LAs responding to the request. Several LAs misunderstood the email request.

One LA did not respond to any requests and responded to the FOI describing that they did not have the resources to run the request. However, in their response the LA demonstrated they had misunderstood the request. The way that the LA had interpreted the request would require an unreasonable amount of work to complete. Subsequently, a second FOI was submitted which removed mention of the research for which the data would be used, and simply requested a historical database of all food outlets. This FOI request was successful. A second LA that had not responded to five previous requests responded to an FOI request within 20 days with a complete dataset. A third LA responded to an FOI by stating that the LA had limited resources. Instead of an appropriate dataset they sent food registration forms from 2005-2016; a refined FOI request was then submitted, resulting in receipt of a full dataset. A FOI was also submitted by one LA, without consultation with, and on behalf of the researcher using the initial email request. The data provided from the FOI requests had to be heavily processed in order to be ready for use in research.

LA	Addresses	XY Coordinates	UPRN	Trading Times	Outlet Type Categorisation	Longitudinal dataset

Figure 3.3. Attributes that were provided by Local Authority

3.4 Discussion

3.4.1 Main Findings

Collecting food outlet data from LAs was a time consuming and challenging process. LAs have different capacities and processes for storing, collecting, and sharing food outlet data and this made collecting data from 22 LAs a complicated and lengthy process. However, starting the data collection journey through a known contact did prove successful in obtaining timely data from over 70% of Welsh LAs. Developing a logic model was an effective tool, but more needs to be done for the LAs that cannot respond to data requests in a timely fashion. Researchers can face many difficulties when collecting data from LAs such as: knowing who to contact; receiving no response; experiencing long time periods before receiving data; being unclear in defining what data are required, LAs not having the time or man power, and historic data not always being maintained. Burgoine [190] commented on how researchers and LA have different agendas and when the perspective of LAs are considered, many of the challenges are justified. LA resources are increasingly stretched and this includes a reduced workforce. Less time and people to contribute to engaging with additional parties such as researchers becomes less of a priority when LAs are working hard to

keep services running smoothly and efficiently. The already stretched resources are likely to have caused the lengthy delays in collecting data from all Welsh LAs.

In addition, delays were also caused from LAs misinterpreting the data requests. Although *all* food outlets were requested, the mention of obesity and fast food outlets in describing the research objectives made the request unclear. In some cases, only fast food outlets were originally sent. In response to this, the email was edited and the mention of fast food outlets was removed from a second request. A second misunderstanding came from a LA where the request resonated with some research that the LA had undertaken themselves. Unfortunately, instead of this being an aide, the LA sent GIS data that was tailored to their investigation. These were fast food outlets within 400m of schools in the LA and buffers surrounding the school. On returning to the LA to request more data, no further responses were received, despite several email requests.

These findings are consistent with what has been reported in the literature [182, 190]. The responses from LAs were not consistent in Wales. Many LAs responded promptly with complete datasets but for some, obtaining data was onerous. For the LAs that were challenging to engage with, email was not an effective form of communication. Emails can be easily ignored, deleted or forgotten about, whereas phone calls or face-to-face contact proved to be a worthwhile effort in obtaining data in a timely fashion.

3.4.2 Strengths

Developing a logic model to guide the data collection from previous studies meant that "lessons learned" from CHALICE and Burgoine et al. were incorporated into this protocol. Implementing a collection strategy that has evolved from previous work meant that ineffective ways to obtain data did not have to be repeated. Making contact with LAs through a senior figure in the LA proved to be a very successful way to obtain the data. With over 70% of LAs providing data through the senior contact, the weight of authority supporting the request greatly improved the likelihood of a quick response and the LA complying. The senior figures are more likely to have contact with researchers and policy makers outside the LA. Having this 'bigger picture' in mind and understanding the benefit of being involved in research conducted outside the LA resulted in the LA responding to the request swiftly.

The logic model helped me to develop a comprehensive understanding of past studies and to consider the priorities and perspectives that LAs may have had towards my request. LAs are a unique and valuable data source and in order to be able to effectively use their data, researchers must carefully plan their data collection process. The process documented in this chapter has proven to be more complicated than just sending an email request off and receiving a complete dataset.

3.4.3 Limitations

An alteration to the data collection protocol would be to follow up emails that have not yet been responded to with phone calls, and adding read receipts to the emails. This may have been helpful in encouraging LAs to respond to the email. A further point of change would be to ensure that the initial email was clear in explaining exactly what data are needed and for what purpose. Receiving feedback on the content of the email would have been beneficial for identifying any aspects of the email that were unclear to people without any spatial research background. Furthermore, the departments that took the longest to respond were limited in their capacity because of restricted resources. The one LA that was unable to provide any data gave the same reason. A great barrier faced by LAs is central government funding cuts. Some departments have reported significant losses to their workforce [337]. Currently, LA resources are stretched and thinned and it was unreasonable to expect immediate access to LA data without at least informing LAs of the benefits that come with partnering or data sharing with academic institutions. Being aware of this context is important and should be a consideration for researchers when approaching LAs.

3.4.4 Implications

The work documented in this chapter has highlighted the unique environment of LAs and how data collection must be a considered process. Planning must go in to it and as part of the bigger picture, LAs and academics have to work more closely together. A closer relationship would be mutually beneficial as LAs will have a better understanding of how important the contribution of their data is and also have access to new research and highly skilled colleagues. For researchers, having a better understanding of how LAs work and their internal structure means that researchers

should: promote the value of LA data; work more closely with LAs; and engage with LAs more readily by actively looking to share skills and knowledge with LAs. Researchers and LAs should look to work together more closely. Researchers have a responsibility to highlight how LA data can be used in research, particularly in health research that is conducted with the aim of improving population health.

Improved communication between departments collecting the 'same' data is also recommended to reduce duplication. A standardised system between EH and planning departments would reduce time and monetary expenditures of LAs. A standardised system that includes high resolution temporal and spatial parameters across the UK would be a powerful tool for public health research. However, the focus of the current system is on implementing legislation effectively and high resolution spatial locators are not currently prioritised in EH data collection. Schemes such as the INSPIRE directive [338] look to complement transparency and open data policies. The INSPIRE framework is used by the government organisation Defra but not by FSA. I received data in varying formats, which highlighted that in spite of the FSA and LAEMS, there are different primary data collection standards for collecting and storing data on food outlets in Welsh LAs. The differences amongst these storing and reporting standards, have shown that food outlet data collection in Wales is inconsistent, not only between LAs, but also within LAs. By following a united framework such as INSPIRE, not only would LA data be consistent throughout the country but inter-country comparison could be undertaken. This recommendation is a long term recommendation as it would be time, money and resource intensive to migrate and merge data collection systems. As a short to long term recommendation, communication between EH and LPA should be encouraged in order to reduce duplication of efforts in data collection, data management and data analysis. This approach would reduce cost and time expenditures of LAs. There are two main departments within LAs that hold information on food outlets; EH and LPA. However, the two departments have different priorities and demands of the data. EH focuses on the licensing and law enforcement of food outlets and LPA are concerned with the classification of premises type and whether the premises type fits their planning agenda for a particular area. This has meant that the different departments have diverse efforts and priorities on what data they collect about food premises. This suggests that there is a likelihood of

overlapping resources and time being spent on collecting and managing the data in LAs.

It is the responsibility of academic institutions to highlight to LAs the 'goldmine' of data that they store and how it can be utilised in innovative and novel research that will help to improve public health. EH departments are in place to improve public health outcomes. In order for EH departments to be best equipped to do this, policy and interventions need to be informed by high quality research. A key role for academic institutions will be to give more incentives to LAs to prioritise and engage with researchers. Incentives such as access to research results, skill training and involving LAs in research teams would work towards a closer partnership between researchers and LAs.

LAs have a unique position where they record longitudinal data that can be used to evaluate whether changes in the environment over time have had an impact on public health. Communication between academics and LAs must be strongly encouraged so that LAs can know of the benefits of their data for research to improve health outcomes. Evidence on improved health is needed for policy development [339, 340] and LAs can benefit from having access to academic evidence. It should be promoted amongst EH and LAs in general that their data can be used to develop high quality and innovative interventions to improve public health. Relevant employees in LAs should be aware that spatial analysis is increasingly being recognised as an important tool in monitoring environmental effects on public health and it is the responsibility of researchers to help non-specialists in understanding this.

3.5 Conclusions

LAs are rich data sources and an effective method of engagement needs to be identified if researchers are to make the most of data that is being collected by these authoritative bodies. Careful planning and an understanding of the nuances for LAs will help in obtaining data from LAs. However, if researchers are to use data from LAs regularly, much work has to be done in building relationships with LAs. Researchers need to endorse the importance of LA data, be considerate of the LA's

perspective, encourage consistent data capture across geographic regions. The promotion of relations between LA and academia which will lead to collaborative and synergistic relationships.

Chapter 4

The Spatial Quality of Secondary Food Outlet Data Sources

4.1 Introduction

The previous chapter documented the collection of food outlet data from an administrative data source; Welsh Local Authorities (LAs). This chapter evaluates the spatial quality of the LA data and also documents a sub analysis of four LAs. The sub analysis investigates the difference between LA data and two other secondary data sources that are becoming more widely used in exposure studies. The secondary data sources are the Food Standards Agency (FSA) [341] and Ordnance Survey Points of Interest (PoI) [342]. Secondary data sources are widely used in GIS-RFE research but not much is known about their quality [188]. This is because they are collected for purposes other than research. Researchers use this data for a secondary purpose as food outlet data in studies exploring the influence of the RFE on obesity.

The spatial quality of the LA dataset was analysed using the five parameters for determining the quality of geographic data that are defined by the International Organisation for Standardisation (ISO) [200, 247]. These are: completeness [201, 202]; logical consistency [203]; positional accuracy [204–206]; temporal accuracy [207]; and thematic accuracy [208, 209]. The sub analysis assessed the accuracy of the secondary datasets by analysing percentage agreement, positive predicted values (PPV) and sensitivity analysis.

The aims of this analysis were to:

1. Assess the spatial quality of the LA longitudinal food outlet datasets;
2. Document the limitations associated with food outlet data sources;

3. Evaluate the most appropriate dataset to use as food outlet location data source in exposure modelling.

Administrative datasets are becoming less popular to use in GIS-RFE studies because acquiring administrative data can be time consuming and substantial data cleaning and processing often needs to be undertaken [190]. Instead, alternative open source secondary spatial data sources are becoming more widely used in exposure modelling. These data sources are free to download and easy to access online. However, the quality of these datasets has not been widely tested and evaluated [10, 188]. Furthermore, some studies have found secondary data sources to be of poor quality [182]. In addition, the limitations associated with the secondary datasets are not well documented in the studies in which they are used. The evaluation of the two secondary datasets contributes to the discussion on using these data sources in exposure modelling. A recent paper suggested that secondary data sources are a "viable alternative to council data" [188] and this chapter seeks to contribute to this conversation.

The use of Ordnance Survey's (OS) Points of Interest (PoI) dataset has been previously validated for a Local Authority in England. However, a far greater geographical area is evaluated here (the whole of Wales) and an additional secondary dataset has been evaluated. The methodology implemented in this study has been reproduced from Burgoine et al. [188] so that results between the geographical locations can be directly compared.

4.2 Methodology

4.2.1 Data Sources

4.2.1.1 Local Authority

LAs are the administrative bodies of local government in the UK. They are responsible for collecting food outlet data for planning, environmental health and business purposes. The organisational structure of LAs was described in Chapter 3.

4.2.1.2 Food Standards Agency

The FSA is an independent Government department that is responsible for ensuring that food sold in the UK is safe and trustworthy [341]. The FSA train and advise Local Authorities in running the Food Hygiene Rating Scheme (FHRS). This scheme involves LAs inspecting businesses that sell food and scoring the level of hygiene. Data collected for this scheme is stored by the FSA on behalf of LAs participating in the FHRS. The FSA publish the food hygiene ratings on their website for all food outlets in the UK [343] .

4.2.1.3 Points of Interest

The PoI dataset was created by and is now maintained by Point X, a company jointly owned by OS and Landmark Information Group [344]. PoI markets itself as "the most comprehensive, location-based directory of all public and privately-owned businesses, education and leisure services in Britain" and contains over four million records [345]. PoI is compiled from more than 170 data suppliers.

4.2.2 Data Collection

4.2.2.1 LA Data

The LA data were address level, longitudinal data. Chapter 3 documents the collection and processing of this data.

4.2.2.2 FSA Data

The FSA data for south Wales were downloaded from the FSA website [346]. The FSA have made postcode level point data available for all outlets that they hold information on in the UK. XML files were downloaded and processed in R to create geocoded data. The data is updated daily and grouped by local authority.

4.2.2.3 PoI Data

The PoI dataset was downloaded for south Wales from Digimap [347] in a csv format [347]. PoI is an address level dataset that is updated quarterly.

4.2.3 Data Preparation

The LA data was received in different formats (see Chapter 3). The formats were unified by converting all LA datasets to csv format. The datasets were imported into R (version 3.3.3), duplicates were removed and text entries were converted to lower case. This was to minimise the difference between the datasets. Two assumptions of the LA data are that: 1. Where there is no entry for when the premises closed, the outlet is open for trading; 2. Where there is no entry for when the premises opened, the outlet has been open since the database began.

The PoI and FSA datasets were imported and prepared for analysis in R. Text entries were converted to lower case. Duplicates were not removed so to replicate current practice of using the datasets as they are downloaded [188].

Food outlets that would sell unhealthy food were extracted from all three data sources using the classifications that were included in their respective datasets (Appendix 2.). These outlets were defined as any outlet where any unhealthy food could be purchased; including bakeries, convenience stores and sweet shops. The outlets were then reclassified according to criteria developed by Lake [189]. These are described in Table 4.1. Restaurants, pubs and specialist shops such as butchers and fishmongers were excluded from the dataset as they are not frequented by school children. Mobile food outlets were also excluded from the dataset because the home address of the owner was given in place of the retail location. The unhealthy food outlets were used to define the food outlet construct of this project to represent the RFE [43, 70].

Table 4.1 Food outlet classification (developed from Lake et. al. [22])

Outlet Type	Examples	Sells
Café/Coffee Shop	• Costa • Starbucks • Independent	• Ready-to-eat food to be taken away • High sugar drinks
Bakery	• Greggs • Independent	• Freshly baked savouries/bread, pre-made sandwiches, baked sweet products and branded products
Takeaway	• Kebab shop • McDonalds • Chip shop • Sandwich shop	• Hot food ordered and paid for at the till • High calorie foods
Supermarket	• Tesco, Asda, Morrisons, Sainsburys • Kwiksave, Netto	• Large, departmentalised, self-service food store selling food and household goods • Smaller, self-service food store selling a limited range of food and household goods for greater convenience
Convenience	• Corner Shop • Newsagents • Petrol Station • Off-licence	• Groceries, newspapers, magazines, snacks, drinks and tobacco products.
Entertainment	• Amusements • Bowling	• Snack foods and drinks
Non Food Outlet	• Boots Pharmacy • Poundland	• Snack foods and drinks
Specialist	• Sweet shop • Thorntons	• Shops that do not fall under the category of convenience or confectioners as sell only bought in sweets

4.2.4 Local Authority Spatial Quality

The ISO defined standards for geographic information have been reported for the 21 LAs that provided food outlet data. These are completeness, logical consistency, positional accuracy, temporal accuracy and thematic accuracy.

4.2.4.1 Completeness

Completeness was assessed by the number of spatial and temporal attributes that the LA provided in the dataset and the number of attributes that were provided as a proportion of the total dataset. The spatial attributes were address, UPRN and address level XY coordinates. Address is an implicit spatial reference, UPRN is a lookup and XY coordinates are an explicit spatial reference.

4.2.4.2 Logical Consistency

Logical consistency was evaluated by assessing the consistency of the data for each LA and between LAs by reporting attributes that were provided and how these attributes were defined by each LA.

4.2.4.3 Positional Accuracy

Positional accuracy was reported as the type of spatial locators that were present in the dataset and the extent to which these were provided. XY coordinates were required to be able to undertake analysis later in the study but if address locators were provided, these implicit locators could be used for geocoding. UPRNs were regarded as the gold standard because they are linked to XY coordinates and contain more detailed information on a premises than XY coordinates.

4.2.4.4 Temporal Accuracy

Temporal accuracy was assessed by investigating the extent to which LAs provided temporal data. This included trading times and when the outlet opened (closed) for trading.

4.2.4.5 Thematic Accuracy

Thematic accuracy was assessing by the method of classification that was provided in the dataset.

4.2.5 Sub Analysis

PoI and FSA were compared to LA data for four LAs; Torfaen, Gwent, Newport and Swansea. These four LAs were analysed as they were representative of urban/rural morphologies and were heterogeneous across socio-economic status. This sub analysis was adapted from Burgoine [188].

Outlets that sell unhealthy food were extracted from the PoI and FSA dataset using their own classification systems (Appendix 2.). The outlets were then reclassified according to classification system reported in Table 4.1 [189].

Outlets from each dataset were matched based on their name and address. The completeness of the secondary datasets was analysed with percentage agreement, positive predicted values (PPV) and a sensitivity analysis in R. These statistics have been widely used in the literature [188, 194, 217, 333]. Percentage agreement represents the percentage of unhealthy food outlets present in the secondary dataset and the LA dataset. PPV is a measure of precision that relates to the percentage of outlets found in the secondary data sources that were present in the LA data. The sensitivity analysis describes the percentage of outlets listed in the LA data that were also in the secondary datasets. Data were categorised according to bands used in the literature [188, 194]: 'poor' <30%; 'fair' 31-50%; 'moderate' 51-70%; 'good' 71-90%; 'excellent' >91%. Differences between percentage agreements, PPVs and sensitivities for all food outlets, as compared by type, were assessed using Fisher's Exact tests. This test was preferred over chi-squared tests due to potentially small expected values.

4.3 Results

4.3.1 LA spatial quality

4.3.1.1 Completeness

LAs provided spatial attributes to varying extents (Table 4.3) Newport provided the greatest number of UPRNs; 2 UPRNs were missing from 1470 food outlets. Seven LAs were unable to provide any UPRNs. On average, XY coordinate provision was smaller than UPRN provision. Generally, LAs that provided a high percentage of UPRNs, also provided a high percentage of XY coordinates. Eleven LAs were unable to provide any coordinates. Five LAs were unable to provide any point spatial locators. All LAs provided addresses. On average, 1% of addresses were missing.

4.3.1.2 Logical Consistency

The datasets provided by each LA were consistent within themselves but there were variations between LA. The extent to which point spatial locators were provided paired by LA, from LAs not providing any point spatial locators to Newport providing more than 99% of the UPRNs and XY coordinates. Every LA provided a consistently high percentage of addresses. However, the format in which these addresses were provided was different between LAs. Addresses were either provided all in one free text field or divided in to 5 parts (Figure 4.1).

Address				
High Street, Llansteffan, Carmarthen, Carmarthenshire, SA33 5JZ				

Address 1	Address 2	Address 3	Address 4	Address 5
High Street	Llansteffan	Carmarthen	Carmarthenshire	SA33 5JZ

Figure 4.1 Example of different formats that the addresses were received in

LAs were also inconsistent in the way that outlets were classified. Two outlets provided just FSA classifications for the food outlets. Fourteen outlets provided their own classifications and three outlets provided FSA and LA classifications. Two LAs did not provide any classifications. These classifications are documented in Appendix 2.

4.3.1.3 Positional Accuracy

The local authority data provided address level locations for all outlets across all LAs. Addresses tended to vary in how they were recorded. Addresses were provided to a high standard. A maximum of 4% of addresses were missing from the Swansea dataset. The Unique Property Reference Number (UPRN) is an attribute that would have ideally been provided for every outlet. It is a nationally approved method of identifying buildings but it was provided to varying completeness. Newport provided the highest UPRN coverage (<2% outlets missing UPRNs). Whereas Monmouthshire, Denbighshire and Flintshire were unable to provide any UPRNs.

Only 9 LAs provided coordinates. The level of coordinate provision tended to correlate with UPRN provision; where UPRN provision was higher, so was coordinate provision. Coordinates that were provided located the center of the outlet.

4.3.1.4 Temporal Accuracy

The local authorities provided historical databases. This longitudinal dataset of food outlets is important for this kind of research because it will allow for future analysis of how outlet type has changed over time. For example, in Swansea, in 1996 there were 53 green grocers and in 2014 there were just 13. Many of these greengrocer properties have been replaced with unhealthy food outlets. The datasets record data from 1996-2015.

Generally, LAs provided at least 10 years of longitudinal data. Table 4.2 shows what year the food outlet dataset begins. Wrexham was the only LA to provide food outlet data that was not longitudinal. The Vale of Glamorgan provided five years of data which was the shortest temporal scale to be provided. Ceredigion, Pembrokeshire and Powys provided 25 years of data (1990-2015) (Figure 4.2).

Table 4.2 The year that the food outlet datasets begin by Local Authority

Local Authority	Year
Anglesey	2001
Blaenau Gwent	1992
Bridgend	2009
Caerphilly	1992
Cardiff	1997
Carmarthen	1999
Ceredigion	1990
Denbigshire	1996
Flintshire	2002
Gwynedd	1999
Merthyr	1996
Monmouthshire	2001
Neath Port Talbot	2004
Newport	1998
Pebrokeshire	1990
Powys	1990
Rhondda Cynon Taff	2003
Swansea	1996
Torfaen	1995
Vale of Glamorgan	2010

Figure 4.2 Number of years of longitudinal food outlet data provided by Local Authorities

The local authority data provided rich datasets that contain valuable information about features that make up the food environment. Newport provided the most spatial locators with only 2 UPRNs and 1 coordinate missing from 1470 outlets. Bridgend,

Pembrokeshire, and Cardiff were classified as moderate quality data. For these local authorities, 22%-34% of URPNs were missing from the dataset; as were high numbers of coordinates. The poorer quality datasets received from Merthyr Tydfil and Monmouthshire did not provide any UPRNs or coordinates. 6 local authorities have not been able to provide any suitable data.

4.3.1.5 Thematic Accuracy

Thematic accuracy between the LAs was variable. As aforementioned, outlet categorisation was inconsistent between LAs. LAs were also inconsistent in the way that outlets were classified. Two outlets provided just FSA classifications for the food outlets. Fourteen outlets provided their own classifications and three outlets provided FSA and LA classifications. Two LAs did not provide any classifications. These classifications are documented in Appendix 2. Trading times were provided by two LAs and twenty LAs provided when the premises opened and closed.

4.3.2 Sensitivity Analysis

The FSA data provided more accurate data than the POI data. Missing FSA records ranged from 12-30% for all food outlets; whereas POI missing records ranged from 63-72% (Table 4.4). The number of matching records was also higher in the FSA data. A maximum match rate between FSA and LA data, for all food outlets, was Torfaen with 85%. The maximum match rate for all outlets amongst POI data was 37% in Blaenau Gwent.

In Torfaen and Blaenau Gwent, Bakers were the best recorded in the FSA data, with 100% match rate. In Newport and Swansea, the highest FSA match rate was in café/coffee shops. Non-food outlets were consistently matched at a low rate. Over three quarters of non-food outlets were missing from the Swansea FSA data. Takeaways consistently had less than 20% missing from the FSA data set across all four LAs.

In the POI data, Bakers have the highest rate of matching. In Torfaen and Blaenau Gwent, 100% of POI Bakers match with LA Bakers. Bakers match the highest in Newport too but with a match rate of 62%. Across all four LAs, the Café/Coffee Shops in the POI data consistently match to less than 35% of the reference data. There were no "Non Food Outlets" identified in the POI dataset.

Table 4.3. Summary of provision of attributes by LA. Conwy did not provide any data

Local Authority	Food Outlets (n)	Names Missing (%)	Addresses Missing (%)	UPRNs missing (n)	UPRNs Missing (%)
Anglesey	1131	20	0	101	10
Blaenau Gwent	1044	0	1	95	9
Bridgend	784	0	0	25	3
Caerphilly	3740	0	2	840	22
Cardiff	4970	0	0	1704	34
Carmarthen	2289	1	0	649	28
Ceredigion	2395	0	3	1340	56
Conwy					
Denbighshire	145	0	2	145	100
Flintshire	175	0	1	17	10
Gwynedd	3387	0	0	3381	100
Merthyr Tydfil	584	0	0	584	100
Monmouthshire	624	0	1	624	100
Newport	1470	0	0	2	0
NPT	1101	0	0	48	4
Pembrokeshire	4603	0	2	1133	25
Powys	6729	0	0	5315	79
RCT	1981	0	0	9	0.5
Swansea	4255	0	4	159	4
Torfaen	645	0	0	61	9
Vale of Glamorgan	2362	8	1	2362	100
Wrexham	1575	0	0	1575	100

An ideal PPV would be 100% as this would show that the outlets in the secondary datasets are present in the LA data. This was not the case, however all outlet PPV values are higher in the FSA dataset; with top value of 91% for Newport. However, the POI dataset produced poorer values with a maximum PPV value for all outlets of 44%. Sensitivity was also higher in the FSA data. POI sensitivity was generally low with median sensitivity values between 29% and 44% (Table 4.5).

Table 4.4. Percentage agreement between LA and FSA and PoI datasets by food outlet type

	LA (n)	FSA (n)	PoI (n)	Matching FSA (n)	Missing FSA (%)	Matching PoI (n)	Missing PoI (%)
Torfaen							
All food outlets	152	203	176	129	15%	39	74%
Cafe/Coffee Shop	40	56	20	31	23%	14	65%
Bakery	4	6	5	4	0%	4	0%
Takeaway	91	87	69	77	15%	22	76%
Supermarket	17	35	39	14	18%	5	71%
Convenience	0	5	40	0	0%	0	0%
Entertainment	0	1	0	0	0%	0	0%
Non-Food Outlet	0	8	0	0	0%	0	0%
Specialist	0	5	3	0	0%	0	0%
Gwent							
All food outlets	226	212	189	184	12%	84	63%
Cafe/Coffee Shop	41	37	22	27	24%	9	78%
Bakery	5	5	6	5	0%	5	0%
Takeaway	84	75	80	69	7%	60	29%
Supermarket	55	57	58	51	11%	24	56%
Convenience	18	17	18	14	17%	7	61%
Entertainment	0	0	0	0	NA	0	NA
Non-Food Outlet	18	16	0	12	22%	0	100%
Specialist	5	5	5	5	0%	2	60%
Newport							
All food outlets	384	313	297	285	26%	109	72%
Cafe/Coffee Shop	51	48	30	45	12%	10	80%
Bakery	21	1	17	1	95%	13	38%
Takeaway	149	134	112	128	14%	46	69%
Supermarket	96	76	77	74	23%	27	72%
Convenience	41	41	59	27	34%	10	76%
Entertainment	0	0	0	0	NA	0	NA
Non-Food Outlet	23	13	0	13	43%	0	100%
Specialist	3	0	2	0	100%	1	67%
Swansea							
All food outlets	624	458	580	439	30%	176	72%
Cafe/Coffee Shop	108	94	91	85	21%	26	76%
Bakery	52	27	29	29	44%	19	63%
Takeaway	177	155	220	142	20%	50	72%
Supermarket	160	109	116	98	39%	111	31%
Convenience	92	57	121	51	45%	19	79%
Entertainment	0	0	0	0	NA	0	NA
Non-Food Outlet	13	3	0	3	77%	0	100%
Specialist	22	13	3	13	41%	2	91%

Table 4.5.PPV and sensitivity analysis for FSA and PoI data sets. NA values indicate where dataset did not include any outlets

	FSA PPV (%)	FSA Sensitivity (%)	PoI PPV (%)	PoI Sensitivity (%)
Torfaen				
All food outlets	64	85	22	26
Cafe/Coffee Shop	55	78	70	35
Baker	67	100	80	100
Takeaway	89	85	32	24
Supermarket	40	82	13	29
Convenience	0	0	0	0
Entertainment	0	0	NA	NA
Non-Food Outlet	0	0	NA	NA
Specialist	0	0	0	0
Gwent				
All food outlets	87	81	44	37
Cafe/Coffee Shop	73	66	41	22
Baker	100	100	83	100
Takeaway	92	82	75	71
Supermarket	89	93	41	44
Convenience	82	NA	39	NA
Entertainment	0	0	0	0
Non-Food Outlet	75	NA	NA	NA
Specialist	100	NA	40	NA
Newport				
All food outlets	91	74	37	28
Cafe/Coffee Shop	94	88	33	20
Baker	100	5	76	62
Takeaway	96	86	41	31
Supermarket	97	77	35	28
Convenience	66	NA	17	NA
Entertainment	0	0	0	0
Non-Food Outlet	100	NA	NA	NA
Specialist	NA	NA	50	NA
Swansea				
All food outlets	96	70	30	28
Cafe/Coffee Shop	90	79	29	24
Baker	100	56	66	37
Takeaway	92	80	23	28
Supermarket	90	61	96	69
Convenience	89	55	16	21
Entertainment	0	NA	0	NA
Non-Food Outlet	100	23	NA	0
Specialist	100	59	67	9

4.4 Discussion

4.4.1 Main Findings

LAs in Wales provided variable quality of spatial referenced food outlet data. All local authorities provided address data but a number of LAs were unable to provide UPRNs or XY coordinates. This may be due to variation in resources available (e.g. time and employees), inter department communications, expertise and processes for storing, collecting, and sharing food outlet data that exist within LAs. These differences between LAs were documented and discussed in Chapter 3, and capacities and processes appear to influence the spatial quality of the data that is recorded. LAs that were difficult to obtain data from generally provided poorer quality spatial data. There is no official guidance or any legal requirements for LAs to record UPRNs or XY coordinates in food outlet datasets. Therefore, in order for these spatial locators to be recorded in the data, the LA must have an appreciation of the benefits of recording these attributes. These benefits include; being able to map the data, ease of identifying and tracking outlets, and the reuse of the data for secondary purposes (e.g. academic research). LAs have the potential to provide high resolution spatial locators via the Local Land and Property Gazetteer (LLPG) – this is a definitive, spatially referenced address list for commercial and residential property in a LA. As discussed in Chapter 3, researchers need to promote the use of LA data and work collaboratively with LAs to share best practices and knowledge with each other. Through good professional relationships, LAs can understand the perspective of researchers and the importance of including spatial locators in their data. Furthermore, sharing results and research findings that are generated from LA data will allow LAs to reap the benefit of their data.

LAs universally record addresses of food outlets. Therefore, LA datasets that do not provide UPRNs or XY coordinates need to be geocoded to generate these high resolution spatial locators. This is a time-consuming process but the result is a reliable longitudinal dataset. LA data have been reported as a reliable data source as an alternative to field validated studies [217, 333]. The completeness of the datasets compared in this chapter is the biggest difference between the LAs dataset and the FSA and PoI datasets.

The PoI dataset performed particularly poorly in the sensitivity analysis. These results contradict those previously reported by Burgoine et al [188]. Burgoine et al compared the PoI dataset with LA data from Cambridgeshire LA. It was reported that PoI had an overall PPV of 74.9% and sensitivity of 59.9% when compared with data from Cambridgeshire LA [188]. The study found PoI data to be a reasonable alternative to LA data. This chapter reports an average PPV of 33.3% for the four analysed LAs and an average sensitivity of 29.8%. The difference in these findings could be because different geographic regions have been evaluated. There could be a spatial bias in the PoI dataset which results in the PoI dataset being less complete and accurate for Wales than in Cambridgeshire. However, PoI data is collated from over 170 sources and is updated quarterly but little is actually known about how much data cleaning is undertaken. Duplicate entries and outlets remaining in the dataset were a large source of the discrepancies between the LA dataset and the PoI dataset. The results here suggest that PoI data for Wales is not suitable for use in research studies. Not only are many outlets not present in the dataset but there are duplicates and outlets that are closed. As there is no record of when the outlet ceased trading, it is not possible to identify these outlets in the dataset. Furthermore, there are no opening times recorded in the PoI dataset.

The FSA dataset reported excellent PPV values and a high sensitivity but not perfect agreement. This was an unexpected result as LAs provide FSA with their data. The discrepancies could be explained by temporal differences between the datasets. The FSA data contained trading outlets but FSA being up to data is reliant on LA being timely in updating the FHRS. On the other hand, FSA data is updated daily and the LA data was obtained a few months prior to downloading the FSA data. A further cause of the difference between the outlets included in the LA and FSA datasets is because the FSA does not record data on outlets that do not prepare food on the premises. This is reflected in the lower PPV and Sensitivity scores of outlet types; convenience, non-food and specialist. Although the FSA dataset showed good agreement with the LA data, the dataset is not appropriate for this study as it is not longitudinal and does not contain UPRNs or address level XY coordinates.

Although LAs are consistently found to be the most comprehensive record of the food environment, it should be noted that mobile food outlets are not included in these records. This is because mobile food outlets are registered to the home address of the

owner and data is not usually recorded regarding where the outlets trade, or the roads that they traverse when trading. Mobile food outlets are a source of unhealthy food for children and are therefore part of the 'retail food environment'. Due to the lack of data available, it is not possible to include this data in this study.

4.4.2 Strengths

The spatial quality of LA food outlet data has not been evaluated in this way before in the published literature. Using an accepted method of evaluating spatial data quality means that this evaluation can be replicated elsewhere.

This study has used a uniform classification [217] in order that the results are comparable with other studies. To give a fair comparison between the datasets, the LA data were collected as closely together as possible. However, there will likely still be temporal discrepancies between the datasets.

This chapter highlights the advantages and disadvantages of LA data and secondary data sources. No dataset is a perfect representation of the real world but it is essential that researchers consider the data source they use in exposure modelling. This chapter contributes to the discussion on equipping researchers with information about spatial datasets so that they can make informed decisions.

4.4.3 Limitations

4.4.3.1 LA Quality Analysis

A limitation of the quality analysis of the LA data is that it assumes no human error. Data at the LA will be input by employees and quality checks are not compulsory. LA data has previously been reported as very good but not perfect [333]. Errors in LAs data usually relate to the temporal accuracy of the data. LA datasets have been found to contain a small proportion false positives and also not recording outlets [217]. It is not surprising that LAs do not necessarily undertake quality checks as public services are under resourced with stretched budgets. On the other hand, this data is recorded for legal purposes (trading businesses must be registered with the LA and LAs must undertake food standards inspections) and the LAs also provide their data to the FSA so there may be a degree of circularity in the data.

This investigation analysed a large geographic area and so field validation was not conducted. Further research should look to sample the datasets and evaluate whether there are any systematic gaps in LA data (such as across socio-economic divides [188]).

4.4.3.2 Sensitivity Analysis

A limitation was the matching method that was used to compare the FSA and PoI with the LA data. It was extremely challenging to compare the datasets because of the different ways that data was recorded within them. Probabilistic or deterministic matching should be explored in future studies.

In addition to the matching method, the comparison was further complicated by the data sources classifying outlets in different ways. Not only were the classification systems different but there is the potential that outlets may be classified as different outlet types. For example, small supermarkets being classified as supermarkets or convenience stores. This makes it difficult to be confident that like for like are always being compared.

4.4.4 Implications

The LA data that was used to calculate exposure to the RFE in later chapters was from Swansea, Carmarthenshire, Cardiff and Rhondda Cynon Taff. These are the regions from where the health data has been obtained. However, at the time of this analysis, Carmarthenshire, Cardiff and RCT data had not yet been obtained. The four LAs that were analysed in the sub analysis were chosen because they are representative of socio economic status and urban/rural morphologies in Wales.

The main implication of the results presented in this chapter on this study is that LA data will be used to calculate exposures to the RFE for teenagers in south Wales. LA data is not without flaws but the results suggest that the LA dataset is of the most reliable quality. This resonates with previous studies where LA data have been reported as a good source of food outlet data and have been used as the 'gold standard' in comparative studies. In addition to the LA data it is also the most appropriate to use

for this thesis project as the dataset is longitudinal, address level and can be relied upon to be complete.

The LA data will have to be geocoded to operationalise it for use in spatial studies, as many of the outlets do not contain XY coordinates or UPRN. However, this is possible as addresses have been supplied for 99% of outlets. The decision to spend time and resources collecting and processing LA data is a decision to be made by researchers. Researchers should weigh up the advantages and disadvantages of data sources and evaluate the appropriateness of the data for the desired purpose. Researchers should be upfront with the sources of data that they use and the limitations associated with the datasets. Policy makers and practitioners should be aware that data sources and methods have limitations and care should be taken when interpreting results from scientific papers.

With regard to research, these results suggest serious caution should be taken if using the PoI dataset for studies in Wales. Further research needs to be undertaken to investigate whether using different data sources influences the results of exposure modelling. The widely known concept of "garbage-in, garbage-out" suggests that the integrity of the dataset will influence exposure results [333]. It is being used because it is easy to access and a paper has reported it as an acceptable alternative to LA data. However, this study reports similar findings to the paper. There is a need to undertake further research to evaluate whether this quality has an impact in exposure studies. One study found that it didn't. However, the study defined environmental exposure at a postcode and LSOA level. More research needs to be undertaken to see the influence on address level studies. Of course, the scale of studies will be important to consider and how exposure environments are defined will also contribute to the degree to which the food outlet dataset would influence exposures. A study published in 2017 [10] reported that there was a small difference in the size of association between exposure to the RFE and BMI for LA data and PoI data. Exposure to the RFE are measured for the three differently defined home neighbourhoods. The three definitions that are used to define the home neighbourhood are: LSOA density; 800m Euclidean buffer; and 2000 m Euclidean buffer around the home postcode location. However, using these methods to define the home environment does not necessarily capture how individuals interact with their home environment [181, 348, 349]. Furthermore, using datasets that have different spatial resolutions of outlet location may introduce Modifiable Unit

Area Problem (MAUP) [350]. Therefore, further research needs to be undertaken to assess whether using PoI influences results of high resolution exposure modelling.

4.5 Conclusions

LA data is currently the most reliable source of food outlet data. This study has collected a longitudinal food outlet dataset for the whole of Wales for 2010-2015.This has not been done before. Unless field validation takes place, LAs can be expected to contain outlets that are currently not trading. However, LA data can be relied up as the most accurate source of food outlet data as it is a legal requirement for traders to register and unregister their food businesses at least 28 days before an outlet opens (or closes). This means that LA datasets are longitudinal and this is a great advantage over FSA and PoI datasets. In order for LA data to be utilised in spatial research, the data needs to be processed. Some LAs provided data that will need to be extensively processed before they are ready to be used in research. Other LAs provided data with a very high provision of UPRNs or XY coordinates. These LAs tended to be larger LAs that are more engaged with research institutions. PoI data had poor PPV and moderate sensitivity to the LA data. This poor quality is likely to be a result of how the data is collated and maintained. FSA data had good PPV and sensitivity with LA data but is not longitudinal. LA food outlet data will be used as the food outlet data source for modelling exposure to the 'retail food environment' for children aged 13-14 in south Wales.

Chapter 5

Exploratory Data Collection of Routes to and from School

5.1 Introduction

This chapter reports the collection of data on walking routes to and from school from two cohorts of teenagers in south Wales. The first stage of data collection recorded data from 11 teenagers aged 13-16 from across south Wales. The primary aim of this pilot data collection was to investigate whether a supervised mapping survey would be an appropriate method to collect this data. The second round of data collection recorded GPS data of walking routes to and from school for children aged 13-14 that attended a secondary school in Neath Port Talbot in south Wales. The aim of this data collection was to collect GPS data on walking routes to and from school for around 100 pupils in south Wales, ages 13-14. The two aims of this chapter were:

1. To establish the most appropriate method of collecting route data from teenagers;
2. To collect a robust dataset of walking routes to and from school for teenagers

An objective of this PhD was to better characterise the daily activity space of children by improving GIS modelled exposures to the 'retail food environment' (RFE) along walking routes. This is so that reliable predictions of children's exposure to the RFE along their route to and from school can be made. Current practice of calculating environmental exposures uses GIS shortest network routes. This has previously been reported as an unreliable method of predicting the environment that a child is exposed to along their route to and from school [11]. By improving the accuracy of GIS modelled exposures, population level exposures can be generated. This will contribute to research of complex systems in obesity and evaluating the effect of the food environment on health at a population level.

In order to improve the accuracy of GIS modelled exposures, route choice should be incorporated in to GIS measures. Currently, shortest network routes are used as a proxy for the distance travelled and exposure along walking routes. However, route choice has been found to be more complicated than a simple shortest route estimation for routes to and from school [351]. To understand more about route choice of children who walk to school, data was collected on the routes that children take to school. This chapter documents a pilot study and a GPS data collection methodology that I developed to collect data and validate the GIS generated exposure model.

5.2 Pilot Study

Studies have collected data on routes travelled using many different methods; travel diaries, interviews, mapping surveys and GPS data [114, 266, 352, 353]. These methods have been discussed at length in the literature review (Chapter 2). The most commonly used method of collecting route data are mapping surveys or GPS data. The literature recommends that GPS data is the most reliable data source from which to generate environmental exposures. However, collecting GPS data can be time consuming and costly. Procuring the GPS loggers, ensuring that participants charge the devices and obtaining the loggers from participants once the study finishes are all challenges of collecting GPS data. Furthermore, several studies suggest that when participants wear GPS loggers, their behaviour is influenced by an awareness of the device [279, 280]. Population level analysis is also not feasible using GPS data.

A study published by Stewart et al. in 2017 [354], suggested that mapping surveys are a reasonable method for collecting route data from adolescents. The study advised that researchers should consider the advantages and disadvantages of methodologies used to collect route data when deciding which method to implement. Mapping surveys are a cost and resource-efficient alternative to GPS data collection and data can often be collected quickly and efficiently [354]. Previous studies that have collected data using mapping surveys have used paper maps [348, 355, 356] and interactive online mapping applications [354, 357, 358].

5.2.1 Methodology

A pilot study was designed to collect data using a supervised mapping survey to assess if this method would provide sufficient information to inform and evaluate the Exposure Model. I worked with ALPHA [359], a research advisory group (Advice Leading to Public Health Advancement) to collect the data for the pilot study. ALPHA was set up by a research centre (DECIPHer) to understand the views of people aged 14-21 and advise investigators of young people's views to improve research relevancy for young people [359]. Of the 40 members of ALPHA, approximately half of the members were eligible for the pilot study as they were aged between 11-14 years and walked to school. Eleven members of ALPHA volunteered to participate in the pilot study. I met with ALPHA in July 2015 to record the routes that the younger group members took on their route to school. Participants undertook the mapping survey and a closed question questionnaire.

5.2.1.1 Participant Recruitment

Contact was made with the leader of ALPHA by email. This person relayed information about this project to the members of ALPHA at a meeting. Eleven members of ALPHA aged between 13 and 16 years old chose to participate in the data collection. The organisers of ALPHA organised parental consent.

The participants lived, and attended school in south Wales; Cardiff, Swansea, Caerphilly and Merthyr Tydfil. Their routes to school were heterogeneous in urban/rural regions and socio-economic status.

5.2.1.2 Data Collection

Route Collection

Home and school postcodes were obtained from the participants prior to the meeting. Using these locations, I calculated the postcode level shortest network routes from home to school using Network Analyst in ArcMap 10.1 using Ordnance Survey Integrated Transport Network Roads [360]. Home postcode centroids were obtained from Ordnance Survey's Code Point [361] and address level school locations were used from Ordnance Survey's AddressBase Premium [362].

The participants were given paper maps that included their home and school locations [348]. The shortest network routes were included on the map in order to guide the pupils and act as an aid. The shortest network routes were represented as a red line which was overlaid on to a Google Maps background. Google Maps was chosen as the base map design because it is a design that many young people would be familiar with. Young people are technologically literate and are familiar with wayfinding applications such as Google Maps [363]. Google maps is the most downloaded wayfinding app for smart phones [364, 365] familiarity with the map would further help the participants to accurately recount their route to school [366].

The participants assessed the route calculated by the GIS and corrected errors where appropriate with a black marker. To ensure that the participants recounted their route as accurately as possible, Google Street View was used so that the participants could virtually 'walk' their route as children have been reported as making fewer mistakes recounting a journey when navigating a digital map [367]. Participants were informed that if they took a route that was not available on the map, that they should mark that despite the omission.

The verified routes were then digitised using GIS. This dataset has been used in Chapter 5 to assess the most appropriate dataset for use in the GIS exposure model. The drawn routes were used instead of GPS data as when conducting the analysis, this dataset was the only dataset available.

Questionnaire

A closed question questionnaire was given to the participants to complete. This questionnaire was designed to ask participants about their method of transport to school and what food/drink they purchased on their way to/from school.

Food Outlet Data

Food outlet locations were obtained from the LA dataset that was documented in Chapter 4. This data is address level data and unhealthy food outlets were extracted from the dataset according to the typology of unhealthy food outlets documented in Chapter 4 adapted from Lake et al [189].

5.2.2 Analysis

The drawn routes were digitised in ArcMap 10.0 and the route lengths and exposures were calculated. Exposure to the RFE was defined as the number of unhealthy food outlets within 100m buffer of the route that was taken on their route to school. All route analysis was undertaken in Arc Map 10.0. Route lengths have been rounded to the nearest 50m in order that the route lengths will not be disclosive.

Summary statistics for the results from the questionnaires were calculated in Excel.

5.2.3 Results

Eleven hand drawn routes to school were collected from the ALPHA group. The average length of the walk to school was 1820m. The shortest distance travelled to school was 700m and the longest distance walked to school was 4900m. The majority of routes were less than 2100m long. The average exposure was 2 unhealthy food outlets. Seven routes had an exposure of 0 to the RFE. The greatest exposure was 10 along the shortest route of 700m.

Eleven participants completed the food survey. Three participants reported buying food on their journey to school and three different participants reported buying food on their journey home. One participant reported buying food on both journeys, on the way to school and on the way home. More than half of the participants reported wanting to have more choices available to buy healthy food on their journey to and from school. The most popular foods to buy were snacks and fizzy drinks.

5.2.4 Discussion

5.2.4.1 Main findings

The routes recorded from the survey were all shorter than 5km in length. This could be because children who live more than 4.8km from their closest secondary school are eligible for free transport to school. Furthermore, previous studies suggest that on average, children aged 14 years that walk to school, walk 3km to school [368].

More than 7 of the participants had an exposure of 0, but 4 participants reported buying no food on their route to and/or from school. This suggests that either the routes have been reported erroneously, or the children who buy food on their route home traverse

a different route home, compared with their route to school. The latter is more likely as the participants were provided with significant assistance to ensure that the routes were reported correctly. Variation in the route to and from school has been reported in the literature and this is an important dimension to understand in children's activity spaces. In order to make meaningful conclusions about the associations between exposure environments and health outcomes, researchers must accurately represent the way that children traverse through the built environment when they are autonomous. I hypothesise that including the nuances of child behaviour along walking routes will produce more accurate exposures to the RFE along children's walking routes to and from school.

The survey results showed that children purchased convenience snacks and sugary drinks when they bought food on their route to and from school. The questionnaire provides insight in to children's purchasing habits along their route to school. However, for future research an interview approach would provide greater insight and more in-depth themes could be discussed. For example, understanding whether teenagers change their route to specifically go to the shops. Larsen el al [348] suggests that children may walk less direct or longer routes to walk with friends or avoiding major roads, for example.

Supervised mapping surveys are a valuable way to collect data. Previous studies have suggested mapping surveys are a viable alternative to GPS data collection depending on the context of the study. In the context of this pilot study, the method was appropriate and produced reliable results. Making use of ALPHA was an important factor in the pilot study being a success. Recruitment of participants can be time consuming but also drop-out rates can be high and the amount of useable data can be low. Working with ALPHA meant that I had readily available participants who were supervised to ensure all of the participants provided useable data.

However, undertaking a supervised mapping survey for a large cohort, would be an inefficient way to collect data. The process is time consuming and requires the supervisor to have expertise. A supervised mapping survey is a reliable way to record data for small cohorts, however, GPS data collection or web-based surveys are reliable alternatives for larger cohorts.

5.2.4.2 Strengths

A strength of this pilot is the reliability of the data provided. The participants were given a lot of support to complete their hand drawn routes. The maps that the participants drew their route to school on were large, contained points of interest and were overlain on a familiar map background which helped participant understanding. These methods have been positively evaluated in the mapping survey literature [348, 355, 356]. It could be said that using Google maps as the base map could have produced errors as Google maps has been reported as being slow to update in the past. However, more recent research reports that Google Maps quality is comparable with other map sources such as OSM [242, 369].

Complementing the mapping survey with a series of questions about what food purchased on the route to school acted as a memory prompt for the participants to improve route precision. It also gave insight in understanding child travel behaviours on their route to school.

5.2.4.3 Limitations

A limitation of this pilot study was that the data collection process assumed comprehensive spatial literacy from the participants to be able to accurately recount the route that they travel. That being an assumption that the participants would understand how to read a map and would be able to recount their route in the context of a map [370, 371]. The data collected was for the route to school and no data was collected regarding the route home from school. Furthermore, the data was just one single route and it has been shown in the literature that routes can vary from day to day [11]. Therefore, it is preferable that data is collected for multiple days.

The maps that were provided to the participants already contained a route printed on the map. This route was the shortest network route generated from the ITN network. Despite having informed participants that this route was for reference only, this may have biased and influenced the results of the drawn routes.

5.2.4.4 Implications

The results of this pilot study have shown that supervised mapping surveys are a reliable method to obtain route data from small cohorts of participants. However, to collect a data from a cohort of over 100 participants through a supervised mapping survey would not have been a viable option for this study. Collecting data in this way would have been too time consuming. The collection of data for the 11 participants took several hours and to get reliable results, required a trained researcher to be present. There are more efficient ways to collect data for a larger sample. The time limitations of this pilot study strongly suggest that GPS data would be more appropriate to collect a larger sample of data that was required for this project. The results and lessons learned from this pilot study informed my GPS data collection strategy reported in the next section.

The implication of these results on this PhD project was that an alternative method of data collection had to be undertaken for the larger cohort. This next section documents the how GPS data was collected for walking routes to and from school.

5.3 GPS Data Collection

The second round of data collection recorded GPS data from teenagers using a mobile phone app. The data collection focussed on year 9 pupils (aged 13 and 14 years) who usually walked to school and owned a smart phone. A GPS-enabled app was used to collect data on the route that a child walks to and from school. A travel diary was designed to capture information on whether participants bought food along their route to and/or from school.

Mobile phones are being increasingly used in healthcare research to record objective measures of physical activity and exposure. A mobile phone app was chosen to record GPS data of the participant's routes to and from school for this study because:

1. Data would not be recorded all day and therefore reduce the volume of data that would have to be processed and extracted

2. It would minimise behaviour change as the participants carry their phone around with them on a daily basis [372]

3. It is a low cost alternative to GPS and accelerometer instruments [373]

4. Students could participate without being conspicuous [372]

A travel diary was also developed from the closed questionnaire (section 5.2.1.2) to collect data on children's purchasing habits on their route to and from school.

5.3.1 Methodology

5.3.1.1 Data sources

GPS data

For android phones, the app chosen was called 'Track My Trip', for Apple phones the 'Easy Trails' app was selected and for Windows phones "Map my Tracks". They had a simple interface, were free to download, and did not require membership to export data. All apps used GPS to record the path, speed and distance of the child's journey when the app was activated. In addition to this, the participants were asked to complete a travel diary.

Travel diary

The travel diary was developed based on an ethically approved collaborative "Travel to Work" study that was undertaken by the University of Bristol, University of Bath and Swansea University [374] (Appendix 4).

Food Outlets

Food outlet locations were obtained from the LA dataset that was documented in Chapter 4. This data is address level data and unhealthy food outlets were extracted from the dataset according to the typology of unhealthy food outlets documented in Chapter 4 adapted from Lake et al [189].

5.3.1.2 School and Participant Recruitment

Consideration was given to the busy environment that schools are and the contact with the students was designed to be kept to a minimum. All contact with participants was designed to be undertaken during three sessions during a lunchtime or registration time. These times in the school day were chosen so that no teaching was disrupted.

In conjunction, the data collection was designed to give the school and pupils the most out of a visitor undertaking scientific research in the school. I offered teachers the opportunity to use this engagement to explore GIS (Geographical Information Systems) with their pupils in workshop-style lesson provided by myself. I made clear that I was enthusiastic to help the school, as GIS is on the national geography curriculum for children of this age. I also designed the sessions to be fun, interactive and age-appropriate. The participants were also given goodie bags that contained healthy treats and vouchers for a local water park.

The first visit introduced students to the study and parental consent forms were given out (Appendix 4). On the second visit to the school the consent forms were collected and the participants were assisted in installing the app on their phone. A final visit was undertaken to collect the GPS data and travel diaries from the students.

To encourage the schools to engage and participate, several different incentives were included in the design of this data collection protocol at school level and pupil level. The intention was to maximise the likelihood of schools engaging with my project. To the Head Teacher and Head of Geography (HoG), I outlined how the study would give their students an introduction to GIS and the benefits of learning these skills (Appendix 4). I highlighted how participating in this study fitted in with the curriculum for KS3 and KS4. I described the wider perspective by including information on how the results of this study would be important in providing evidence to local councils and the government to make changes to provide a healthy local environment. As a final school level incentive, I provided a report of the pupil's results that would be beneficial to their involvement with the Healthy Schools scheme that is promoted by Welsh Government. In appreciation of participation, the pupils were given vouchers to a water park in Swansea and a healthy snack goodie bag. The healthy snack goodie bag was provided by healthy snack companies that I had contacted for sponsorship. All children who completed the data collection were also entered into a draw for a chance to win vouchers to a trampoline park in Swansea. These incentives were developed based on recommendations from the PEACH and SPEEDY studies that have collected GPS data on children during the school day [47, 273].

Following the design of the data collection strategy, ethical approval was obtained from Swansea University Medical School Ethics Committee. Several comments had

to be addressed regarding how the data would be safeguarded and protected before approval.

After ethical approval was obtained, four secondary schools from Swansea were approached to enrol with the study. These four schools were the same schools that had previously taken part in a Swansea University Medical School study that collected child Body Mass Index data (BMI) for year 9 pupils (aged 13 and 14). The same schools and age group were approached for the GPS study in order to maintain spatial and social homogeneity between the spatial and health data that would inform the exposure predictions of this study and the BMI data from a previous cohort that would be used for analysis in Chapter 8. Pupils that were recruited walked to school.

Head teachers at the selected schools were sent a short invitation letter (Appendix 4). An information pack was also delivered to the school which contained a copy of the ethical approval and information leaflets for teachers and students. In this pack, details were provided about the aims of the study, data collection procedures, the degree of school involvement required, and benefits of participation. A meeting with the head of geography was requested at each visit. The schools were then contacted by phone and contact details for the head of Geography were obtained.

A fifth school was also approached in Port Talbot. The deputy head of this school was a family member and was an enthusiastic advocate of the research. An in depth explanation was given to the deputy head teacher so that they were able to sell the project to the pupils. An announcement was made by the deputy head teacher in a year 9 assembly and students volunteered to participate in the study. The deputy head teacher brought the participants together and explained the premise of the research and discussed any questions that the children had. The consent forms were then handed out and returned by the time I first visited the school. A sixth school was approached where I attended secondary school. However, after a period of 1 month, whilst a response from this school was being waited for, I decided that no more time could be given to GPS data collection on this project. Instead, I decided to use GPS data from a large-scale study that had already been collected by trained researchers.

5.3.1.3 Data collection

In June 2016, I visited the participating school to refresh the participants about the study and to help them download the app on to their phone. Step by step, visual, instructions were produced on how to download the app. A set of instructions was generated for three different operating systems: Windows; Apple and Android. These instructions were given out to the pupils and two IT teachers assisted during the session. The session was for half an hour during registration period. The participants gave their pupil ids and phone numbers so that a text could be sent out to remind the pupils to record their data in the morning. A second text was sent just before the end of the school day.

The raw GPS data from participants was transferred from participant mobile phone devices on to an encrypted file by connecting the phone to a password protected laptop with a lead. CSV or GPX formats were downloaded depending on which operating system was used to collect data. The data points were imported into PostGIS [375] database and transformed to Mercator projection (EPSG 4326). The gpx files were imported to PostGIS through the command line tool ogr2ogr. CSV files were imported to PostGIS using the command line tool pgfutter.

5.3.1.4 Route Distances and Exposure to the Retail Food Environment

Any spurious points were manually cleaned from the GPS point dataset and line geometries were created from the point data. The length of each route was then calculated.

Exposures were calculated for each route. Using Local Authority data as the food outlet data source (Chapters 3 & 4). Exposure to the 'retail food environment' (RFE) was defined as the number of unhealthy food outlets within 100m of the route that a child walked on their route.

Route distances and exposure to the RFE were calculated in PostGIS [375] and pgRouting [376]

5.3.2 Results

No schools that were approached in Swansea participated in the study. Two of the schools did not respond to any of the multiple phone calls or emails. The remaining two schools engaged in planning a visit to begin data collection but later pulled out due to high workloads. The one school that did participate in the study recruited 35 children. This cohort contained pupils across a range of academic abilities and socio-economic backgrounds.

5.3.2.1 GPS data

Seventeen children provided data from the thirty-five participants that were recruited for the study. Those who participated provided varying amounts of GPS data on their routes to and from school. The other participants did not provide data as they reported that they forgot to record their route to and from school using the GPS app, despite being sent a reminder text in the morning and afternoon.

The number of participants that provided GPS data was 10, 4 of whom provided data for their walk to school and their walk home. Four children also provided data for multiple days, three of these participants were the students that provided data on their walk to school and their walk home.

To account for some participants having multiple readings, and others not, an average value was taken for each participant and these were compared. Morning route distances were, on average, 889m shorter than the average afternoon route distance (Table 5.1) The average morning exposure for the participants was 4 outlets, the average afternoon exposure was 10.

A comparison was made between the four participants that provided data for their walks to school and home. The difference in the distance they travelled in the morning compared to afternoon was small (between 13 and 58m). However, exposure to the RFE doubled on their walk home.

Table 5.1. Summary of GPS route distances and associated exposure to the RFE by participant

Route to School			**Route Home**		
id	Distance (m)	Exposure (n)	id	Distance (m)	Exposure (n)
2	4558	12	1	2587	12
3	306	1	2	4586	24
4	1106	4	3	248	2
5	1402	4	5	1363	4
9	108	1	6	4549	26
10	1450	4	7	1905	7
			8	3683.5	6
			9	95	2
Mean	**1488**	**4**		**2377**	**10**

Figure 5.1 Word cloud to highlight the most popular food types bought by teenagers. Larger words show more popular foods

The survey data yielded a similar response rate. Of the eleven travel diaries completed, two travel diaries were completed by the participants who provided GPS data. The remaining nine participants completed and returned their travel diaries but provided no GPS data. Out of the 44 days of data collected across all participants, on nineteen

of these, food was bought on the way to school and on seven days, food was bought on the way home. Three children bought no food throughout the week. Children who bought food on their way home had recorded longer times to walk home. Figure 5.1 highlights the foods that the participants recorded buying on their daily commute. No hot food was purchased.

5.3.3 Discussion

5.3.3.1 Main Findings

Participants were exposed to more opportunities to buy unhealthy food on their way home compared with the walk to school. This is because a greater number of food outlets were open for trading on commute home compared with the walk to school. Many independent, hot food outlets are not open early in the morning when children are travelling to school. However, no hot food was recorded in the travel diaries. The most popular foods to purchase on the journey to or from school being sweets, chocolate and water. These food types are more likely to be purchased from convenience stores and supermarkets.

The results show that the longer the route, the greater the exposure. However, the travel surveys suggest that this greater exposure does not necessarily lead to children buying more food. The results from the survey show that children prefer to buy snack food rather hot food takeaways on their walk to and from school. This could be for a number of reasons. Hot food takeaways tend not to be open during the commute to school and the majority of hot food takeaways do not open until 5pm, which is after the majority of children have walked home. It may also be that children do not buy hot food on their way home if they know that they are going to have a meal when they get home.

Recruiting schools to participate in the study proved to be very challenging. Designing the data collection strategy, organising incentives for the participants and school and obtaining ethical approval was an extensive process which yielded a low response rate. One out of five schools approached engaged with the study, and 48% of participants who signed up to be involved with the study provided data. Schools are already very busy places and even though the data collection strategy was designed to be as attractive to schools as possible, the schools were simply too busy to engage with the study. On reflection, if I were to repeat the study, I would look to approach more

schools and recruit more participants from each school. In addition, I would plan more time for data collection and make more visits to the school in order to make face to face contact with the relevant staff. Another change to the data collection method would be to approach schools from areas that are further away from Swansea University. The schools that I approached are regularly invited to be involved in research as they are in close proximity to the university. The final school that participated in the study was located in a neighbouring county 10-12 miles away from the university. The school found this an exciting opportunity and appreciated how their pupils would benefit from being involved in the study.

5.3.3.2 Strengths

A main strength of this study was that ready-made mobile phone apps were used to collect GPS data. This was a novel concept and to my knowledge has not been used in this field of research before. Owning a phone is now no longer an indicator of socio-economic status [377] and the students that attended the first introductory session covered a range of academic abilities and socio-economic backgrounds. The students were happy to engage in the data collection strategy and informal conversations with the participants suggested that wearing GPS loggers would have put some of them off participating.

Although the sample size was small (20% of planned sample), it has been valuable to learn about the challenges of working with schools in research.

5.3.3.3 Limitations

A limitation of this study is the small sample size. The target sample size of 100 was not reached because of the relatively small pool of children that was available to me from just one school. Future recruitment should anticipate a low uptake and look to approach a greater number of schools and explore working with schools that are located further from the university. However, the percentage of children that provided data is consistent with other studies [378].

The small sample size means that the results are not robust enough to be reliably scaled up. The large difference in average distance of travelled in the morning compared to

the afternoon distances is likely to be skewed. This is because two pupils who travelled an average 3.5km and 4.5km in the afternoon did not provide data for their morning walk to school. On discussion with the participants, the intermittent provision of GPS data was explained as either due to forgetting to record the data (despite having a reminder text each morning and afternoon) or because they did not walk on that occasion. Although it was specified that participants should walk to school, there were exceptions.

5.3.3.4 Implications

The GPS data collection strategy that I designed and implemented here provided me with some data. However, in order to have a cohort of GPS data sufficiently large to design my exposure model, other avenues had to be explored. Obtaining a larger GPS dataset is documented in the next section of this chapter.

The results of this study also give an insight into the food children buy on their route to and from school. The results suggest that hot food takeaways may not be where school children purchase food from on their route to and from school. Policies that limit hot food takeaways around schools are being explored in the UK as a possible public health intervention to reduce obesity. However, these preliminary results suggest that children do not necessarily buy food from hot food takeaways. The foods that were most prevalent were foods that can be bought in convenience shops and small supermarkets. Further research should be undertaken to further understand child purchasing habits. This would not only help to give insight on whether exposure to the RFE influences purchasing habits but also whether food outlet location influences route choice on the walk to and from school. Policies that limit hot food takeaways around schools are being explored in the UK as a possible public health intervention to reduce obesity. However, these preliminary results suggest that children do not necessarily buy food from hot food takeaways along their route to and from school. In addition, the "Healthy schools" scheme in Wales [379] means that school children cannot leave the school premises to purchase lunch off site. The foods that were most prevalent were foods that can be bought in convenience shops and small supermarkets. The results reported here suggest that policies and interventions may need to turn their focus to these outlet-types in specifically targeting childhood obesity and healthy

attitudes to food. Interventions and policies akin to the "Healthier checkout lanes" that have been implemented in supermarkets [380–382] could be explored in smaller stores where children frequent on their way to and from school.

The results of this chapter suggest that using smart phone apps is a feasible alternative to GPS loggers for collecting GPS data for academic studies, particularly when the participants are involved. Further qualitative research could be undertaken to investigate whether children's participation in scientific research could be increased by using mobile phones. Collecting data using their mobile phone allowed participants to be inconspicuous and this may have encouraged a broader demographic of participants to become involved in the study.

Working with the participants suggested that future research could involve developing a purpose-made app for teenagers recording their routes with GPS. There was no app common to different operating systems that was free to download and easy to export data. Three different apps were used and it was therefore more challenging to help the participants when they encountered problems because of the different apps being used. Detailed instructions were created for three operating systems, but in the time pressured environment of the first session with the participants, it was challenging to help every participant who encountered an issue. The purpose of many GPS enabled apps is fitness focussed. Therefore, although the apps were simple to use, there were many functions available that were not relevant to this study. This was distracting and confusing for some participants. A final point is that differences in mobile phone hardware would be the main source of variation in GPS accuracy between individuals.

5.4 Conclusions

The supervised mapping survey produced reliable results and proved to be a trustworthy method of collecting data for a small cohort of individuals. This method is too resource intensive to collect route data for large cohorts, however, a web based mapping survey or GPS data collection provides a way to collect GPS data on a larger scale. Collecting GPS data using a mobile phone app provided detailed data on the

routes to and from school without collecting unwanted GPS data for the rest of the participants' day. This minimised processing and extraction of data. It proved very difficult to recruit schools as they are such busy environments and the particular schools that were approached can become inundated with requests from the university to become involved in research. Future recruitment should explore working with schools that are located further from the university. More pupils should also be recruited in order to offset dropout rates. A dataset that was large enough to robustly analyse what environmental characteristics are associated with route choice in teenagers walking to and from school was not collected through the reported methods. As a result, a cohort large enough to undertake this analysis was obtained from researchers that had collected GPS data for a large scale project (reported in Chapter 7).

The food questionnaire and travel diaries suggest that children do not necessarily buy food from hot food takeaways. The foods that were most prevalent were foods that can be bought in convenience shops and small supermarkets. Further research needs to be undertaken to evaluate child purchasing habits so that policies and interventions targeting childhood obesity can be informed by child activity spaces and purchasing habits.

Chapter 6

Evaluating OpenStreetMap Data Accuracy and Completeness for Environmental Exposure Modelling

6.1 Introduction

Modelling exposures to the 'retail food environment' (RFE) requires; food outlet data, home and school locations and a road network. Obtaining food outlet data has been documented in Chapter 3 and the quality of this data evaluated in Chapter 4. This chapter evaluates whether Open Street Map (OSM) is a reliable alternative to an authoritative dataset to use as the road and footpath data source in modelling exposure to the RFE. This chapter also evaluates the influence of including footpath data in the road network on modelling exposure to the RFE along walking routes to school. The primary aims of this chapter are to:

- Evaluate the accuracy of OSM in Wales to determine whether it is a reliable alternative data source to an authoritative dataset
- Assess the influence of including footpath data in the road network when modelling exposure to the RFE

To evaluate the quality of OSM road and footpath data in Wales, I have calculated positional accuracy and completeness of OSM. The OSM dataset has been compared with the Integrated Transport Network road (ITN) and Integrated Transport Network urban paths (ITNUP) dataset. Completeness and positional accuracy metrics have already been used to successfully evaluate OSM road data elsewhere in the UK and

other locations around the world [204, 205, 383, 384]. Previous studies have found main roads (Motorways, A roads and B roads) recorded in OSM to have a high positional accuracy and completeness in the UK [208, 385]. Minor roads and footpaths included in the OSM dataset have not been evaluated before. The results have been reported for the whole of Wales and stratified by rural and urban regions. The quality of OSM in Wales has not been documented before and to my knowledge, studies also do not stratify results by urban and rural regions. This chapter has further evaluated OSM by comparing GIS generated shortest network route lengths and exposures with walking routes to school collected from a cohort of 11 teenagers aged 13-16.

Studies modelling access or exposure to an aspect of the environment use road data to represent how people travel around. Previous UK studies that have modelled access or exposure to the environment have used the ITN dataset as the road data source. For example, Harrison et al. calculated GIS modelled shortest network routes and associated exposures for children travelling to and from school [11]. The study compared the GIS routes and exposures with GPS data and concluded that GIS modelled exposures are not sufficiently accurate to represent the environment a child is exposed to on their routes to and from school [11]. However, no footpath data has been included in the networks used in these studies. The literature states that children prefer to traverse quieter roads when they walk to and from school [114, 238].To accurately model the route that children will walk on their way to and from school, footpaths should be included where possible in GIS models. This chapter assesses whether including footpath data in network modelling makes a significant difference to modelled route lengths and associated exposures. To my knowledge, this work has not been previously undertaken.

The impact of including footpath data in the road network when modelling exposure to the RFE was assessed by calculating routes to school and associated exposures along the OSM, ITN and ITNUPs datasets. These modelled route lengths and exposures were compared with route data recorded from a survey of 13-16 year olds. In addition, the percentage of the modelled routes that intersected with the drawn routes have also been reported.

ITN is regarded as an authoritative dataset, or gold standard, as it is compiled by mapping and survey experts that work for the UK's national mapping agency,

Ordnance Survey (OS). Unless you are eligible for an academic license or Public Sector Mapping Agreement (PSMA) [386], OS products can have high procurement costs. However, the ITN and ITNUP datasets should be used with caution for the purposes of modelling walking routes. This is because OS footpath data records footpaths in urban areas of the UK greater than $5km^2$. However, footpaths elsewhere in the country are not recorded. This data is recorded in a dataset called ITNUP; the name being indicative of its coverage.

OSM is an open source map of the world, created by the public contributing data through the internet [387]. Generating data in this way is termed Volunteered Geographic Information (VGI) and it has allowed OSM to record detailed, world-wide data. The novel nature of OSM data collection and the uncertainties associated with this method [203, 239] means that quality checks are important. The first studies that evaluated OSM data quality used traditional methods of spatial data quality evaluation [199, 200, 246] which compared OSM to an authoritative dataset, which is deemed as the gold standard. The widely cited parameters for the quality of geographic data are defined by the International Organisation for Standardisation (ISO) [11]. These are: completeness [201, 202]; logical consistency [203]; positional accuracy [204–206]; temporal accuracy [207]; and thematic accuracy [208, 209].

In recent years, as the OSM dataset has grown, researchers have become concerned that traditional spatial quality indicators such as those reported by the ISO are not necessarily the most appropriate way to evaluate OSM. These concerns centre on the methods used to collect OSM data [258]. The view is that VGI data may need to be evaluated using different methods to those used to evaluate spatial data that is collected by professional mappers and surveyors. However, the discussion on the most appropriate method(s) to evaluate VGI continues and the current literature is equivocal. Determining an appropriate method of evaluating OSM was not within the scope of this project, so I decided to apply methods that have been previously undertaken on evaluating OSM data in the UK and other locations in the world. Many of the studies that have evaluated OSM quality have been implemented in more developed countries such as the UK , Germany [207, 249, 261], Austria [383], Ireland [248], Greece [208], France [203], the US [201, 202] and Japan [263]. These studies found that the quality of OSM data was homogenous across the study areas. These study areas were cities or regions, but I have assessed OSM for a whole country which

is a much larger geographic region than has been previously undertaken in developed countries. Elsewhere in the world the quality of OSM has been found to be a good source of spatial data but note that there will be limitations to its applications because of its heterogeneity (South Africa [388], Tehran, Iran [262]).

6.2 Methodology

6.2.1 Data Sources

6.2.1.1 OpenstreetMap

The OpenStreetMap project aims to create a map dataset of the world that is free to use, editable and licensed under copyright schemes that allow data to be accessed without any licences. OSM follows a "peer production" model whereby members can edit and contribute data to the project via the internet. In 2015, OSM had more than 2 million registered members [389]. These members are known as "contributors" and they edit and upload data through an online editor, Potlatch [385]. Contributors tend to supply data by uploading GPS tracks or digitising aerial data. In addition, OSM has taken advantage of mapping agencies making data open source. For example, the Netherlands became the first country in the world to have its roads fully mapped in OSM following a commercial navigator company donating a complete street map dataset to the OSM project. In the UK, OS released a suite of open source data including open roads data and in the USA, release of the Tiger dataset provided a detailed street and highway coverage for the entire country. This data has been incorporated in the OSM dataset.

Although OSM boasts a vast membership, different types of OSM contributors were identified by Neis and Zipf [249] and there remains a core group of contributors who dedicate their time to contributing data; improving OSM's infrastructure and developing and maintaining services [385].

6.2.1.2 Integrated Transport Network

OS' Integrated Transport Network is a "current and complete" network of roads in Great Britain [360]. The dataset contains rich routing information such as access

restrictions, one-way roads and toll indicators. The dataset consists of centrelines in a structured vertices-and-node network.

6.2.1.3 Integrated Transport Network Urban Paths

OS ITNUP [390] compliments the ITN as it is also a dataset of centrelines in a structured vertices-and-node network. The dataset details man made paths in urban areas of the UK that are greater than 5km^2. This includes subways, steps, footbridges and cycle paths.

The three road datasets were reclassified so to be comparable. The classifications are documented in Table 6.1.

6.2.1.4 Hand drawn routes

The collection and evaluation of this dataset is reported in full in Chapter 5. I met with a group of 11 volunteers, aged 13-16. The participants lived, and attended school in Cardiff, Swansea, Caerphilly and Merthyr Tydfil. Their routes to school were heterogeneous in urban/rural regions and socio-economic status.

6.2.1.5 Food Outlet Locations

Food outlet locations were obtained from the FSA dataset that was documented in Chapter 4. This data is postcode level data and was downloaded from the FSA website [346]. Unhealthy food outlets were extracted from the dataset according to the typology of unhealthy food outlets documented in Chapter 4 adapted from Lake et al [189]. This data is postcode level data and was downloaded from the FSA website.

6.2.2 Data Acquisition

The OSM dataset for Wales was downloaded as a database file from Geofabrik [391]. The ITN dataset was obtained from OS in a shapefile format. Both datasets were then imported into a spatially enabled database. How the OSM data was downloaded and imported into database management software is shown in Figure 6.1. The collection of the food outlet data that was used to calculate exposures is documented in Chapters 3 and 4. The collection of the routes to school data is documented in Chapter 6.

Download OSM data:	Convert database file to .osm format:	.osm file imported in to pgAdmin:	Road type selected:
OSM data was downloaded from Geofabrik in October 2015. The data is updated daily by the data providers.	Command line tool osmosis was used to convert the data from pbf to osm format. Osmosis was used to split the large OSM datasets into more manageable datasets.	.osm file was then imported in to a PostgreSQL spatially enable relational database (pgAdmin). The command line tool osm2pgsql was used.	The imported data included information on entities such as traffic lights, pedestrian crossings and rivers. The road data were copied in to a new tablet that only contained road data.

Figure 6.1. Methodology of downloading OSM data and importing the data into a spatially enabled database

6.2.3 Positional Accuracy

To investigate the positional accuracy of OSM, the test dataset, I generated a number of different sized buffers around each segment of road from the authoritative ITN dataset in PostGIS (Figure 6.2). The buffers were 3m, 5m, 8m and 10m. I generated buffers of different distances based on previous research that found these distances to be reasonable distances to expect OSM to intersect with authoritative datasets [206]. I then calculated the percentage of OSM roads that intersected with the ITN buffer. Positional accuracy was calculated for each category of road type for the whole of Wales. Positional accuracy was also calculated for each LSOA to investigate whether positional accuracy was associated with rurality. Urban and rural morphologies have been defined by the Office for National Statistics rurality index at LSOA level [392].

6.2.4 Completeness

Completeness was measured as the difference in length between OSM and ITN road datasets. I decided to measure completeness this way after considering the concept that in some cases OSM is more up to date than authoritative datasets. By representing completeness as the difference between OSM and ITN road lengths, it was possible to see where ITN had greater road coverage and where OSM had greater coverage.

6.2.5 Route Comparison

The suitability of OSM was further assessed by comparing mapping survey routes to school with GIS-generated routes. The comparison used a small sample of hand drawn routes to school compared with postcode level shortest network routes generated from three different road network datasets. The road network datasets were: Ordnance Survey Integrated Transport Network Roads; Ordnance Survey Integrated Transport Network Roads and Paths; and Open Street Map. The ITN dataset with and without footpaths was used to see whether the inclusion of footpaths improved the accuracy of route modelling. Previous research, investigating GIS generated routes compared with GPS generated routes, used Ordnance Survey Integrated Transport Network Roads as the road network dataset [11]. GIS generated routes have so far been found not to be an accurate representation of the environment that a child experiences along their route to school [11, 393]. I wanted to see whether including footpaths would improve exposure modelling.

Shortest network routes from home to school were calculated for the three road networks using the Network Analyst tool in ArcMap 10.1. The modelled routes were compared to the hand-drawn routes to evaluate the three datasets. The datasets were evaluated by analysing the difference in length of the modelled routes compared to the hand drawn routes. Due to the non-normal distribution of the hand drawn routes, a Spearman's rank correlation was calculated to measure the strength of correlation between the modelled route length and the hand drawn route length.

Exposure to the RFE was defined as the number of food outlets that had the potential to sell unhealthy food within 100m of the route that a child walked past on their route to school. Exposure scores were calculated for each of the four routes for each participant. The three modelled scores were compared to the hand drawn route exposures using a Spearman's rank correlation.

Table 6.1. Road types from OSM and ITN that were included in the road classification

Road Classification	OSM	ITN
Motorway	Motorway Motorway_link	Motorway
A Road	Trunk Primary Trunk_link Primary_link	A Road
B Road	Secondary Secondary_link	B Road
Minor Road	Tertiary Unclassified Residential Service Tertiary_link Living Street	Alley Local Street Minor Road Private Road - Publicly Accessible Private Road - Restricted Access
Footpath	Cycleway Footway Path Pedestrian Steps Bridleway	Pedestrianised Street Path Links
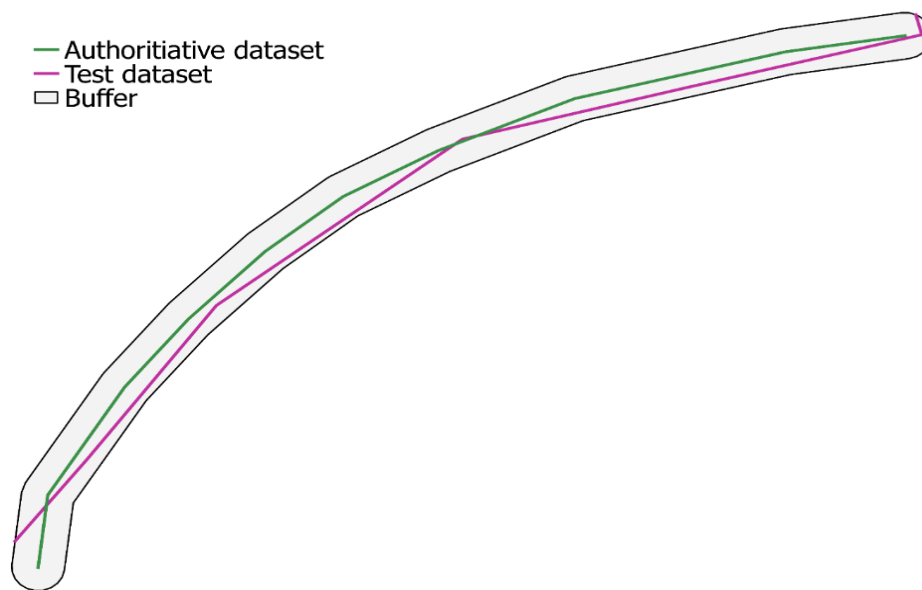
Figure 6.2 Goodchild and Hunter (1997) buffer comparison method. A buffer of width x is created around the authoritative dataset. The percentage of the test dataset that falls within the buffer is calculated. Based on visualisation from Haklay et al. (2010)

6.3 Results

6.3.1 Positional Accuracy

The road type with the greatest positional accuracy was A roads (Figure 6.3). 81% of A roads in Wales were within 3m of the ITN dataset where 81% of OSM A roads were found within 3m of the ITN dataset and 98% within 10m of ITN. OSM Motorways was the road type with greatest percentage of road within 10m of ITN (99%). Minor roads were the road type with the smallest percentage within 10m of the ITN dataset (82%). Footpaths had the smallest positional accuracy of the whole road network with 31% of footpaths within 3m of the ITN dataset. These results suggest that the main roads, have the greatest positional accuracy in the OSM dataset.

There was no systematic difference between the positional accuracy for roads in urban and rural areas. The greatest difference between positional accuracy for major roads in urban and rural areas was 5% for B roads, within a 3m buffer. The difference between footpath positional accuracy in urban and rural areas is more distinct. Footpaths in urban areas had a positional accuracy of 41% within a 3m buffer and footpaths in rural areas had a positional accuracy of 21% for the same buffer size. For

each buffer size, the positional accuracy of urban footpaths is, on average, double the positional accuracy of footpaths in rural areas (Table 6.2).
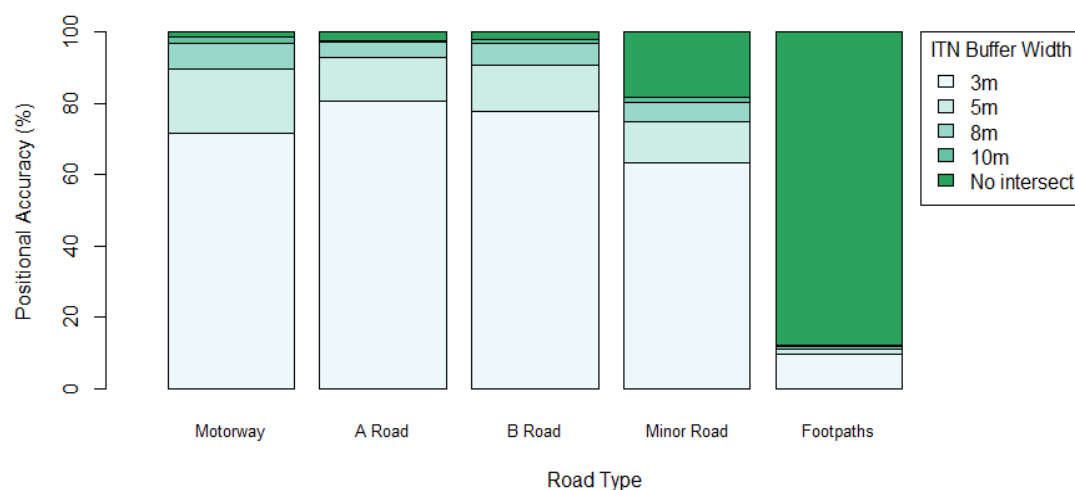
Figure 6.3. OSM Positional accuracy by road type for the whole of Wales

Table 6.2. OSM positional accuracy (PA) for Wales by urban and rural regions

	Motorway	A Road	B Road	Minor Road	Footpaths
			Urban		
3m PA (%)	73	86	84	76	41
5m PA (%)	90	96	95	87	46
8m PA(%)	97	98	98	91	50
10m PA (%)	99	98	98	92	50
			Rural		
3m PA (%)	74	82	80	71	21
5m PA (%)	92	94	92	83	24
8m PA (%)	98	98	97	89	26
10m PA (%)	99	99	98	90	26

6.3.2 Completeness

The completeness of OSM for major roads (Motorway, A roads, B roads) was very high (Figure 6.4). The total length of B roads recorded in the OSM dataset was 30km shorter than the total length of B roads in the ITN dataset. However, OSM recorded a greater length of Motorway and A roads than the ITN dataset; 8km and 46km respectively. Completeness for Minor roads was poorer. The OSM dataset was 9111 km shorter than the ITN Minor roads (Figure 6.5). This is a relatively small difference of 2.3 km of road per km^2 of Wales. The vast majority of the omitted roads (93%) were in rural regions.

In contrast, the OSM dataset contained 6,807 km more footpaths than the ITN dataset. Urban regions recorded 508 km more footpaths in OSM than ITN. The total length of OSM footpaths was three times longer than ITN footpaths, recording 7,000 km more footpath than the authoritative ITN dataset. The ITN dataset records over 3000 km of footpaths but the length of shared ITN and OSM footpaths is 1,230 km. OSM records more than 3500 km of footpaths for rural regions (Figure 6.6).
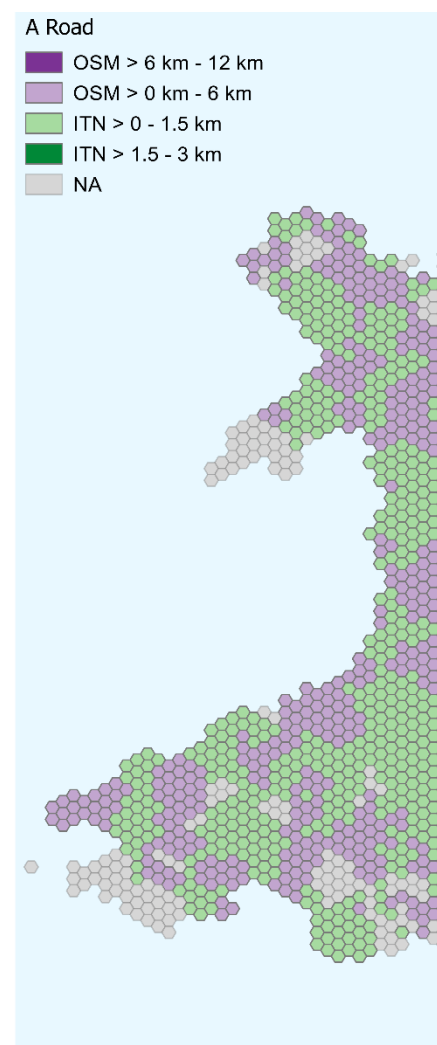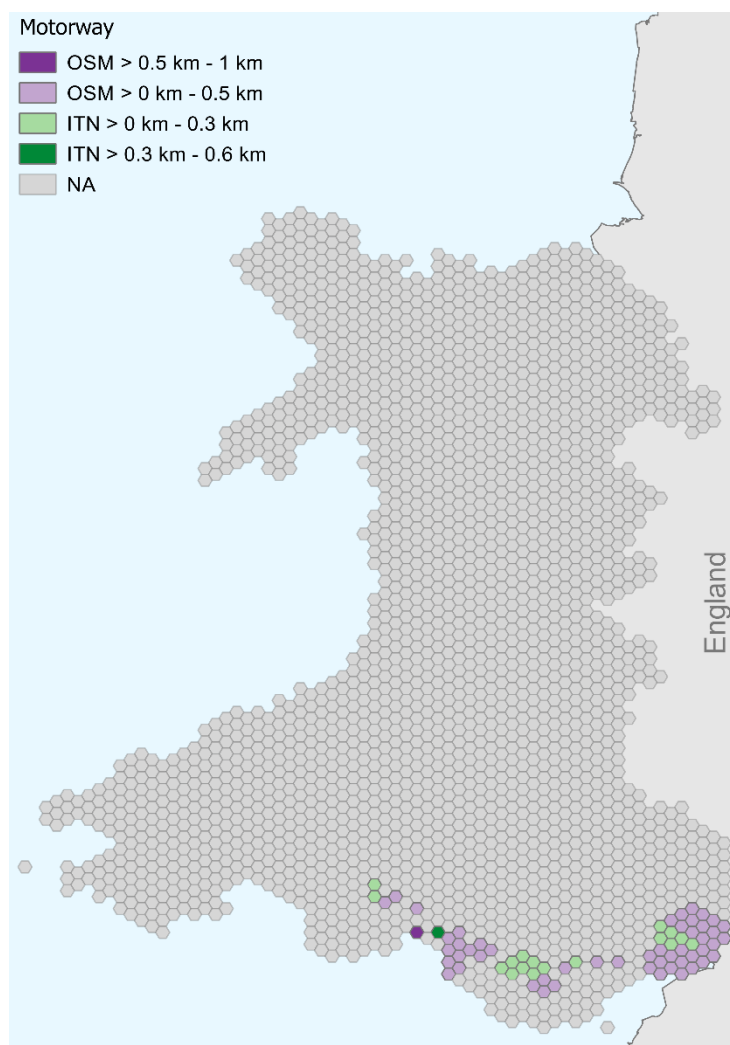
Figure 6.4. Hexbin map shows where either ITN or OSM record greater length of Motorway and A roads in Wales.
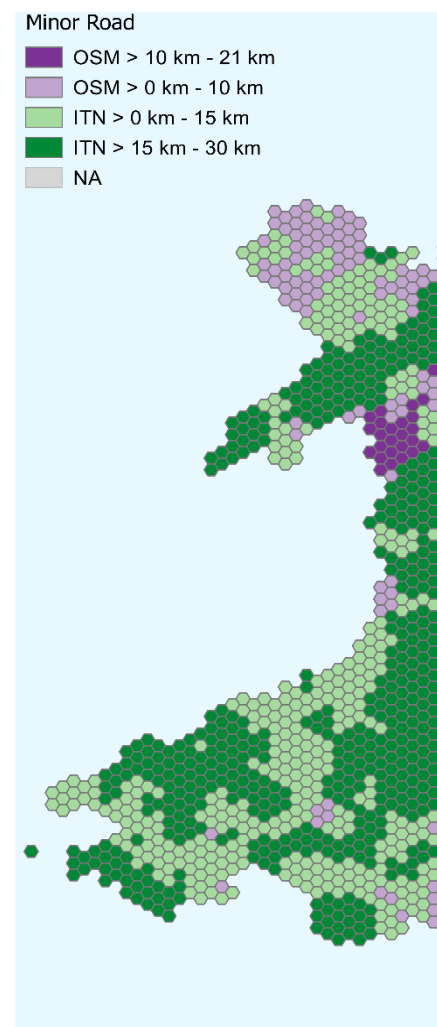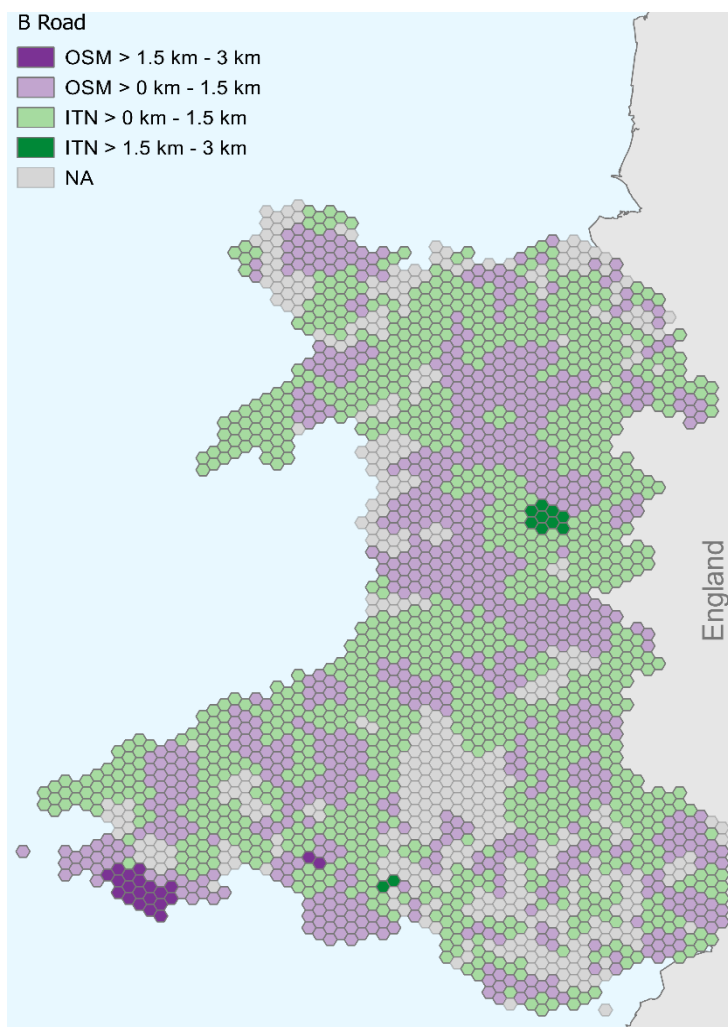
Figure 6.5. Hexbin map shows where either ITN or OSM record greater length of B roads and minor roads in Wales
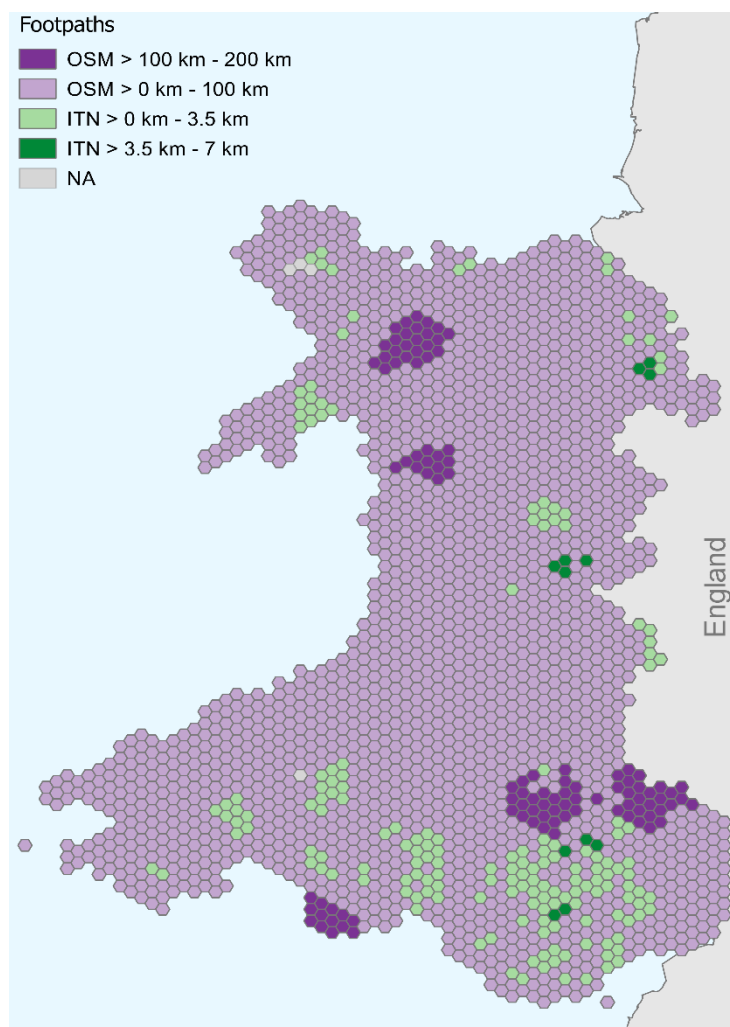
Figure 6.6. Hexbin map shows where either ITN or OSM record greater length of footpaths in Wales.

6.3.3 Route Exposure Evaluation

Eleven, hand drawn walking routes to school were collected from the ALPHA group. The ALPHA group and this data collection methodology are detailed in Chapter 5. The average length of the drawn walking routes to school was 1820m. The shortest distance walked to school was 706m and the longest distance recorded was 4936m. The majority of routes were shorter than 2100m. The average exposure to the RFE was 2 unhealthy food outlets.

The smallest average difference between the modelled route lengths and the drawn route lengths were for the routes modelled with the OSM dataset (245 m). The average distance of the OSM modelled routes was 1955m. The distribution of the OSM modelled route lengths were the most similar of the modelled routes to the drawn routes lengths. Two routes calculated along the OSM road network were the same length as the drawn routes (*Figure 6.7*).

The greatest difference in route lengths were in the routes modelled along the ITN Roads network. The greatest difference in length of routes calculated using the ITN dataset when compared with the drawn routes was 1,691 m. The average length of the routes calculated using the ITN dataset was 1832m. The shortest distance was 711m and the longest distance was 3789m.

The average length of the ITNUP generated routes was 1894m. The shortest distance was 710m and the longest route was 3846m. The greatest difference between the routes calculated using the ITNUP dataset and the drawn routes was 1147m.

For all four groups of routes, the most frequent route lengths were between 1500 and 2100m (Table 6.3). When compared with the drawn routes, OSM ($\rho = 0.85$, $p < 0.01$) and ITN ($\rho = 0.70$, $p = 0.02$) modelled routes lengths were significantly correlated with the drawn route lengths. ITNUP ($\rho = 0.50$, $p = 0.11$) route lengths were weakly correlated with the drawn route lengths and the results were not statistically significant.
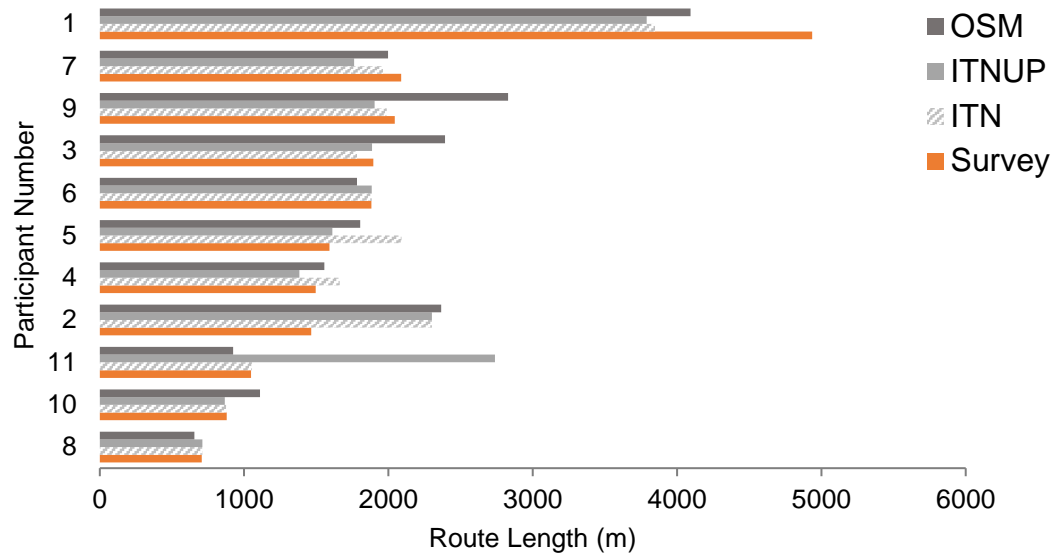
Figure 6.7. Route length for each network compared to the survey data grouped by participant. The graph is ordered by ascending survey route length

Table 6.3. Difference in metres between the drawn route lengths and modelled route lengths. Drawn route lengths rounded to nearest 50m to avoid disclosure

Route length (m)	OSM difference	ITN difference	ITNUP difference
700	-51	5	5
900	230	-3	-13
1000	-123	7	1692
1500	899	835	835
1500	61	167	-113
1600	212	499	20
1900	-99	2	2
1900	498	-113	-8
2000	784	-55	-140
2100	-91	-127	-325
5000	-843	-1090	-1147

The average exposure to the RFE along the OSM modelled routes was 1. Seven routes had no exposure to the RFE. The average exposure to the RFE along the ITN modelled routes was 1. Seven routes had an exposure of 0. The greatest exposure was 10 outlets along the shortest route of 701m. This was for the same participant as the hand drawn route. The average exposure to the RFE along the ITNUP modelled routes was 3. Six routes had no exposure. The greatest exposure to the RFE was 12.

Seven drawn routes had an exposure of 0 to the RFE. The greatest exposure along the drawn routes was 10 unhealthy outlets. This was along the shortest route of 700m. Exposure to the RFE was over-estimated by each of the three modelled routes. The exposures calculated from OSM (ρ = 0.28, p= 0.41), ITN (ρ = 0.42, p=0.20) and ITNUP (ρ = 0.13, p= 0.69) datasets were weakly correlated and not statistically significant.

6.4 Discussion

6.4.1 Main Findings

The evaluative assessments have shown that OSM has excellent positional accuracy and completeness for main roads (Motorway, A roads and B roads) in Wales. These findings are consistent with what has been found elsewhere in the UK and other locations across the world [204, 205, 260]. The length of motorway and A roads recorded in the OSM dataset was greater than the length of motorway and A roads recorded in ITN. However, this should not suggest that the ITN dataset is incomplete for main roads. Rather, the relatively small discrepancy in length could be due to several reasons:

1) Differences in the labelling of road type between the datasets; this resulted in a greater length of road re-classified as motorway and A road in OSM for this study;
2) How OSM data is collated; OSM will have much of its data contributed through GPS data, and errors may be introduced due to positional accuracy of the GPS data.

In contrast, minor roads recorded in the OSM dataset were more than 9000km shorter than minor roads recorded in the ITN dataset. The minor roads that are recorded in the OSM dataset do, however, have a high level of positional accuracy (86% within 5m of ITN). Minor roads recorded in the OSM dataset have not been evaluated before in the UK, but the findings here are consistent with other European studies [259]. The majority (93%) of minor roads not recorded in the OSM dataset were in rural regions, and this suggests an urban/rural bias in Wales. That is, OSM records data on rural regions to a lesser extent than urban areas, this has also been reported elsewhere [394–396]. Rural regions that lack OSM coverage are very sparsely populated; 26 people per km^2. Sparsely populated areas have fewer contributions [251] and this directly impacts completeness of the OSM dataset [203, 261]. In addition, the long, winding roads in rural Wales are less like to be contributed than straight, short roads [397].

A further key finding of this evaluation was that the OSM dataset has more than 508 km of footpath data for urban regions than the ITNUP dataset. A reason for this difference may be because of the purpose that ITNUP is designed for. ITNUP records data on footpaths in urban areas greater than $5km^2$. This means that footpaths elsewhere in the UK are not recorded in the ITN dataset. The footpaths that are recorded in ITNUP were compared with footpaths recorded in OSM and OSM returned a poor positional accuracy. Given that we know that OSM has recorded a far greater *length* of footpaths compared with ITNUP, this suggests that the footpaths recorded in ITNUP may in fact not be recorded in OSM. Something to explore in the future may be to combine the ITNUP and OSM dataset to improve completeness of the road dataset. This would be akin the method that Dessing et al. [238] implemented in investigating route choice during active transportation to school; the Dutch authoritative road dataset was combined with OSM.

The greatest coverage of OSM footpaths was seen in the areas in and around Wales' National Parks and Area of Outstanding National Beauty. These are sparsely populated areas but receive high numbers of visitors that enjoy outdoor activities. This demographic have previously been linked with being significant contributors to OSM [201]. Furthermore, there are specific projects dedicated to recording official walking routes for OSM in the UK [398]. Projects like this will focus contributor's efforts on specific projects and reduce the urban-rural bias. This extensive rural coverage will

contribute to the overall footpath coverage of Wales which is more than 6000 km greater than the ITN.

It would be unmerited to simply claim that OSM footpath coverage is better than ITN footpath dataset. The ITNUP dataset does not claim to be a country-wide dataset but clearly states its coverage. However, the findings of this chapter suggest that OSM may be more appropriate to use in modelling exposure to the RFE along walking routes. This is because footpaths are far more widely recorded in the OSM dataset, compared with the ITN dataset. Conversely, the poor positional accuracy of OSM suggests that OSM and ITNUP may have different coverages as OSM does not intersect with a high percentage of the footpaths that ITNUP records.

Routes modelled with the OSM dataset produced very strong correlations with the drawn route lengths. This may be because in the routes where children did use footpaths, OSM had these paths recorded in the dataset. Including footpaths in the network captures more alternative networks along which children may tend to traverse and therefore better represents the length of the route that they walk. The ITN routes were also strongly correlated with the drawn lengths. Where children did not traverse footpaths along their route, ITN was a reliable data source to predict the distance that the participant travelled to school. This supports findings elsewhere [11].

With regard to modelling exposure to the RFE along children's walking routes, these results do not clearly show that including footpaths in the road network improves modelled environmental exposures. However, the difference in the strength of correlation between route lengths and exposures modelled from different road networks, highlights that the road network used to model walking routes influences the route that is modelled. Food outlets tend to be located along main roads and the literature proposes that where possible, children prefer to traverse quieter roads and use "unofficial" networks. Therefore, all aspects of road data are important in modelling children's exposure to the RFE.

It may be that the method of calculating the route by the shortest network route may not be appropriate for capturing children's route choice on their walk to and from school. This was demonstrated when one of the participants travelled almost a kilometre further than the modelled routes predicted. This was because they traversed a footpath along a river to a friend's house. The literature states that GPS routes are

currently the only method of accurately investigating exposure environments [393] as they can record this "selective mobility bias" [267, 358]. Although GPS routes are excellent tools for providing accurate route length and route exposure methods [267, 281], it is important to explore the opportunity to improve the accuracy of GIS modelled walking routes. This is because GIS offers the potential to routinely generate population level exposure analysis and enable the large-scale evaluation of environment and health. This will be a powerful tool for public health researchers and practitioners.

6.4.2 Strengths

A strength of this investigation was that a large geographic area was analysed; most previous studies have focussed on cities or regions. This chapter promotes the notion that data quality evaluation is important before using any dataset. No dataset perfectly represents the real world, but seeking to use the most appropriate data available and being aware of its limitations is fundamental for high quality research.

The results in this chapter show that using different networks yields different results when modelling route lengths and route exposures. This has not been reported before and highlights that the road network used is an important consideration when modelling exposures. The results show that OSM has a greater footpath coverage than ITNUP and this should be considered when modelling child active travel. OSM contain more "unofficial" paths which are likely to be traversed by children.

Evaluating OSM and reporting its appropriateness for modelling walking routes will be a valuable addition to the literature as OSM provides international coverage, derived data from authoritative sources, access for non-PSMA users in the UK and daily updates. Using OSM in studies means that it possible undertake national and international comparisons between locations i.e. applying the same method in a different geographical location; and also between studies i.e. differences between methodologies can be explored using the same data source. This reproducibility is an important aspect of research, particularly public health as it allows for comparison, sharing of knowledge and reduces duplication.

6.4.3 Limitations

In deciding whether comparing OSM with an authoritative dataset is an appropriate evaluative method, a further point to consider is that open source data from OS have been used to contribute to the OSM dataset [264, 399]. Although the open source data is not as detailed at the ITN dataset, when mapping agencies and companies make data available as open source, it will often be used to contribute to OSM. It could therefore be argued that in comparing OSM with the OS authoritative data set, I am just comparing like for like.

A further limitation is associated with the classification of OSM roads. The results show that OSM has poorer coverage of minor roads than ITN. However, a contributing factor to the discrepancy between OSM minor road and ITN minor road length is that OSM and ITN do not have directly comparable road types. OSM is an international dataset so not only are there different names for road types compared with the ITN, there may be differences in how users classify roads. OSM do provide classification guidance, however, this is still open to interpretation by the contributor. Brovelli et al. have attempted to appease the technical challenges of comparing OSM with an authoritative dataset and designed open source software to quantitatively compare OSM with an authoritative dataset [389].

The sample of hand drawn routes to school was a morphologically diverse sample across rural and urban areas. However, the sample was small and only routes to school were requested. To more fully understand children's route choices on their walk to and from school, a larger sample will need to be collected. There are limitations with using drawn routes to represent the route a child travels. This method depended on the students being spatially aware and map literate. I did, however, take these limitations into account when I designed the sample. This was to ensure that the route that the students had documented were as accurate as possible.

6.4.4 Implications

There are several implications from the results from this chapter on my PhD project and also for the wider literature.

With regard to my PhD project, OSM will be used as the road dataset for modelling children's exposure to the RFE. As already discussed, OSM has reliable coverage of

main roads and records far more length of footpaths than ITNUP does. The regions that lacked minor road coverage were in very sparsely populated regions of Wales (mid-Wales, national parks and Areas of Outstanding National Beauty (AONB)); regions that will not be included in this study. In addition, OSM provides the opportunity for this research to be reproduced elsewhere in the world. Sharing research and resources is fundamental for public health researchers to avoid duplication of efforts. Not sharing knowledge and skills will prolong finding potential interventions for issues that harm the public's health.

For the wider literature, this chapter has displayed results that suggest including footpath data in the road network may provide a way to improve the accuracy of GIS modelled exposures. The results do not overwhelmingly express this but the change in correlation of the route length and exposure between road networks suggests that different road datasets produce different modelled routes. Added to what we know about children preferring to use footpaths and "unofficial" networks when walking [113], I propose that on a larger scale, including footpaths will contribute to more accurate modelling of children's walking routes in a GIS.

Including footpaths in to the road network, did not however, convincingly improve the correlation between modelled exposures to the RFE and exposures calculated from hand drawn routes. This suggests that child route choice is more complex than traversing the shortest route; for example children often choose to avoid busy roads on their way to school [238].

Future studies should not be discouraged from using OSM by the results reported in this chapter for minor road coverage. This is because OSM is regularly updated and a paper published in 2017 found that OSM is more than 83% complete globally and more than 40% of countries have a fully mapped street network [400]. Studies looking to use OSM should asses the appropriateness of OSM for their specific project.

Due to the small sample of hand drawn routes, I did not want to overstate the value of the comparison with the GIS modelled routes. The use of hand drawn routes was appropriate for this preliminary study. However, the collection of hand drawn routes was a resource and time intensive practice. This preliminary cohort of hand drawn routes provoked me to decide that I needed to collect a larger sample of GPS data to: a) better understand a child's route choice on their way to and from school; b) develop

a better-informed exposure mode that included other characteristics of children's route choice. For example, including in the model to avoid steep roads [401]; avoid woodlands & major roads [56]; along pavements where possible [127]; and past or through parks [402]. This process is documented in the next chapter, Chapter 6.

6.5 Conclusions

When compared with the ITN dataset, the OSM dataset is complete for main roads. Caution should be taken in rural areas as minor road coverage can be poor as roads are not documented in the dataset. However, the OSM dataset has a more comprehensive coverage of footpaths in Wales than the ITNUP dataset. For the purposes of this study, OSM will be used as the road data source. Footpaths have been found to be important in children's route choice in their walk to and from school, and therefore OSM is the most appropriate dataset to use in this study. Furthermore, OSM provides international coverage, derived data from authoritative sources, access for non-PSMA users in the UK and daily updates. This will allow for national and international comparisons to be made from this study.

Future work would should investigate a hybrid network that contains both ITN and OSM data. However, an updated comparison may be first advised as a paper published in 2017 found that OSM is more than 83% complete globally and more than 40% of countries have a fully mapped street network [400].

The results in this chapter show that using different networks yields different results when modelling route lengths and route exposures. Therefore, the road network chosen is an important consideration when modelling exposures. The results show that OSM has a greater footpath coverage than ITNUP. This is an important factor to consider when modelling child active travel. OSM appeared to contain more "unofficial" paths; paths that are likely to be traversed by children.

Chapter 7

GPS Data Acquisition

7.1 Introduction

The cohort of data collected in Chapter 6 was not large enough to make assumptions that could be generalised to the UK population. A large, detailed sample of GPS data was required to identify environmental characteristics associated with route choice in teenagers aged 13-14 years along their route to or from school. This GPS data will also be used to validate the GIS modelled exposures to the RFE.

I obtained a large sample of GPS data for 982 children aged 13-14 from researchers who had worked on a large-scale study called the PEAR study [403]. The Physical Environment and Activity Relationships (in adolescents) (PEAR) study was an MRC funded project that builds on PEACH study. It is a cross-sectional study of year 9 students from Bristol, South Gloucestershire, North Somerset and Bath and North East Somerset. The study was run by Bristol University in partnership with University of East Anglia, Queen Mary University and Cardiff University. Further details on the PEAR study are documented in Appendix 5.

The PEAR study had substantial funding (£798,304) and time allocated to the project (3 years). This meant that the project had significant resources to spend time recruiting individuals and to record data for such a large number of participants. A team of 14 people worked on this project with two analysts committed to collecting this data. In contrast, for my project, I was working on my own with a limited time frame, compared to the PEAR study, to collect GPS data (~6 months).

This chapter documents the preparation of the raw GPS data for analysis. The aim of this analysis is to produce a criteria based algorithm that can:

1. Clean a GPS dataset;
2. Extract walking routes to and from school;

Several methods of cleaning and extracting have been used in the literature but they all tend to split the raw GPS data into smaller partitions which represent single journeys [281]. Following this, analysis can be undertaken on these smaller segments. Criteria-base methods are the least technically demanding of these approaches, but they have been successfully implemented other studies [288, 289].

A cohort of 982 children was extremely valuable and collecting a dataset as large or as detailed as the PEAR study was out of the scope of this PhD. The GPS data collected by the PEAR study was a reliable source of GPS data because the data was collected by trained researchers that are experts in GPS data collection and analysis. The GPS data were sent to me from the data custodians via password protected cloud sharing. The data was imported, stored and managed in pgAdmin using postgreSQL.

7.2 Methodology

7.2.1 Data sources

The raw GPS PEAR data was received from the data providers in csv format. Home locations of the PEAR participants were provided as XY coordinates in a csv file from the data providers. OS Mastermap Topography Layer [404] provided school and building footprints that were downloaded from DigiMap [347].The csv files were imported in to PostGIS database using the command line tool pgfutter. Home and school locations were imported into PostGIS for analysis.

A criteria-based extraction process was produced to prepare the PEAR dataset for analysis. The method was based on other successful criteria-based methods [288, 289]. The extraction method is shown in Appendix 6. There were three main stages to the preparing the data: pre-processing; processing and post-processing. This methodology is summarised in Figure 7.1.

7.2.2 Pre-processing

Criteria were developed to remove points that were not from the route to or from school. These criteria were advised by a researcher that has previously manually extracted walking routes to and from school [11]. Participants were excluded if no home and/or school locations were recorded (n=5). The chosen criteria were:

1. Remove data from weekends;
2. Remove GPS points that were not recorded during prime commuting times (before 07:30, between 9:30 and 14:30, after 16:30);
3. Remove points within school building footprint;
   a. School footprints were provided by OS Mastermap [404]
4. Remove points within 40m of home;
   a. Point home locations were provided by PEAR research team
5. Remove points that were within building foodt prints.
   a. Polygon building footprints were used from OS Mastermap [404]

7.2.3 Processing

GPS points were indexed to indicate each individual trip made by each participant. For each trip, GPS points were ordered by their timestamp. A trip was defined as consecutive points on a single day during a morning commute time (between 7:30 and 9:30), or during an afternoon commute time (between 14:30 and 16:30). Each trip was assumed to represent a participant travelling from home to school, or from school to home. The GPS points for each trip were then used to create line geometries. Home and school locations were appended on to the line geometries because they had been removed in the pre-processing stage. Walking routes were extracted from the data based on average speed of the points that made up each trip (av speed < 10m/s).

7.2.4 Post-processing

Manual cleaning removed any remaining motorised transport trips that had been wrongly classified as a walking route. Remaining outliers were removed by selecting points that were more than 100m from any other GPS points belonging to that trip. Routes were manually inspected to ensure that only the route from home to school and school to home had been captured. Any points representing other journeys were

removed. For example, some children once they arrived home, left home again before 16 30. These points were removed.
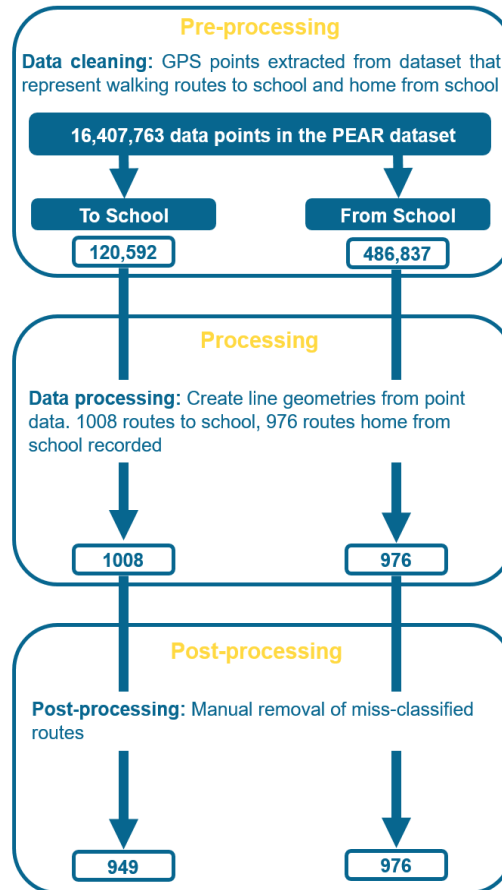
Figure 7.1. Summary of criteria based algorithm to clean GPS data and extract walking routes

7.3 Results

A total of 16,407,763 GPS points were recorded in the PEAR dataset. The number of points removed at each stage of data processing are reported in Figure 7.2. The majority of GPS points were recorded on week days (72%). Approximately 1.5 million points were recorded during the time of the morning commute and more than 3 million points were recorded during the time allocated for the route home. Of these, more than 1.4 million points were removed from the morning commute and 2.5 million from the afternoon commute as the points were recorded inside buildings. 1,628 points were

removed manually to leave points that recorded walking routes to school. For walking routes home from school, 7,960 points were removed manually.

After processing, 120,592 and 486,837 points remained for trips to school and trips home, respectively. These points represented 949 trips to school and 976 trips home. The trips to school were recorded by 377 individuals and the trips home were recorded by 409 individuals. In total, 884 individuals provided GPS data on their walking route to school *or* home. 333 children provided GPS data on their walking route to school *and* walking route home.
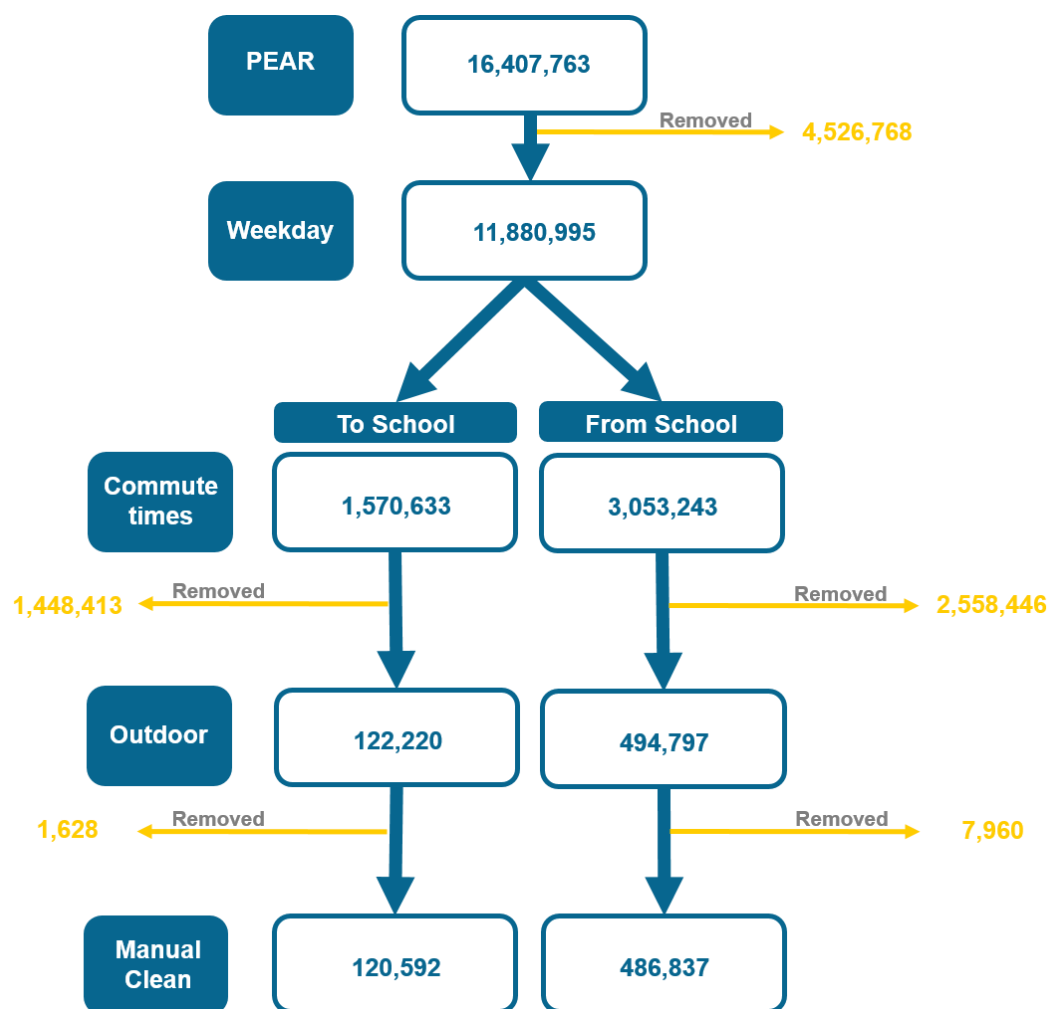
Figure 7.2 Points removed during data cleaning (n)

7.4 Discussion

7.4.1 Main Findings

Automating the data cleaning and processing has resulted in a detailed and robust dataset that can be used to investigate environmental characteristics associated with children's routes choice along their walk to and from school. Distinguishing between the walk to and walk home from school has not, to my knowledge, been documented before. The literature suggests that children behave differently between their the walk to and walk home from school. Therefore, this dataset will be a powerful tool to analyse the extent to which routes, and thus exposure to the 'retail food environment' (RFE), differ between children's walking routes to and walking routes home from school. Now these data have been extracted, many different types of exposure environments can be explored. This thesis will explore exposure to the RFE, but using the relevant environmental data, the route data could also be used to explore exposure to greenspaces, pollution etc.

This algorithm will be shared with the PEAR team as this dataset had not been previously cleaned and processed. The criteria-based algorithm could be modified to process the raw data to extract routes for different modes of transport.

Processing the GPS data was a time intensive process which required substantial expertise. The literature agrees with this finding. It states that cleaning GPS data is a technical process and that there can be difficulty when interpreting the data. The majority of studies working with GPS data, have so far manually cleaned the data or used rudimentary criteria. Furthermore, studies have been criticised for not differentiating between travel modes when calculating environmental exposures from GPS routes. As an example, researchers may only be interested in foot and cycle journeys as passengers in motorised vehicles will potentially have lesser opportunity to buy food on their journey. Cetateanu et al. [281] stated that this problem has largely been ignored. A study by Wheeler et al. [275] used crude travel speed criteria to differentiate between travel modes, and Harrison et al. [11] manually cleaned the GPS data. Other studies have attempted to develop more robust cleaning algorithms in order to be able to extract meaningful information from the data, such as travel mode [283–287]. Feng et al [286] used a Bayesian Belief Network model to identify travel mode

from GPS data and they found that accelerometer data can make an important contribution to imputing the correct travel mode. Furthermore, Lin et al [287] used an unsupervised method to detect travel mode and also reported that accelerometer data is important for distinguishing travel mode, particularly between high speed models. Extending my method to distinguish travel mode for each PEAR participant would need to make use of the accelerometer data in the dataset. This is because applying criteria that are based on expert-chosen rules may not be applicable between datasets [281]. Arguably, the most sophisticated method of GPS cleaning was published by Cetateanu et al. [281] whereby the researchers had developed an unsupervised algorithm to clean GPS data and distinguish between travel modes.

7.4.2 Strengths

A strength of this method is that I have automated the cleaning of the raw data and extracted walking routes without the need for additional training data. The criteria based method has successfully extracted walking routes to and from school from the PEAR dataset. Several studies have recently developed machine learning tools to extract motorised transport modes from raw GPS data. However, without any training data, this was not possible.

This criteria based method does not classify slow moving transport as a walking route because the speed criteria is based on average speed of the whole journey, rather than the speed of each individual point. This allowed the algorithm to discount routes that consisted of more than one mode of travel and therefore were not a walking route. This is a success of the algorithm as this has been a criticism of previous studies that have used a criteria based method. The inclusion of the home and school location in the extraction of the walking routes also resulted in the algorithm being able to identify direct routes to and from school and indirect routes. In addition, by appending the home and school locations to the route geometries, during manual cleaning, it was easy for me to identify routes that were not eligible to be classified as walking routes. The success of the algorithm is shown in the very small percentage of the points that were manually removed.

This algorithm has been written using an open source programming language and therefore it can be freely shared and developed. This will contribute to the

development of more accurate travel mode identification methods and also prevent replication of work.

7.4.3 Limitations

A limitation of this method is that if a child does not complete their journey from home to school or school to home within the defined commuter hours, then the journey was not included in the final routes. Therefore, if pupils travel to school from a location other than their home, or stop at another location on their way home (e.g. a friend's home, a sports club) this activity is not captured by the algorithm. This should be developed in future research as these routes could hold important insight in to the way that children travel to and from school.

7.4.4 Implications

The implications of these results on this PhD project is that my project has a robust, detailed, dataset that can be used to identify environmental characteristics associated with route choice in teenagers aged 13-14 years along their route to or from school. GPS data are objective behaviour measures and very rich data sources and therefore this GPS data can also be used to validate the GIS modelled exposures to the RFE. Population level research is not possible using GPS data. However, current GIS modelling methods do not provide us with enough detail to understand complex processes, such as children's autonomous travelling behaviours. It is important that researchers draw from both data types and use them in a complimentary way to produce research that can inform evidence-based policy.

To my knowledge, this is the first criteria based method to specifically extract walking routes to and from school. This is a powerful tool as it adds value to the GPS by specifically extracting walking routes to and from school. As researchers are becoming increasingly interested with accurately characterising daily activity spaces, this is a valuable addition to the data.

Future developments of this research could include developing the criteria based method to extract other travel modes to and from school and also to explore the use of

machine learning methodologies to undertake supervised or unsupervised extraction of GPS journeys.

7.5 Conclusions

A criteria-based algorithm has been produced to clean GPS data and extract walking routes to and from school from the data (see Appendix 6.). Further work could include developing a machine learning method that uses supervised or unsupervised methods to extract walking routes to and from school, such as that developed by Cetateanu et al. [281].

Chapter 8

Weighted Network Exposure Model

8.1 Introduction

This chapter documents a methodology that I have developed to generate population level exposure to the 'retail food environment' (RFE) along children's walking commute to school and home. The aims of this chapter are to:

- Investigate which environmental characteristics differ between GPS routes to and from school, and shortest network routes (SNR) to and from school;
- Incorporate the environmental characteristics that differ to GIS methodology by using a weighted network to emulate route choice more realistically;
- Use the weighted network to model routes to and from school in a GIS for individuals that provided GPS data;
- Evaluate the accuracy of the routes to and from school modelled using the weighted network.

These aims will be achieved by fulfilling several objectives. Firstly, SNR will be calculated for all participants from the PEAR study that provided GPS data (cleaning and preparation of this data was documented in Chapter 7). Environmental characteristics will be calculated for SNR and GPS routes, and the differences between SNR and GPS routes will be analysed using a conditional logistic regression. The results of the conditional logistic regression will be used to inform the impedance values of two weighted networks; a weighted network for the route to school, and a weighted network for the route home. Routes to and from school will be calculated using the weighted networks for the PEAR participants. The accuracy of the weighted network routes (WNRs) will be analysed by comparing their length and exposures

with the GPS routes. A multilevel regression model will also be run to evaluate the association between the GPS routes and WNRs.

The RFE is receiving greater attention from researchers and policy makers as a way to explore obesity-related behaviours. How the RFE is influencing childhood obesity is not a widely explored research area. Many studies that have investigated the link between the RFE and obesogenic behaviours have tended to focus on a single activity space that individuals will interact with in their day to day lives. For example, studies have focussed on the RFE either around the home, or their school [68, 89, 92, 223, 234]. The majority of studies that have focussed on the link between childhood obesity and the RFE have focussed on the RFE surrounding schools, and policies have been adopted by Local Authorities (LAs) in the UK limiting fast food outlets within 400m of school premises [405]. However, a systematic review published in 2014 [14] found little evidence to suggest that the RFE surrounding schools influences food purchases and consumption. The review did find some evidence that the RFE may have an effect on body weight but the authors suggested that this may be a result of residual confounding. There is a conflicting evidence base amongst RFE studies and this has been attributed to the methodological inconsistencies between studies. Studies have used different data sources, defined the RFE in different ways, and used different GIS methods to define the RFE [70].

Children's journeys through the RFE on their way to and from school have been investigated by two studies in the UK [11, 169]. These studies have used GIS to predict exposure to the RFE along a child's commute to and from school. Harrison et al. [11] and Griffiths et al. [169] modelled children's commutes to and from school using shortest network analysis in a GIS. Griffiths et al. concluded that there "is little support for the notion that exposure to food outlets in the home, school and commuting neighbourhoods increase the risk of obesity in children". However, GIS-generated commutes have been found to be a suitable proxy for the distance a child travels but not for calculating the RFE they are exposed to [11].

The Foresight report [9] highlighted the complexity of the obesity epidemic by stating that change is required at individual and population levels, across multiple sectors. Therefore, studies investigating obesity and obesogenic behaviours should acknowledge the complexity of the issue and the study design should reflect this. I

decided to investigate whether developing a more sophisticated GIS method could be used to improve the accuracy of GIS modelled RFE exposures along children's routes to and from school. This could allow population modelling of individual level environment and health data.

8.2 Methodology

All GPS processing and GIS analysis was undertaken in PostGIS (using pgadmin3 version 9.5). The methodology is summarised in Figure 8.1.
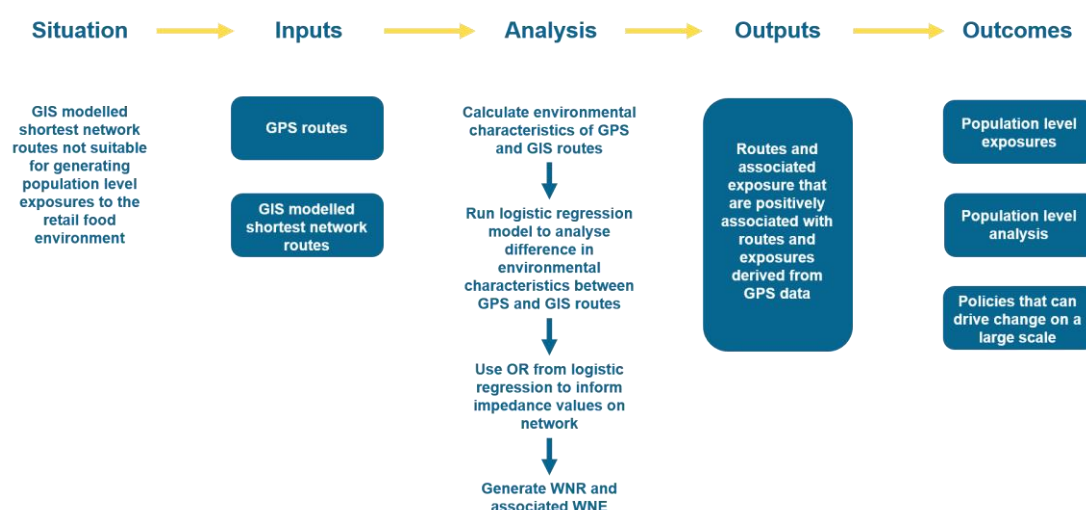
Figure 8.1. Summary of methodology

8.2.1 Data Preparation

8.2.1.1 GPS routes

The cleaning and processing of GPS data from the PEAR project was documented at length in Chapter 7. The PEAR data provided 949 walking routes to school and 976 walking routes home from school, for 884 individuals.

8.2.1.2 GIS generated shortest network routes

Shortest Network Routes were also calculated for the 884 PEAR participants that walked to school. The SNRs were calculated from home to school locations. The home point locations were provided by the PEAR research team. The name of the school attended by each participant was also provided by the PEAR research team. School point locations were represented by the school access point closest to the home location, based on Euclidean distance. These access points were downloaded from OS Mastermap [404].

8.2.1.3 Road network

The road network used to calculate the GIS modelled routes was generated from the OSM dataset documented in Chapter 5. The road dataset was noded using pgRouting. This process is documented in Appendix 7.

8.2.2 Environmental Characteristics

Characteristics of the environment that have been documented in the literature as influencing route choice [56, 238, 406] were calculated for both the GPS walking routes, and the corresponding GIS generated SNRs. Table 8.1 defines the environmental characteristics that were calculated (see Appendix 7.).

Greenspace has been defined as accessible green spaces that have been shown in the literature to influence child route choice. These green environments were accessible grass fields, meadows, parks and recreation grounds. Greenspaces that were not accessible were not included as I was interested in the green environments that are discussed in the literature as influencing child route choice. I have not accounted for the different ways that children who live in urban or rural areas may be influenced by green environments. This is because a small number of children in the PEAR study that lived in rural areas walked to school. A reason for this may be that children living in rural areas usually live further away from their school (>5km) and rural roads often do not have pavements which means it can be more difficult to safely walk to school.

Table 8.1. Environmental characteristics calculated along GPS and GIS modelled routes in PostGIS; method and data source

Environmental Characteristic	How to measure	Data Source
Length of route (m)	Calculate length of route	
Green space along route (%)	Percentage of route within 25m of a green space	Greenspace polygon data were downloaded from OpenStreetMap (OSM) [391] which included all green environments that have been shown influence children's route choice.
Blue space along route (%)	Percentage route within 25m of a blue space	Bluespace polygon data were downloaded from Ordnance Survey (OS) Open Rivers [407], Meridian 2 [408] and OSM [391]. The three datasets were merged to include areas where there were gaps in each dataset's coverage
Traffic lights (n)	Count of points along route	Traffic light point data were obtained from OSM [391]
Accidents (n)	Count of points along route	Road traffic accident data was downloaded from Stats19 [409]. Accidents that occurred between the school commuting hours (7:30-9:30 and 14:30-16:30) were extracted and represented as point data

Type of street (%)
  a. Main road
  b. Residential
  c. Minor Road
  d. Footpath
	Percentage of route with this road type. Using OSM road classification	The OSM road types were aggregated into four road classifications that have been used in the literature [238]: main road, minor road, residential road and footpath

Woodland	Percentage of woodland along route (% within 25m)	Woodland polygon data were downloaded from OS Meridian 2 [408]
Exposure to RFE	Count of unhealthy food outlets within 100m of route. Chapter 4 documents how the unhealthy food outlet data were collated and prepared	Postcode level food outlet point data were downloaded from the Food Standards Agency [346]. A typology of opening times is documented in Appendix 8. This was so to account for outlets not being open all day.

8.2.3 Conditional Logistic Regression

A backward stepwise method was used to fit the regression model. The independent variables included in the model are the environmental characteristics listed in Table 8.1. Independent variables that were not useful (i.e. not statistically significant) in discriminating between the GPS routes and the SNR were not included in the final model (greenspace, accidents, woodland). This reduced the set of independent variables that were included in the final model [410]. This is an advantage of using a stepwise approach when there are a large number of independent variables to include.

Two separate conditional logistic regression models were fitted. The first model was fitted to discriminate between the environmental characteristics along the GPS routes to school and the SNR **to school**. The second model was run to discriminate between the environmental characteristics along the GPS routes home from school and the SNR home **from school**.

In logistic regression, the probability of Y occurring is predicted given known values of X. When there are several predictors, the equation from which the probability of Y is predicted is given by:

$$P(Y) = \frac{1}{1 + e^{-(b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots b_n X_{ni})}}$$

In which P(Y) is the probability of Y occurring, e is the base of natural logarithms, b_0 is the constant, X_n is the predictor variables, and b_n is a coefficient attached to the relevant predictor.

Table 8.2, Table 8.3 and Table 8.4 show the descriptive statistical summaries of the variables included in the initial model.

The fit of the model was assessed by using "goodness of fit" criteria; the Akaike information criterion (AIC) and the Bayes information criterion (BIC). The contribution of each independent variable was assessed by calculating the z-statistic.

*Table 8.2 Descriptive statistics of variables included in initial logistic regression model for routes **to school**. Greenspace, Accidents and Woodland were not included in the final model.*

Variables	Mean	SD	Median	Min	Max	Skew	SE
Length (m)	1552.95	876.43	1407.18	66.29	5182.31	0.79	28.45
Greenspace (%)	18.94	20.12	14.21	0.00	92.29	1.21	0.65
Bluespace (%)	1.79	5.12	0.00	0.00	42.61	4.44	0.17
Traffic Light Count (n)	0.47	1.25	0.00	0.00	12.00	4.04	0.04
Accidents (n)	2.36	3.39	1.00	0.00	21.00	2.04	0.11
Main road (%)	26.95	26.33	19.54	0.00	100.00	0.72	0.85
Residential road (%)	54.64	25.33	51.86	0.00	100.00	0.11	0.82
Footpath (%)	27.96	22.13	22.84	0.00	93.78	0.83	0.72
Minor road (%)	11.06	15.77	5.41	0.00	97.50	2.32	0.51
Woodland (%)	0.00	0.00	0.00	0.00	0.00	NA	0.00
Exposure (n)	5.94	9.44	3.00	0.00	70.00	3.00	0.31

Table 8.3 Descriptive statistics of variables included in initial logistic regression model for routes home from school. Greenspace, Accidents and Woodland were not included in the final model.

Variables	Mean	SD	Median	Min	Max	Skew	SE
Length (m)	1512.42	913.78	1371.93	103.25	7172.21	1.15	29.25
Greenspace (%)	75.46	18.03	78.29	10.96	100.00	-0.68	0.58
Bluespace (%)	1.85	4.97	0.00	0.00	39.36	4.11	0.16
Traffic Light Count (n)	0.40	1.16	0.00	0.00	13.00	4.78	0.04
Accidents (n)	2.08	3.36	1.00	0.00	24.00	2.70	0.11
Main road (%)	26.46	25.81	19.51	0.00	100.00	0.76	0.83
Residential road (%)	53.44	25.77	51.24	0.00	100.00	0.21	0.82
Footpath (%)	29.22	24.09	22.79	0.00	98.02	0.77	0.77
Minor road (%)	11.03	15.91	5.70	0.00	97.57	2.46	0.51
Woodland (%)	0.00	0.00	0.00	0.00	0.00	NA	0.00
Exposure (n)	5.02	8.38	2.00	0.00	63.00	3.30	0.27

Table 8.4 Descriptive statistics of variables included in initial logistic regression model for shortest network routes. Greenspace, Accidents and Woodland were not included in the final model.

Variables	Mean	SD	Median	Min	Max	Skew	SE
Length (m)	4061.07	3411.75	2932.50	107.26	27343.17	1.80	108.43
Greenspace (%)	80.89	15.38	83.61	11.09	100.00	-0.98	0.49
Bluespace (%)	2.41	4.41	0.00	0.00	34.77	3.21	0.14
Traffic Light Count (n)	2.36	4.61	0.00	0.00	39.00	3.20	0.15
Accidents (n)	5.29	7.38	2.00	0.00	58.00	2.52	0.23
Main road (%)	38.81	28.94	36.62	0.00	100.00	0.25	0.92
Residential road (%)	28.51	27.01	20.40	0.00	100.00	1.02	0.86
Footpath (%)	15.53	14.74	11.41	0.00	77.05	1.13	0.47
Minor road (%)	13.26	18.71	5.63	0.00	100.00	1.71	0.59
Woodland (%)	0.11	0.86	0.00	0.00	11.97	9.04	0.03
Exposure (n)	10.75	19.82	3.00	0.00	174.00	3.45	0.63

8.2.4 Weighted Network

The results of the logistic regression models were used to assign costs (also known in the literature as impedances) (Figure 8.4) to the road network. Line geometries (vertices) that made up the road network were labelled based on the environmental characteristics they contained. A column was generated to label each vertex with a road type and a column was generated so that each vertex could be flagged if traffic signals were positioned within 25m of the vertex. This process was repeated for every environmental characteristic. The vertices were then allocated impedance values. The results of the conditional logistic regression were used to inform the impedance values. Specifically, environmental characteristics that were greater on the GPS routes compared with SNR (e.g. traffic lights) resulted in a lower cost on the network. Environmental characteristics that were fewer along GPS routes compared to SNR were given a higher cost on the network (e.g. food outlets).

I then took a manual, iterative approach to modify the impedance values along the network. Once the first impedance values were assigned, I compared the weighted network routes with the GPS routes. I did this by measuring the percentage of the route overlap and the correlation between the GIS modelled route exposures and the GPS modelled exposures (Spearman's Rank). I then made small, incremental changes to the impedance values on the network until the optimum overlap and correlation was reached.

8.2.5 Weighted Network Routes

Routes to school and routes home from school were generated in PostGIS using the Dijkstra algorithm [411] in pgRouting [376] using the cost weighted network. The routing function, pgRouting, is an open source tool that extends the PostGIS and PostgreSQL geospatial database to provide routing functionality. This routing software was chosen over other products (e.g. ArcGIS network analyst) because:

1) It is free to download and use;
2) Code can be made publicly available and reused;
3) Different routing algorithms can be selected.

The Dijkstra algorithm is a shortest path algorithm that was developed by Edsger Dijkstra in 1959 [412]. This routing algorithm was selected for use in pgRouting because it is the routing algorithm that is currently used in studies to model commuting routes in a GIS (child and adult routes). I wanted to adapt current practice by developing the underlying cost network rather than use the standard routing algorithm which is known to not adequately capture child route choice.

The routes to and from school were then used to calculate the associated exposures to the RFE along the commute to and from school. The exposure to the RFE was calculated in PostGIS and defined as the number of outlets along the route within 100m of the route. This is common practice for defining access to the RFE along a commute [11, 56, 113, 268]. Routes and exposure to the RFE generated through this method will be referred to in the rest of the text as WNR for weighted network routes and WNE for weighted network exposures.

8.2.6 Modelled Routes Validation

8.2.6.1 Route Overlap

The WNR were compared with the GPS data to see the similarity in route shape. The intersection of the WNR within 50m of the corresponding GPS route was calculated in PostGIS. A buffer of 50m was used because this distance has been applied elsewhere when comparing GIS generated children's walking routes with GPS data [11]. Descriptive statistics of the differences between the GIS WNR and the GPS routes were calculated in R (version 3.3.3).

8.2.6.2 Multilevel Regression Model

A multilevel regression model was fitted to assess the association between the GPS exposures and WNE. A multilevel model was used because the route data was hierarchical; i.e. the model assumed that the cases were not independent of one another. This was appropriate as the data had two different levels; route exposures for individuals, and individuals were clustered within 15 schools.

Two random effects regression models were run. One regression model was fitted for the GPS exposures and WNE for routes to school. A second regression model was fitted for the GPS exposures and WNE for routes home from school. Table 8.5 shows the describtive statistics of the variables included in the models; GPS exposures and WNE.

The linear model is represented as:

$$Y_{ij} = (b_0 + u_{0j}) + b_1 X_{ij} + \varepsilon_{ij}$$

Where Y is the outcome, X is the predictor, b_0 is the intercept, u_{0j} is the random intercept, ε is error.

Table 8.5 Descriptive statistics of variables included in the two multilevel regression models. GPS am refers to the exposures along the GPS routes to school and WNE am refers to the modelled exposure along the modelled route to school. GPS pm refers to the exposure along the GPS routes home from school and WNE pm refers to the modelled exposure along the modelled route home from school.

Exposure	Mean	SD	Median	Min	Max	Range	Skew	SE
GPS am	5.94	9.44	3	0	70	70	3	0.31
WNE am	7.5	12.2	3	0	116	116	3.84	0.39
GPS pm	5.02	8.38	2	0	63	63	3.3	0.27
WNR pm	9.67	18.24	3	0	149	149	3.83	0.58

8.3 Results

8.3.1 Environmental Characteristics

The GPS data yielded 949 walking routes to school recorded by 377 individual children and 976 walking routes home, recorded by 409 children. Overall, 884

individuals provided GPS data on their walking route to school *or* walking routes home. There were 333 children who provided GPS data for both their walking route to school *and* walking route home. Table 8.6 and Table 8.7 summarise the environmental characteristics of the GPS walking routes recorded in the GPS dataset.

The mean distance recorded by the GPS data for walking routes to school was 1553m. The shortest route was recorded at 66m and the maximum distance travelled walking to school was 5182m. There were 10 routes longer than 4000m in length. The proportion of the routes that traversed main roads, minor roads and footpaths were skewed distributions (Figure 8.2 and Figure 8.3). The median values of the routes to school were, 20% of the route was along main roads, 5% along residential roads and 23% along footpaths. On average, 52% of the route to school was along a residential road. The median exposure to the RFE on the walk to school was 3 outlets. Some individuals' routes were entirely along main roads or residential roads.

The mean distance travelled on the route home was 1512m. The minimum distance walked on the route home was 103m and the maximum distance walked was 7172m. Five routes home were recorded as greater than 5000m. The median percentage of road type traversed did not differ greatly from morning routes for main roads, minor roads and footpaths (20%, 6% and 23% respectively). The percentage of the route along footpaths was greater for routes home from school compared with routes to school (Figure 8.3). The median exposure to the RFE on the walk home from school was 2 outlets. The range of route length for routes home was greater than routes to school by 1900m.

*Table 8.6. Summarises the environmental characteristics along the walking **routes to school** recorded in the PEAR dataset. Green*
percentage of the route that is within 50m of Green and Bluespaces. Main Road, Residential Road, Footpath and Minor Road descri
these road types. Traffic lights, Pedestrian Crossing, Accidents and Exposure are counts of these features. Orange highlights value

	Mean	Median	Min	Max	Range	Skew
Distance (m)	**1553**	1407	66	**5182**	**5116**	**0.79**
Greenspace (%)	19	14	0	92	92	**1.21**
Bluespace (%)	2	0	0	43	43	4.44
Traffic Lights (n)	0	0	0	12	12	4.04
Pedestrian Crossing (n)	1	0	0	9	9	2.69
Accidents (n)	2	1	0	21	21	2.04
Main Road (%)	27	20	0	100	100	0.72
Residential Road (%)	**55**	52	0	100	100	0.11
Footpath (%)	28	23	0	94	94	**0.83**
Minor Road (%)	11	5	0	97	97	2.32
Exposure (n)	6	**3**	0	70	70	3.00

*Table 8.7. Summarises the environmental characteristics along the walking **routes home from school** recorded in the PEAR da*

describe the percentage of the route that is within 50m of Green and Bluespaces. Main Road, Residential Road, Footpath and Min

that traverses these road types. Traffic lights, Pedestrian Crossing, Accidents and Exposure are counts of these features. Orange h

	Mean	Median	Min	Max	Range	Skew
Distance (m)	**1512**	1372	103	**7172**	**7069**	**1.15**
Greenspace (%)	21	17	0	91	91	0.91
Bluespace (%)	2	0	0	39	39	4.11
Traffic Lights (n)	0	0	0	13	13	4.78
Pedestrian Crossing (n)	1	0	0	10	10	2.66
Accidents (n)	2	1	0	24	24	2.70
Main Road (%)	26	20	0	100	100	0.76
Residential Road (%)	**53**	51	0	100	100	0.21
Footpath (%)	29	23	0	98	98	**0.77**
Minor Road (%)	11	6	0	98	98	2.46
Exposure (n)	5	**2**	0	63	63	3.30
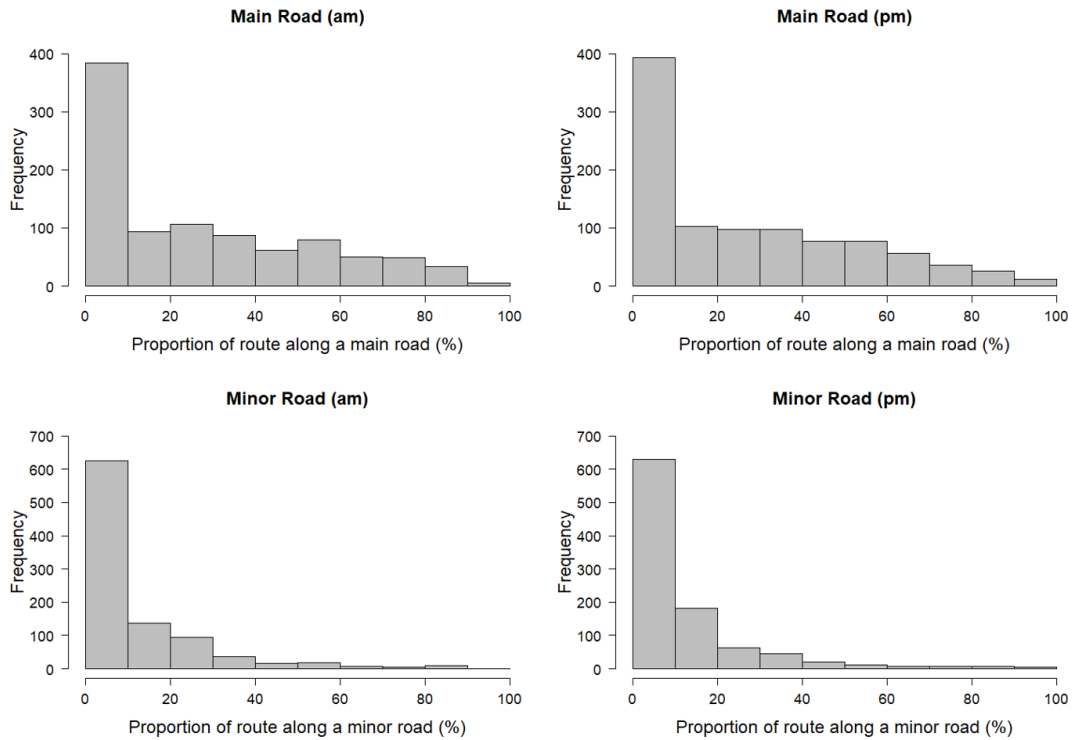
Figure 8.2Proportion of walking route along main roads and minor roads. On the walk to school (am) and the walk home from school (pm)
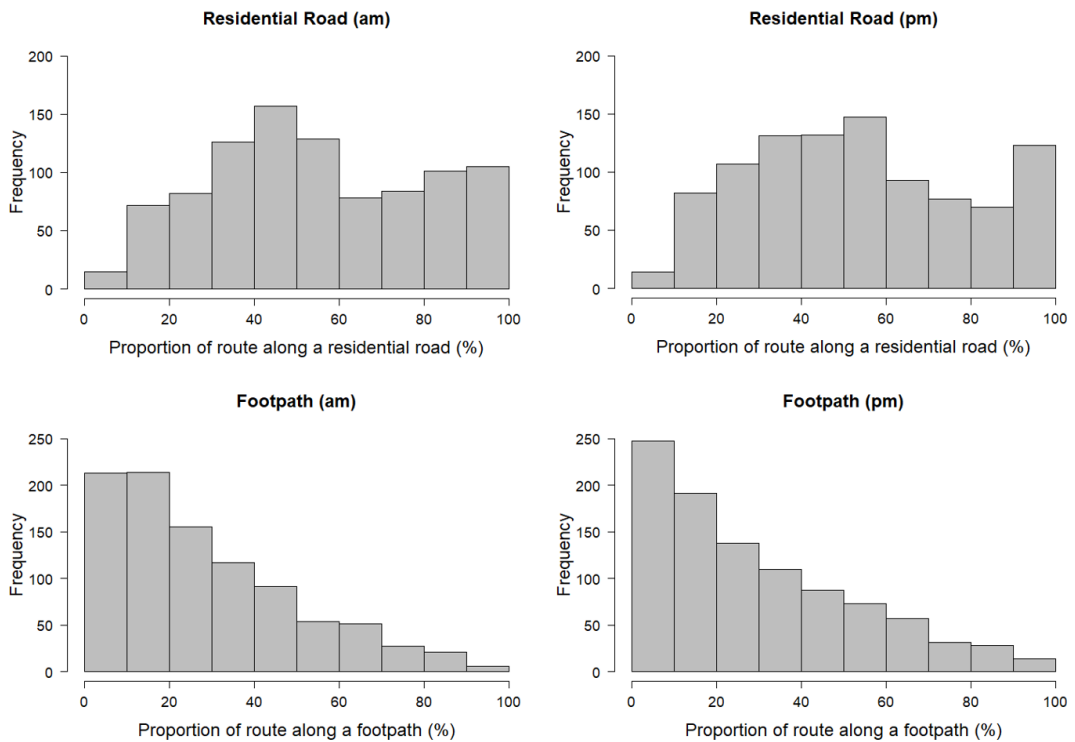
Figure 8.3 Proportion of walking route along residential roads and footpaths. On the walk to school (am) and the walk home from school (pm)

8.3.2 Conditional Logistic Regression

Regression modelling was completed to discriminate between the GPS routes and the GIS modelled SNR. Table 8.8 shows the results of the conditional logistic regression for walking routes to school. Table 8.9 shows the results of the conditional logistic regression for walking routes home from school. The conditional logistic regression models followed a backward selection process. Only significant predictor variables were included in the final model.

8.3.2.1 Routes to school

The results showed that for routes to school, there were significantly more blue spaces and traffic lights along the GPS routes. The Odds Ratio (OR) for main roads, residential, footpaths and minor roads showed that GPS routes contained a significantly smaller percentage of the route along these road types. The logistic regression results also showed that GPS routes had significantly less exposure than the SNR for the walk **to school** (OR 0.918, 95% CI 0.834,0.976) and the walk home **from school** (OR 0.901, 95% CI 0.842,0.964).

8.3.2.2. Routes home from school

Similarly, there were significantly more blue spaces and traffic lights along the GPS routes than the SNR home from school. The percentage of the routes along main roads, residential, footpaths and minor roads was significantly less for GPS routes compared with the SNR. GPS walking routes home had significantly smaller exposures than the SNR (OR 0.918).

Table 8.8. Conditional Logistic Regression results for environmental characteristics along walking routes to school

	B (SE)	p	Odds Ratio	95 % CI for OR	
				Lower	Upper
Length	0.002 (0.001)	< 0.01	1.002	1.002	1.003
Blue space percentage	0.141 (0.037)	< 0.01	1.151	1.072	1.237
Traffic Light Count	0.377 (0.140)	0.07	1.458	1.108	1.920
Main Road	-0.098 (0.012)	< 0.01	0.907	0.885	0.929
Residential Road	-0.127 (0.012)	< 0.01	0.881	0.861	0.901
Footpath	-0.127 (0.013)	< 0.01	0.871	0.849	0.894
Minor Road	-0.099 (0.016)	< 0.01	0.905	0.878	0.934
Exposure	-0.085 (0.022)	0.1	0.918	0.834	0.976

Table 8.9. Conditional Logistic Regression results for environmental characteristics along walking routes home from school

	B (SE)	p	Odds Ratio	95 % CI for OR Lower	Upper
Length	0.001 (0.001)	< 0.01	1.001	1.000	1.002
Blue space percentage	0.089 (0.033)	< 0.05	1.093	1.025	1.165
Traffic Light Count	0.572 (0.148)	< 0.05	1.772	1.326	2.370
Main Road	-0.087 (0.011)	< 0.05	0.917	0.897	0.936
Residential Road	-0.116 (0.011)	< 0.05	0.890	0.871	0.909
Footpath	-0.122 (0.011)	< 0.05	0.885	0.866	0.905
Minor Road	-0.092 (0.017)	< 0.05	0.912	0.883	0.942
Exposure	-0.104 (0.034)	< 0.05	0.901	0.842	0.964

8.3.3 Weighted Network

Table 8.10 reports the final impedance values that were assigned to the road network. Impedance values were informed by the OR of the conditional logistic regression and the final values were derived from the iterative approach used to find the optimum network weights described in (8.2.4). Each vertex was assigned a value based on the road type. Network vertices that had traffic lights or outlets along them were reassigned the impedance value set for traffic lights and outlets, as road type impedance values were assigned first. For example, a vertex representing a main road that contained traffic lights was reassigned a value of 0.5. If the vertex was within 25m of an outlet, the vertex was reassigned a value of 0.8. This was an iterative process. Figure 8.4 visualises this process.

Table 8.10. The impedance that was assigned to each vertices

	Environmental characteristics					
	Main Road	Residential Road	Footpath	Minor Road	Traffic Lights	Outlet
Routes to school	0.8	0.5	0.8	1.5	0.5	0.8
Routes home from school	0.8	0.4	0.8	1.2	0.7	0.5

Figure 8.4. Example of how costs were assigned to each edge in the road network. Each colour represents a different cost

8.3.4 Modelled Routes Validation

8.3.4.1 Overlap of GPS and GIS

For routes to school, the mean intersection between the GIS generated routes and the GPS routes was larger for the WNR than for the SNR. On average, 53.49% of WNR to school and 18.24% of shortest network routes took the same path as the GPS routes (Table 8.11). For the GIS generated WNRs to school, 26% intersected with 100% of the corresponding GPS routes. 56% of the GIS generated WNRs home from school intersected 100% with the corresponding GPS routes (Table 8.12). Compared with 1.6% of shortest network routes to school and 1.84% of shortest network routes that intersected 100% with the corresponding GPS routes home from school. Table 8.13 and Table 8.14 shows that for routes to school and routes home, the mean percentage of intersection decreases as the route length increases. Based on the GPS recorded length, for routes home from school less than 3km long, 61% of WNR had an intersection with GPS routes of 100%. In contrast, 28% of WNR routes intersected

144

with 100% with the GPS routes to school less than 3km in length. Intersection of 100% occurred in the shortest network routes that were up to 1km long. For the SNR, 7% of routes to school intersected 100% with GPS routes shorter than 1km; compared with 43% of the WNR to school intersecting 100% with GPS routes shorter than 1km. One GPS journey to school was recorded as greater than 5km in length. The route intersection of the WNR sharply decreased for route lengths greater than 4km. However, there are only 10 routes in the dataset that are greater than 4km in length. The distribution of the WNR intersection was slightly skewed to the left (skew = 0.02) (Figure 8.5). The negative kurtosis value of -1.65 suggests that the data has a more platykurtic distribution and the data is therefore distributed across the tails and the peaks (Figure 8.5). The SNR intersections were positively skewed. Figure 8.5 shows that the most frequent rate of intersection is less than or equal to 10%.

For routes home from school the mean intersection between the GPS routes and GIS generated routes was 91% for the WNR and 18.40% for SNR. The smallest percentage of intersection for the WNR home from school was 10% for a route between 2km and 2.5km long. The negatively skewed distributions of intersections for the WNR suggest that higher intersection rates were more frequent (Figure 8.5). The positively skewed distribution for the shortest network intersections highlight that the intersection rates were skewed to the left and poorer rates of intersection were more frequent.

*Table 8.11. Summary of SNR and WNR overlap with GPS routes for walking **routes to school***

	Mean	SD	Median	Min	Max	Range
GPS route length (m)	1553	876	1407	66	5182	5116
SNR length (m)	3143	1844	2785	163	14027	13864
SNR intersect distance (m)	183	210	112	0	1651	1651
SNR intersect with GPS (%)	18	23	8	0	100	100
WNR length (m)	1679	826	1644	151	4403	4252
WNR intersect with GPS distance (m)	773	689	566	9	3359	3350
WNR intersect with GPS (%)	53	39	44	0	100	100

Table 8.12. Summary of SNR and WNR overlap with GPS routes for walking routes
home from school

	Mean	SD	Median	Min	Max	Range
GPS route length (m)	1512	914	1372	103	7172	7069
SNR length (m)	3144	1869	2801	295	14027	13732
SNR intersect distance (m)	178	208	103	0	1651	1651
SNR intersect with GPS (%)	18	23	8	0	100	100
WNR length (m)	1707	869	1580	151	6915	6764
WNR intersect with GPS distance (m)	1457	739	1400	113	4077	3964
WNR intersect with GPS (%)	91	17	100	10	100	90
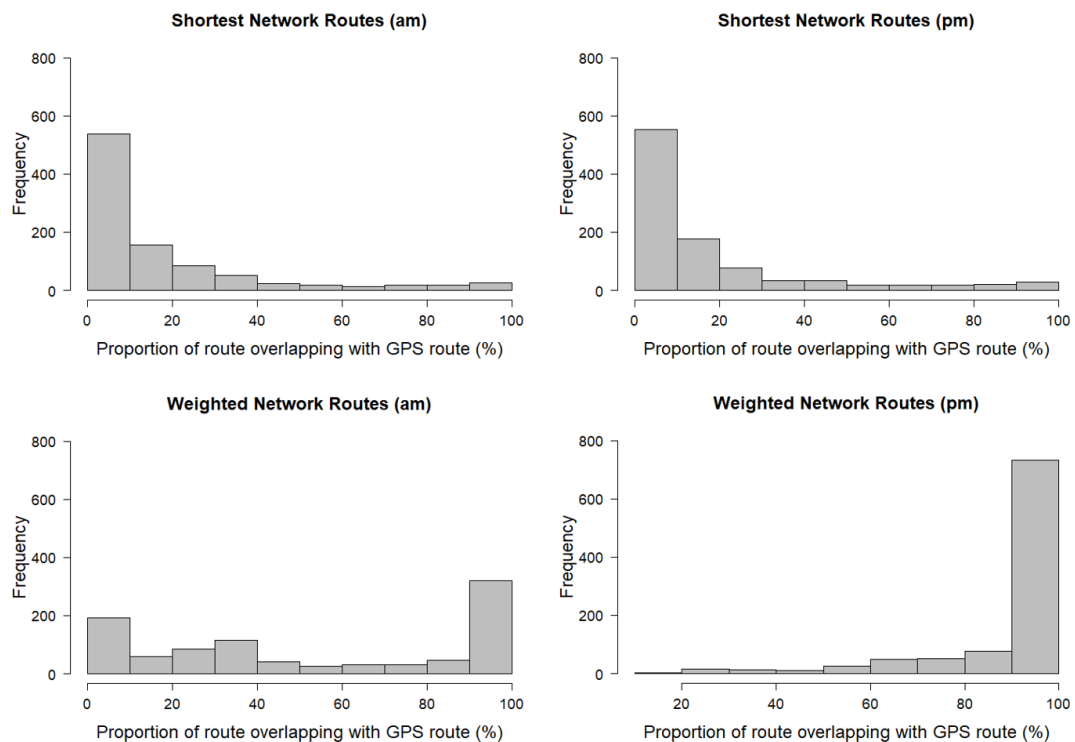
Figure 8.5. SNR and WNR overlap with GPS route for routes to school (am) and routes home from school (pm)

146

*Table 8.13. Route overlap (%) of SNR and WNR with GPS routes **to school**. Displayed as 500m intervals of GPS route length (m).*

	route length (m)	mean	median	min	max	range
SNR intersect with GPS (%)	<500	54	40	0	100	173
WNR intersect with GPS (%)		77	100	2	100	98
SNR intersect with GPS (%)	> 500 <= 1000	27	14	0	101	101
WNR intersect with GPS (%)		59	77	1	100	99
SNR intersect with GPS (%)	> 1000 <= 1500	14	8	1	52	51
WNR intersect with GPS (%)		59	63	1	100	99
SNR intersect with GPS (%)	> 1500 <= 2000	13	8	0	91	91
WNR intersect with GPS (%)		44	32	0	100	100
SNR intersect with GPS (%)	> 2000 <= 2500	6	4	0	46	45
WNR intersect with GPS (%)		46	35	1	100	99
SNR intersect with GPS (%)	> 2500 <= 3000	7	4	0	56	55
WNR intersect with GPS (%)		48	33	1	100	99
SNR intersect with GPS (%)	> 3000 <= 3500	10	4	0	55	54
WNR intersect with GPS (%)		34	28	1	99	98
SNR intersect with GPS (%)	> 3500 <= 4000	2	1	0	6	6
WNR intersect with GPS (%)		23	4	1	93	92
SNR intersect with GPS (%)	> 4000 <= 4500	2	2	2	2	0
WNR intersect with GPS (%)		23	23	20	27	8
SNR intersect with GPS (%)	> 4500 <= 5000	6	6	6	6	0
WNR intersect with GPS (%)		20	20	20	20	0
SNR intersect with GPS (%)	> 5000 <= 5500	3	3	3	3	0
WNR intersect with GPS (%)		7	7	7	7	0

*Table 8.14. Route overlap (%) of SNR and WNR with GPS routes **home from school**. Displayed as 500m intervals of GPS route length (m).*

	route length (m)	mean	median	min	max	range
SNR intersect with GPS (%)	<500	52	30	0	179	179
WNR intersect with GPS (%)		94	100	42	100	58
SNR intersect with GPS (%)	> 500 <= 1000	26	13	0	108	108
WNR intersect with GPS (%)		96	100	26	100	74
SNR intersect with GPS (%)	> 1000 <= 1500	12	7	1	51	50
WNR intersect with GPS (%)		94	100	11	100	89
SNR intersect with GPS (%)	> 1500 <= 2000	11	7	0	82	82
WNR intersect with GPS (%)		89	97	21	100	79
SNR intersect with GPS (%)	> 2000 <= 2500	9	5	0	77	76
WNR intersect with GPS (%)		88	97	10	100	90
SNR intersect with GPS (%)	> 2500 <= 3000	8	3	1	56	56
WNR intersect with GPS (%)		81	86	52	100	48
SNR intersect with GPS (%)	> 3000 <= 3500	6	2	0	55	55
WNR intersect with GPS (%)		84	92	31	100	69
SNR intersect with GPS (%)	> 3500 <= 4000	2	2	0	4	4
WNR intersect with GPS (%)		87	91	64	100	36
SNR intersect with GPS (%)	> 4000 <= 4500	2	2	0	7	7
WNR intersect with GPS (%)		69	69	22	94	73
SNR intersect with GPS (%)	> 4500 <= 5000	2	2	2	2	0
WNR intersect with GPS (%)		65	65	64	65	1
SNR intersect with GPS (%)	> 5000 <= 5500	1	1	1	1	0
WNR intersect with GPS (%)		61	61	61	61	0
SNR intersect with GPS (%)	> 5500 <= 6000	0	0	0	0	0
WNR intersect with GPS (%)		44	44	44	44	0
No GPS routes > 6000m < 7000m						
SNR intersect with GPS (%)	> 7000 <= 7500	0	0	0	0	0
WNR intersect with GPS (%)		34	34	34	34	0

8.3.4.2 Multilevel Regression Model

A multilevel random effects model was fitted between exposure scores along the GPS route and exposure scores calculated along the weighted network. Routes were nested within individuals who were nested within schools.

The WNR exposures, both along routes to school and along routes home, were significantly associated with the exposures calculated from the GPS routes ($p< 0.001$). The regression coefficients and standard error (SE) are shown in Table 8.15.

Table 8.15. Final multilevel regression model

Fixed Effect	Value	SE	t-value	p-value
Route to school				
Intercept	0.81	1.17	0.48	0.63
WNR Exposure	1.42	0.07	19.43	< 0.001
Route home from school				
Intercept	0.66	0.34	1.95	0.05
WNR Exposure	1.15	0.04	25.65	< 0.001

8.4 Discussion

8.4.1 Main Findings

Pupils aged 13-14 years who walk to and from school tend to travel less than 4800m. Current government policy states that Local Authorities are obliged to provide free travel arrangements for pupils aged 8-16 years who live more than 4800m from their nearest school [413]. Therefore, this shows that the GPS data is reliable because we would not expect children in the UK to generally walk further than 4800 m to school when they are eligible for free travel.

The quieter residential streets were the most frequently traversed road types for routes to school and home from school (on average, 55% of the route to school and 53% of

the route home was along a residential street) and this is consistent with the literature [238]. Whereas the most traversed road type for the shortest network route was main roads (on average, 39% of route was along main roads). The shortest network route will often route individuals along main roads because this is usually the most direct route. However, based on the results of this chapter and previous studies [11, 238], we can see that children often choose a short route but not necessarily the most direct or absolute shortest route. This is why I chose to adapt the shortest network method, by applying impedances along a network, to more realistically emulate route choice. In the literature, assigning costings to a network is popular in transport planning. However, to my knowledge, this approach has not previously been used to model children's walking routes and therefore this is an important addition to the literature.

There is a plethora of literature of factors and behaviours associated with active travel [348, 402, 414, 415] but only a small number of studies have investigated what determines route choice when children actively travel to school. Environmental characteristics associated with route choice during active travel to school were investigated in Holland for children aged 8-12 years [238]. My current study has contributed to the literature by clearly demonstrating the environmental characteristics that influenced route choice for children in the UK aged 13-14 years. The most influential environmental factors associated with child walking routes to and from school were length, bluespace, traffic light count, exposure and the proportion of the journey that was made up of a particular road type (main road, residential road, footpath and minor road). Detailed environmental exposures have not been reported previously for child walking commutes in the UK.

There were significantly more traffic lights and blue space along the GPS routes than the shortest network routes. For routes to school, the traffic light count had a smaller odds ratio and narrower confidence interval compared with routes home from school; OR = 1.4 (95% CI 1.11-1.92) and 1.8 (1.33-2.37), respectively. This may be because traffic lights are placed on main roads at busy junctions. On the route to school there is a time constraint. Children must arrive at school by a certain time, therefore they choose to take a more direct route. Usually a more direct route is via main roads. Furthermore, traffic lights are usually placed for pedestrians to safely cross the road and so children may choose to travel via traffic lights when they are traversing main roads in order to safely cross. Previous research suggested children may avoid wooded

areas because they are perceived to be unsafe [416, 417]. This study did not find woodland to be significantly negatively associated with route choice. Woodland was not significantly associated with route choice, either to or from school. However, the research that reported children may avoid wooded areas was actually based on interviews with adults. Woodland not being significantly different between GPS routes and shortest network routes does not support the findings in the literature. This suggests that woodland is not generally an important factor for children deciding which route they chose to take.

Greenspace was not included in the final conditional logistic regression model because it masked interactions of the other characteristics. This may be because the difference between the values of greenspace in the GPS data was vastly different to the greenspace values for the shortest network routes. The mean greenspace percentage for the GPS routes to school and GPS routes home were 19% and 21% respectively. The mean greenspace percentage was 75% for the shortest network routes to school and home from school. For all other characteristics, the mean values did not vary to this degree between the GPS and shortest network routes. This disrupted the model and greenspace was removed from the analysis.

This study has reported that exposure to the RFE was significantly less along the GPS routes when compared with the shortest network routes. This supports the findings by Harrison et al. [11] that shortest network routes overexpose children to the RFE. Dessing et al. [238] and Harrison et al. [11] both found that the shortest network distance was a suitable proxy for the distance that a child travels. The results of this chapter, support this because the OR for length was very close to 1 for routes to school and routes home from school; OR = 1.001 (95% CI 1.002 – 1.003), OR = 1.001 (95% CI 1.000 – 1.002, respectively). This suggests that children traverse a short route on their commute to and from school but their route choice is not necessarily determined by choosing the absolute shortest route.

Improving the shortest route calculation by using a weighted network to generate walking routes and the associated exposures, produced exposures that were significantly associated with exposures calculated from GPS data. Both route to school exposures and route home exposures were significantly associated with exposures calculated from GPS data. It is an important finding that the modelled exposures

showed agreement with the GPS exposures. This has not been reported before and supports the notion that modelled population level data could be used instead of small samples of people who have consented to wear devices. GPS data is often used as the 'gold standard' for representing 'real-life' behaviour. However, it has been suggested that wearing GPS devices may influence the behaviour of the participant [279, 280]. GPS data is a valuable resource for researchers but it is also important to develop reliable GIS methods that can be used for population scale research because it is not possible to collect GPS data on a population scale.

Population level analysis is essential for developing successful interventions that could drive a large scale reduction in the prevalence of obesity [8]. Methods such as data linkage, provide a population-level resource and a cost-effective alternative to cohort studies [295, 296] to undertake this essential research. Linking together health, administrative and environment data allows researchers to answer important societal questions that may otherwise not be evaluated [330]. GIS generated exposures to alcohol outlets have previously been modelled and linked with health data for population level analysis [151, 182]. The GIS model reported in this chapter has been used to calculate routes and associated exposure to the RFE for children in south Wales. This is reported and discussed in the following chapter.

8.4.2 Strengths

This investigation evaluated a new way to model exposure to the RFE along walking routes to and from school. The results of this chapter demonstrated that GIS can be used to model exposures to unhealthy food outlets along children's walking routes to school and home that are significantly associated with environmental exposures calculated from GPS route data. In practical terms, for every outlet that a child was recorded walking past in the GPS data, the model predicts a child walking past 1.41 outlets on their route to school and 1.15 outlets on their walk home from school. The GPS routes had a median exposure of 3 and 2 respectively. The GIS weighted network model predicts that the median exposure is 2 for the route to school and the route home from school.

This is an important result because modelled exposures along children's walking routes to and from school can be used to build a more complete picture of a child's

daily exposure to the RFE. Exposures to the RFE along the walk to and from school have been combined with home and school exposure to investigate the association between daily exposure to the RFE and BMI in teenagers aged 11-13 years in south Wales. These findings are reported in the following chapter.

In this study population level exposures were generated using GIS. This is a powerful tool, particularly in public health research, where the results of population level research contribute to the evidence base that policymakers draw from.

This methodology has the potential to be applied across a range of contexts. For example, environmental exposures to unhealthy food along commutes to school in other modes of transport and also calculating exposures to other environmental measures such as pollution and green space.

8.4.3 Limitations

Like any model that seeks to emulate the real world, there are limitations of and assumptions made by the GIS model. The model predicts routes that are directly from home to school and school to home. The GPS data that was used to derive the impedance values for the road network only contained walking routes to school that began at the home location and ended at the school location; and walking routes home that began at school and ended at the home location. It is likely that there will be children who will not walk the absolute same route to school and route home every day. For example, children that walk to school will often travel via a friend's home or they may not live in just one home, but two, and so have different start and end locations. This model calculates one exposure per child for their route to school and their route home. However, as this is a population level model, this is a reasonable assumption to make and the majority of route scenarios have likely been captured.

The GPS data contained a different number of routes for the routes to school and the routes home from school. These data contained multiple journeys for individuals and individuals were not consistent in providing the same number of routes for their journeys to school and their journeys home. For example, one participant may have provided two walking routes to school and one walking route home; whilst another participant provided three walking routes to school and no walking route home. Therefore, comparisons between the characteristics of the routes to school and routes

home from school should be made with caution. The purpose of producing the descriptive statistics of the environmental characteristics was for exploratory purposes. However, the conditional logistic regression and multilevel linear regression models take into account that individuals provided multiple routes and that individuals are nested within schools.

Investigating exposure to the RFE at home, school and along walking routes has provided an insight into children's daily exposure to the RFE. However, other locations or journeys that expose individuals to the RFE have not been included in the study. For example, children attending after school clubs at school, or other locations.

Accidents during commute times were included as a proxy for how safe a road may be perceived which may influence a child (or parent's decision) to walk along a given road. However, accidents at other times of the day may influence a child's decision, or parents decision to allow their child, to walk along a particular road.

8.4.4 Future work

Further work should explore the impedance values of the network and the routing algorithm that is used in the current model. This model uses the Dijkstra's algorithm and a weighted network to generate routes that represent walking routes to and from school. The Dijkstra algorithm was the only network routing algorithm explored in this investigation. There are however, other shortest network routing algorithms that have been developed for network analysis, such as A* algorithm K-shortest path. A suggestion for future work would be to explore whether the routing algorithm makes a significant difference to modelled exposures, and whether one of these algorithms may be more appropriate to use.

Based on the GPS and WNR intersections and the regression results, routes home were more closely associated with the GPS data than routes to school. The impedance values applied to the network were informed by the results of two logistic regression models (route to school and route home from school). Further investigation is needed to explore whether the impedance values assigned to the walking routes home from school are more appropriate than the impedances assigned for routes to school. The route intersection was far greater between the GPS data and the WNR than the intersection between the GPS data and the shortest network routes to school; and the exposures along the route to school generated from the weighted network were

significantly associated with the GPS exposures. However, I did not explore whether the impedance values applied to the network are the optimum impedance values for modelling children's walking routes to and from school. Future work could explore optimum network impedance values through the application of machine learning principles.

The greenspaces explored in this analysis were informed by previous literature. However, future research should investigate whether different types of greenspace are associated with route choice in children. Furthermore, it should be investigated whether green environments that influence route choice differ between children who live in urban areas and children who live in rural areas.

Purchasing data was not collected from the PEAR participants to see whether they bought food from the outlets in the RFE that they walked through. Research has found associations between higher exposures to fast food outlets and greater amounts of fast food being eaten [168]. Furthermore, a greater exposure to RFE will influence attitudes to food and the normative eating patterns; social factors that are known to influence obesity levels [418]. Therefore, future research should look to collect both purchasing data or food diaries and GPS data.

This model has been designed to predict exposures along walking routes to school and walking routes home from school. When children walk to school and/or home unaccompanied, they have more autonomy than if they are being driven in a car or travelling on a bus. Children are able to choose their route. Although a child's route will be influenced by their parent's thoughts and opinions; particularly where safety is concerned. At age 13-14 years, it is likely that they will have access to money to purchase food. Children who are driven to school will not have this freedom. When children are driven or accompanied by an adult, buying unhealthy food will often be determined by whether a parent or guardian is willing for the child to purchase unhealthy food or to purchase the food for them. It would be interesting to explore how to collect these behaviour data to model this behaviour accurately. It would also be interesting to compare children who are driven to those who generally walk to school. Covariates on parents would ideally be included in the future model, including parent education and work location.

It would be interesting to see how exposure to and engagement with the RFE is different from children that commute to school on foot or by car. For children that use public transport, their exposure to the RFE would be governed by the route that the mode of transport traverses. Public transport routes would need to be obtained.

8.4.5 Implications

This chapter has demonstrated that GIS can be used to generate population level exposures with known accuracy. Generating routes using a shortest network algorithm and a novel weighted street network has produced exposures to unhealthy food that are significantly associated with exposures calculated from GPS data. The results documented in this chapter support the concept that GIS routing models can reliably emulate real life behaviours at a population level. Population level research is an important aspect of the evidence base that policy makers use. Being able to generate children's exposure to the RFE along their routes to and from school will be a powerful tool for researchers and policy makers as they attempt to combat increasing obesity rates.

This chapter has presented a GIS model that can be developed and applied in different contexts. For example, the model could be developed to account for different travel modes such as motorised transport or cycling routes to school. This methodology could also be applied in measuring adult exposure along commutes to work and recreational journeys.

This model has been developed using GPS data from a previously funded UK study [419]. The GPS data was representative of a large geographic area that covered urban and rural areas. The model could therefore be applied, to school children around the UK. Children that live in rural areas tend to live further away from the school that they attend and are more likely to drive or use public transport to commute to school. For children who live within 4800 metres of the school that they attend, this model is reliable across urban and rural regions in the UK. However, it should be acknowledged that fewer children in rural regions actively travel to school. This supports the need and feasibility to develop exposure models for other methods of transport. Furthermore, the principle of the model stands to be applied elsewhere in the world: calculate environmental characteristics that are associated with route choice; use these

results to inform the impedance values on the network; produce routes from weighted network.

The exposures that have been generated by this model were subsequently linked with routinely collected health data and other environmental exposures to investigate whether a higher exposure to the RFE is associated with higher BMIs. This is documented in the next chapter.

8.5 Conclusions

The most influential environmental factors associated with child walking routes to and from school were length, traffic light count, exposure and the proportion of the journey that was made up of a particular road type (main road, residential road, footpath and minor road). These characteristics have been used to inform impedances along a network in order to emulate route choice for children walking to and from school. This chapter has developed a methodology that can produce predictions of children's exposure to the RFE along their walk to and from school with known accuracy. This is a novel methodology that provides great potential for developing the model to account for other modes of commuting to and from school or workplaces. This provides the potential to target public health interventions to the people who are most likely to achieve active travel to school and work [374].

Chapter 9

Using Data Linkage to Investigate the Association Between Exposure to the RFE and BMI

9.1 Introduction

This chapter reports the linking of household level environmental exposure data with individual level health data in the SAIL databank. For every home within 4.8km of 10 schools in south Wales, four exposures to the 'retail food environment' (RFE) have been calculated. These are:

- Home exposure to the RFE (also referred to in the literature as home neighbourhood exposure)
- School exposure to the RFE
- Exposure to the RFE along the walk **to** school (from here on, referred to as **exposure am**)
- Exposure to the RFE along the walk home **from** school (from here on, referred to as **exposure pm**)

These four environmental exposures were calculated in a GIS and then linked with individual level health data in the SAIL databank. The health data set was a cohort of individuals aged 11-13 years from south Wales and includes BMI data. A multilevel regression model has been fitted to investigate whether exposures to the RFE are positively associated with BMI for children aged 11-13. I have also investigated differences between children who have the potential walk to school and children who do not walk.

The process of linking datasets from different sources is known as data linkage. This practice brings together diverse information and allows researchers to build a detailed picture of a population without undertaking trials [420]. Data linkage is a population-level resource and a cost-effective alternative to cohort studies [295, 296]. This has led to health data linkage research becoming a rapidly growing research field and it is acknowledged as having the potential to revolutionise health research [293].

The SAIL Databank at Swansea University brings together billions of anonymised, person-based records in a secure environment, which allows trusted researchers to link together health, administrative and environment data [292, 328]. It contains routinely collected health and administrative datasets about the population of Wales [329]. The detailed data stored in SAIL allows researchers to answer important societal questions that may otherwise not be possible to answer [330].

To provide a safe haven for sensitive data, before data arrives in SAIL, identifiable details are anonymised by a trusted third party (TTP), NHS Wales Informatics Service (NWIS) [331]. The process of anonymisation is called the split-file process (Figure 9.1) [331, 332]. This process describes how identifiable details are removed from the dataset and replaced with an Anonymous Linking Field (ALF) or a Residential Anonymous Linking Field (RALF). An ALF is assigned if the data is individual level records or a RALF assigned if the data is address level data.

The health dataset that was used in this data linkage study is a cohort of 1147 children aged 13-14 from south Wales. No BMI data, nor any other overweight/obesity indicators are currently routinely collected for teenagers in the UK. The Child Measurement Programme (CMP) records the BMI of 4-5 year olds in England and Wales, and 10-11 year olds in England. However, the programme does not extend to secondary school aged children (11-16). Therefore, SAIL does not contain population wide coverage of BMI data. This chapter looks to promote the routinisation of collecting BMI data for teenagers and also highlight SAIL's capabilities of facilitating the linkage of environmental exposure data with sensitive health data.
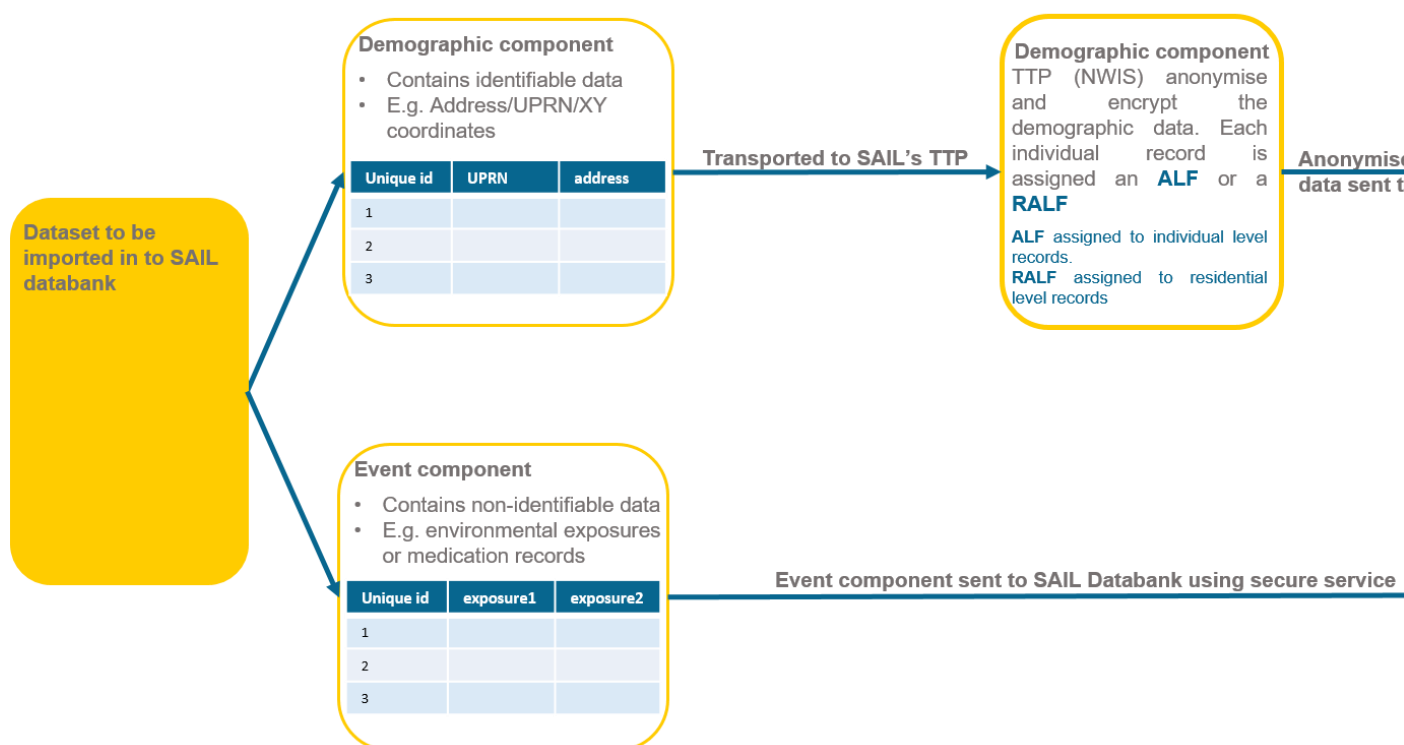
Figure 9.1. Summary of the anonymisation process for SAIL databank

9.2 Methodology

9.2.1 Health and Administrative Data

Two datasets held in the SAIL databank were used in this study; the Health Survey Wales (HSW) dataset and the Welsh Demographic Service (WDS) dataset.

9.2.1.1 Health Survey Wales

The HSW dataset was collected for a study that examined the factors associated with obesity in children aged 11-13 in the UK [421]. The data was collected between 2009-2010 by trained researchers. The dataset contains anthropometric measures, blood samples and fitness test scores for 1147 children across 10 schools in south Wales in the United Kingdom. The dataset contains 490 males and 657 females. One third of the individuals were overweight [421]; this is representative of the UK population for this age [422]. The study population was heterogeneous across socioeconomic status [421]. Five of the schools were defined as "affluent" and the other five as "deprived". The deprivation index was based on the percentage of pupils attending the school that were eligible for free school meals (FSM) $(< 21\% = $ affluent, $> 21\% = $ deprived). This index was calculated by the researchers that collected the HSW data. A detailed methodology of how this index was calculated is documented elsewhere [421].

9.2.1.2 Welsh Demographic Service

The WDS dataset contains individual level administrative data on the Welsh population from 1990 to the present day [423]. The dataset contains demographic data such as age, gender and date of death. It is used as a population register within the SAIL databank because it contains anonymised residences so movement of the population can be traced.

9.2.2 Environmental exposure data

9.2.2.1 Data Sources for Calculating Exposure to the RFE

Home & School locations

Home and school locations were obtained from Address Base Premium (ABP) [362]. ABP is Ordnance Survey's (OS) address database for the United Kingdom. It is a spatial database that, amongst other data, contains point data for current and historic buildings in the UK. The dataset also contains Unique Property Reference Numbers (UPRNs) which are anonymised by NWIS and used to generate RALFs. These are the residential address unique linking fields that allowed me to link the environmental exposure data with the health data.

Food Outlet Locations

Food outlet locations were obtained from Welsh Local Authorities. The collection and preparation of this data was documented in Chapters 3 & 4. Data for Carmarthenshire, Cardiff, Rhondda Cynon Taff and Swansea was used in this analysis. Outlets that did not contain XY coordinates or a UPRN were geocoded using Locator Hub software [424]. This accounted for less than 3% of outlets. Outlets that were open for trading during 2009 and 2010 were included in the exposure counts as this was the year that the health data was collected. A typology of opening times is documented in Appendix 9. This was so to account for outlets not being open all day.

Road network

The road data used was OpenStreetMap. The evaluation and preparation of this dataset are documented in Chapters 5 and Appendix 5.

Walking route to and from school

In order to calculate exposure am and exposure pm, walking routes to and from school had to be obtained. These routes were generated using a weighted network and the methodology is reported in Chapter 7. The routes to and from school were calculated from residential address to school locations and therefore the routes are unique to each individual.

9.2.2.2 Calculating Exposures to the RFE

Home, walking route to school, school and walking route home exposures to the RFE were calculated for each residential address. The environmental exposures were calculated in PostGIS [375] and they are defined in Table 9.1.

The environmental exposure data had to be retrospectively linked to the health data. As the health data had already been anonymised and stored in SAIL, it was not possible to know where participants lived and what school they attended. To overcome this, environmental exposures were generated for all residential addresses within 4.8km of the 10 schools involved in HSW.

Exposure to the RFE was calculated for all individuals that lived within 4.8km of the 10 schools involved in HSW. Pupils who lived in homes within 4.8km of each school were assumed to walk to the school that they attended. This was assumed because:

1. in the UK, children that live more than 4.8 km from their nearest school are entitled to free transport [413];
2. the characteristics of the GPS data collected in Chapter 7 suggest that children tend not to walk further than 4.8km to school. 99% of the walking routes were shorter than 4.8km.

Table 9.1. Definitions of GIS generated environmental exposures to the RFE and distances

Environmental exposures	Definition
Home Exposure to RFE	Number of unhealthy food outlets within 50m of an 800m network buffer of the home location
School Exposure to RFE	Number of unhealthy food outlets within 50m of a 400m network buffer of school location
Exposure to RFE along walk to school (exposure am)	Number of unhealthy food outlets within 100m of route to school. Only outlets open during 7:30 – 9:00 were included in the count
Exposure to RFE along walk home from school (exposure pm)	Number of unhealthy food outlets within 100m of route to school. Only outlets open during 14:40 – 16:30 were included in the count
Distance travelled to school	Distance of the walk to school in metres
Distance travelled home from school	Distance of the walk home from school in metres

9.2.3 Linking Exposure, Health and Administrative Data

9.2.3.1 Preparing the Health Survey Wales Dataset for Linkage

The HSW dataset was already anonymised and held in SAIL for this study. Individuals in the HSW datasets had a preassigned ALF as the primary key variable for record linkage.

9.2.3.2 Preparing the Welsh Demographic Service Dataset for Linkage

The WDS dataset was already anonymised and held in SAIL for this study. Individuals in the WDS datasets had a preassigned ALF as a primary key variable for record linkage. Individuals recorded in the WDS also have a RALF assigned to them.

Given the size of the WDS dataset, a member of the SAIL team subset the dataset on my behalf. This meant that for the WDS dataset, only individuals included in the HSW dataset were made available to me. At this stage, multiple entries for individuals remained. This was because if an individual had moved home (a RALF) then multiple distinct RALFs remained for an individual (an ALF).

9.2.3.3 Preparing the Environmental Exposure Dataset for Linkage

The exposure data was securely transferred into SAIL using the 'split-file' method. This means that the exposure data was split in to two parts. Part 1 contained identifiable data (UPRN and XY coordinates) which was sent to NWIS (the TTP) to anonymise the data. The identifiable UPRN and XY coordinates were replaced with a Residential Anonymised Linkage Field (RALF). Part 2 contained the non-identifiable environmental exposures. Both files were then combined by the SAIL team into a complete pseudonymised dataset (Figure 9.1).

9.2.3.4 Linking HSW, WDS and Environmental Exposure Datasets

The three datasets were re-encrypted before being made available in my workspace and ready for linkage. This meant that the ALF and RALF fields became ALF_E and RALF_E.

The households in which the HSW participants lived in at the time of data collection were flagged in the WDS dataset. The HSW and WDS data were then linked using the ALF_E field. The WDS data contained a RALF_E field and so the linked HSW/WDS dataset was then linked to the exposure dataset using the RALF_E field. This process is summarised in Figure 9.2. The corresponding environmental exposures were linked with the HSW/WDS dataset using the RALFs and a school lookup table.

Two cohorts were established within the linked dataset: "walkers" and "non-walkers". Individuals that were linked with the exposure data were assumed to walk to school as they lived within 4.8km of the school they attended. This was deemed a reasonable assumption because of the aforementioned policy in place [413]. Concurrently, the characteristics of the GPS data collected in Chapter 7 strongly support this assumption as 99% of the walking routes were shorter than 5km. This group was named "walkers". The unlinked individuals were assumed to not actively travel to school because they lived too far away and were classified as "non-walkers".

Environmental exposures dataset containing:
- Address/UPRN/XY coordinates (identifiable data)
- Environmental exposures (non-identifiable data)

Part 1

Anonymised by trusted third party. Point location replaced with RALF

Part 2

Data sent directly to SAIL

SAIL DATABANK

Anonymised environmental exposures dataset

Anonymised linking field: RALF

WDS dataset

Anonymous linking fields: ALF & RALF

HSW dataset

Anonymous linking field: ALF

Linked health dataset

Anonymised linking fields: ALF & RALFs

Figure 9.2. Summary of data linkage process. Split file process for anonymising the environmental data is described. Linking f

167

9.2.4 Research Ethics and Information Governance

This study involved the analysis of retrospective and routinely collected anonymised health data. The anonymisation process within the SAIL databank meant that consent did not have to be given by each individual.

Trained analysts at SAIL prepared the HSW and WDS datasets for me by ensuring that no individuals could be identified through small numbers. In addition, an experienced information governance reviewer ensured that k-anonymisation standards were adhered to before the data was released in to the SAIL Gateway.

Before the work was undertaken, an application was submitted and approved by an Information Governance Review Panel (IGRP). The IGRP includes members from the British Medical Association (BMA) [425], the National Research Ethics Service (NRES) [426], the Public Health Wales NHS Trust and Informing Healthcare [427] as well as lay members. The IGRP approved that the proposed analysis met the strict information governance arrangements set out by SAIL.

9.2.5 Statistical Analysis

a) Environmental Characteristics of the "Walkers" Group.
Descriptive statistics of the environmental exposures were calculated and reported.

b) Differences in Exposures by Deprivation.
Unpaired t-tests were undertaken to investigate the difference of the means of the environmental exposures by deprivation. Unpaired tests were run as the observed values in the "walkers" group were independent of the observed values in the "non-walkers" group.

Deprivation was defined as affluent or deprived; as recorded in the HSW dataset. Schools that participated in the study provided the researchers of HSW with information regarding the number of children receiving FSM and this was used to assess the deprivation of the school. FSM are given to children based on family income and are a marker of the individual child's deprivation. Schools were identified as deprived if more than 21% of pupils were eligible for FSM. This indicator of deprivation was used because individual level deprivation was not available. More

widely used deprivation indicators such as Welsh Index of Multiple Deprivation (WIMD) and Townsend were available to use however, I felt that it was not appropriate to mix small-area deprivation indices with the individual level environmental exposures. The small-area geographies used as the basis for the deprivation indices were not necessarily spatially representative of the home, school and walking routes.

c) Descriptive Statistics of "Walkers" and "Non-walkers"

Unpaired t-tests were also undertaken to investigate differences in BMI and number of risk factors between the cohort of "walkers" and "non-walkers".

d) Multilevel Regression Analysis

A mixed effects linear regression model was fitted to the linked dataset in order to explore whether environmental exposures to the RFE were positively associated with BMI. The outcome variable was a continuous measure of BMI. The model was built up by adding in each main effect separately, rather than a forced entry method. School attended was included as a random effect within the model in order to account for clustering by school. The model was adjusted for age, gender, maternal BMI, ethnic minority and deprivation. Deprivation was a school level confounder that defined each school as affluent or deprived.

Age and maternal BMI were continuous variables. Gender and deprivation were defined as binary variables and ethnicity as a categorical variable. 'Hours of TV watched' and 'number of risk factors' were not included as confounders as studies report that screen time does not predict obesity for this age group [421, 428], and 'number of risk factors' is not a predictor for obesity [421]. Furthermore, the t-test results showed that there was no difference between the "walkers" and "non-walkers" for hours of TV watched and number of risk factors.

All statistical analysis was undertaken within the SAIL gateway, using R (version 3.3.1) and packages: nlme; psych; dplyr. Only data available were analysed, no estimate of missing data variables were included.

9.3 Results

9.3.1 Linking Exposure, Health and Administrative Data

Of the 1147 HSW participants, 916 were successfully linked to the WDS dataset. Erroneous BMIs were then removed (n = 33). The anonymised exposure data was then linked with the 883 individuals. Household level exposure data (home exposure, school exposure, exposure am, exposure pm) were linked to 431 individuals who were allocated to the "walkers" group. The remaining 452 unlinked individuals were classified into the "non-walkers" group. The linkage is summarised in Figure 9.3.
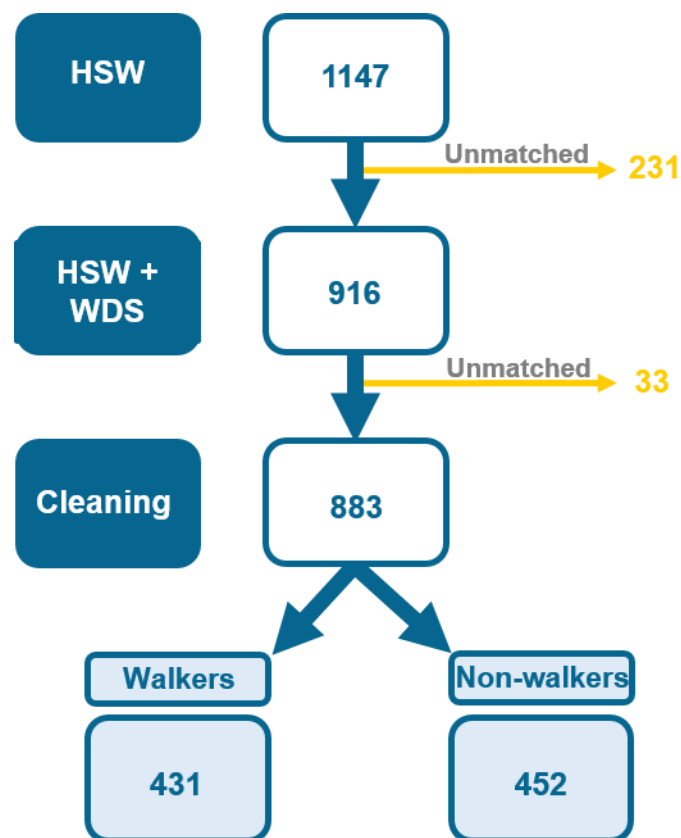
Figure 9.3. Data linkage process shows number of records
matched at each stage of data preparation

9.3.2 Statistical Analysis

a) Environmental Characteristics of the "Walkers" Group

On average, home exposures were the densest exposures. The home exposure was more variable than the other exposures with a standard deviation of 11.8. The mean school exposure was 3 with a maximum exposure of 14. Children walking home from school were exposed to more food outlets than on the way to school. The maximum number of outlets passed on the walk to school was 14, and doubled for the walk home. The average distance travelled was longer along the walk home compared with the walk to school and the distance travelled on the way home was more variable than the walk to school. The maximum distances travelled were 5885m (to school) and 6190m (from school). The distance that a child travelled from school was on average 78m longer than their walk to school. For a single individual, the maximum difference between the walk to and from school was 1225m longer on the walk home compared with their walk to school (Table 9.2).

Table 9.2. Environmental exposures and distances in metres for "walking" group

Exposures	Mean	SD	Median	Min	Max	Skew
Home Exposure	9	11.8	5	0	129	4.1
School Exposure	3	3.9	3	0	14	1.9
Exposure am	2	2.2	1	0	14	1.8
Exposure pm	3	3.9	2	0	28	2.4
Distance am	2143	1237	1855	40	5885	0.7
Distance pm	2220	1325	1890	40	6190	0.8

N.B Environmental characteristics were not calculated for the non-walking group. This would have required an excessive computational effort to model routes from all homes to all schools in south Wales

b) Differences in Exposures by Deprivation

When categorised by deprivation, there were statistically significant differences between home exposure, exposure am, and BMI. Exposure pm and school exposure were not different for affluent and deprived individuals (Table 9.3).

Table 9.3. Differences in environmental exposures and BMI by deprivation

Home Exposure	n	Mean	St Dev	St Error	
Affluent	244	6	1.2	0.78	t = -5.3
Deprived	187	12	10.5	0.8	**p = < 0.05**
School Exposure					
Affluent	244	3	4.6	1.0	t = 0.8
Deprived	187	1	6.1	2.7	p= 0.5
Exposure am					
Affluent	244	1	2.2	0.1	t = -2.8
Deprived	187	2	2.2	0.2	**p = < 0.05**
Exposure pm					
Affluent	244	3	4.7	0.3	t = 0.1
Deprived	187	3	2.7	0.2	p = 0.9
BMI					
Affluent	244	19.7	3.2	2.1	t = -2.5
Deprived	187	20.5	3.56	0.3	**p = < 0.05**

c) Descriptive Statistics of "Walkers" and "Non-walkers"

28% of "walkers" were overweight compared with 33% of "non-walkers". This difference was not statistically significant (chi-squared = 2.60, p = 0.11). Only 13% of "walkers" were obese compared with 22% of "non-walkers". This difference was statistically significant (chi-squared = 12.3, p < 0.05). Figure 9.4 shows the proportion of healthy, overweight and obese individuals in the study. The number of individual

risk factors did not differ between the "walkers" and "non-walkers" either (t = -1.61, p = 0.12). The BMIs of "walkers" were smaller than the BMIs of "non-walkers". This difference was statistically significant (t= -2.88, p < 0.05).

Table 9.4.Student t test results comparing the difference between BMI, hours spent watching TV and number of risk factors in the "walkers" and "non-walkers" groups

BMI	n	Mean	St Dev	St Error	
Walkers	431	20.1	3.4	0.2	t = -2.9
Non Walkers	452	20.8	4.1	0.2	**p < 0.05**
Number of risk factors					
Walkers	431	1.5	1.8	0.1	t = - 1.6
Non Walkers	452	1.7	1.9	0.1	p = 0.1

Figure 9.4. Difference in proportion of walkers and non-walkers that are of a healthy weight, overweight or obese

d) Multilevel Regression Analysis

Table 9.5 shows the unadjusted model for BMI, home exposure to the RFE, school exposure to the RFE and exposure to the RFE along the walk to and from school. The

model was adjusted for age, gender, maternal BMI, ethnic minority and deprivation (Table 9.6). Exposure pm was positively associated with BMI. Conversely, exposure am was negatively associated with BMI. Home and school exposures to the RFE were not associated with BMI.

Table 9.5. Unadjusted mixed effects linear regression model

Fixed Effect	Value	SE	t-value	p-value
Intercept	20.25	0.25	80.17	< 0.05
Home Exposure	2×10^{-3}	0.02	-0.13	0.89
Exposure am	-0.86	0.22	-3.73	< 0.05
School Exposure	-0.08	0.05	-1.40	0.19
Exposure pm	0.67	0.20	3.10	< 0.05

Table 9.6. Fully adjusted mixed effects linear regression model. Adjusted for: age; gender; maternal BMI; deprivation and ethnicity

Fixed Effect	Value	SE	t-value	p-value
Intercept	10.09	3.79	2.66	< 0.05
Home Exposure	-0.01	0.02	-0.81	0.42
Exposure am	-0.93	0.28	-3.27	<0.05
School Exposure	-0.00	0.07	0.00	0.99
Exposure pm	0.53	0.22	2.38	<0.05
Age	0.52	0.31	1.67	0.09
Gender	-0.92	0.38	-2.43	<0.05
Maternal BMI	0.14	0.04	3.85	<0.05
Deprivation	1.07	0.48	2.19	0.11

9.4 Discussion

9.4.1 Main findings

This chapter has used data linkage to investigate whether exposure to the RFE is positively associated with a BMI in children aged 11-13, living in south Wales.

Home exposures were the densest of the four environmental exposures. The home environment provides access to a greater number of outlets compared with the three other environments. This is because of the amount of time assumed to be spent in the home environment. Home exposures were also the most variable of the four exposures. This may be representative of the spatial distribution of food outlets. That is, homes in urban environments will likely have a greater access and exposure to food outlets. Furthermore, the results from this study suggest that exposure to the RFE is greater in deprived regions. This supports what has been reported elsewhere in the literature [223]. Home exposure to the RFE was not associated with BMI and this also resonates with what has been previously reported in the literature [232]. In this study, the home environment is defined as an 800m network buffer from the home location, which is intended to reflect the distance that could be travelled in a 10-minute walk [378]. Studies that have measured the home environment in this way have found little or no association between exposure to the RFE and BMI [56, 232]. This suggests two things: a) the home environment may not be as important in the pathway of obesity for children; b) the way researchers measure the home environment does not capture how children interact with the RFE there.

On average, exposure am was less than exposure pm. This may be because fewer outlets are open when children walk to school compared with when they walk home from school. Also, children walk shorter distances on their walk to school (Table 9.2) and therefore have a reduced chance of walking past a greater number of outlets. Like home exposure, exposure am was greater for children living in deprived catchments, compared with children living in affluent school catchments. This finding supports previous studies [175, 183, 222, 430]. However, a greater exposure to the RFE in deprived areas is not necessarily a key contextual driver for the poorer diet related health outcomes observed in deprived communities [178]. Studies show that a greater exposure to the RFE can influence the amount of unhealthy food eaten but this does

not necessarily equate to higher BMIs [168]. This is further supported by the finding that exposure am was in fact negatively associated with BMI. This finding does not show causality and the reason for this relationship is yet to be explored. It may be that on the route to school, children are less influenced by environmental exposures as a higher exposure on the walk to school is not associated with a higher BMI. Pupils have to arrive at school on time and have been shown to take a more direct route when walking to school in the morning. However, for children that skip breakfast, it may be that they choose to visit a specific outlet. Children who skip breakfast have been found to be more likely to have a higher BMI and be frequent consumers of unhealthy food. This is described as "selective daily mobility bias" in the literature [267, 358]. This suggests that a route is selected based on personal preferences, such as the availability of food, which are linked to BMI [268]. Further research needs to be undertaken to explain why children with a lower exposure to the RFE are more likely to be obese than children who walk past a greater number of unhealthy food outlets.

Exposure to the RFE in the school environment was on average, relatively small. However, one school had a much greater exposure to the RFE (n =14) than the other nine schools. This school was in a dense urban environment and within 400m of the school was a busy high street where many cafes, takeaways and convenience stores were located. For the other nine schools the range of exposure to the RFE was 5 (min = 0, max = 5). No association between a greater school exposure to the RFE and a higher BMI may have been found because outlets within 400m of the school premises may not be representative of how children interact with the RFE before and after school. More research needs to be undertaken to understand purchasing habits of secondary school pupils [68]. In Wales, the Healthy Schools scheme [431] prevents children from leaving the school premises at lunch time to purchase food. The menus that schools provide for lunch are designed to be healthy and promote a good relationship with food – whether children enjoy the food and the effectiveness of this policy has not be evaluated or reported. Research should be undertaken to see whether this policy has reduced the influence of the RFE surrounding schools as children are not able to buy lunch outside of the school premises. Of course, outlets surrounding schools still give children the opportunity to buy unhealthy food when they are arriving at school and when they are on their way home but 400m surrounding the school may not reflect where children choose to purchase unhealthy food. In addition, the

difference in exposure to the RFE at school between affluent and deprived areas was not statistically significant. This may again be related to how the measure of school exposure was calculated. 400m around the school premises may not be large enough to capture the difference in the spatial bias of the distribution of food outlets that has been reported elsewhere in the literature [432].

Exposure to the RFE can double on the walk home compared with the walk to school. This will likely be because a greater number of outlets will be open for trading during the time that children walk home from school. Exposure pm was also more variable than exposure am and average exposure pm does not differ between deprived and affluent students (t=0.07, p = 0.93). This may be because children walk longer distances on their route home compared with the more direct walk to school (Table 9.2), therefore increasing the likelihood of walking past more outlets. Finally, without the time constraint of having to be somewhere, children can choose to take a longer route home via outlets (results from Chapter 6 & 7 show this). This supports injury research where it has been reported that children are more independent along the walk home [433].

Exposure to the RFE along the walk home was the only environmental exposure positively associated with a higher BMI. A previous study found no association between the walk home from school and BMI but this study characterised the route home as a Euclidean 2km buffer. Characterising the walk to and from school in a more accurate way using GIS has not been reported until now. Walking past a greater number of outlets on the journey home from school may be associated with a higher BMI because the greater availability of unhealthy food promotes the consumption of unhealthy food which results in a higher BMI. A higher exposure to the RFE along the walk home may be influencing consumption of unhealthy food and thus contributing to childhood obesity. Previous studies have reported that a greater exposure to takeaways are associated with a greater consumption of unhealthy food [168, 434]. Although there is a lack of evidence whether this greater consumption translates to a higher BMI; longitudinal studies suggest that a greater exposure to fast food outlets is associated with weight gain over time in young adults [435, 436]. On the other hand, as aforementioned, the literature suggests that a greater exposure to the RFE in other activity spaces, such as the home environment, is not necessarily associated with a higher BMI. Therefore, further research needs to be undertaken to explore whether

there is a causal relationship between exposure to the RFE on the walk home and an increased BMI. Differentiating between the walk to school and the walk home has not been reported before using GIS methods. The findings suggest that the walk home from school may be an opportunity to develop interventions and policies. The difference between the exposure to the RFE along the walk home from school for deprived and affluent individuals was not statistically significant (p = 0.93). This would suggest that if interventions are to be considered, they should be developed in deprived and affluent areas.

The "walkers" and "non-walkers" groups showed differences that were statistically significant. The average BMI for "walkers" was less than "non-walkers" and the number of obese individuals in the "walkers" was less than the number of obese individuals in the "non-walkers" group. These differences may be because walking to and from school increases children's overall levels of physical activity [414] and this results in a lower BMI. However, it may be that this relationship is not so simple. A systematic review concluded that children who actively travel to school tend to generally be more physically active but there is limited evidence to suggest that actively commuting to and from school promotes healthy body weights [437]. This is supported by the results from this thesis that showed the difference in the number of overweight individuals between the walkers and non-walkers was not statistically significant. Conversely, a study conducted in America found a negative association between body weight and school neighbourhood walkability [268].

The number of risk factors an individual had was not different between the "walkers" and "non-walkers" groups either. Studies consistently show that active school travel (AST) has many benefits, but the literature is mixed whether AST alone, increases physical activity levels enough to reduce sedentary behaviour in the home environment and risk factors of obesity. Policy should continue to encourage AST as it encourages and supports a healthy lifestyle. AST cannot be relied upon as a single intervention to reduce childhood obesity. However, incorporated as part of a systems approach, AST can help to promote increased physical activity and promote good health.

9.4.2 Strengths

The main strength of this study is that multiple aspects of the daily exposure environment (home, school and walk to and from school) have been linked with health data. This has not been reported before. Policy to promote healthy eating in the USA and UK focusses on the spatial access to food outlets around home neighbourhoods and school premises [181]. The results presented in this chapter suggest that children's purchasing habits on the way home from school should be better understood. This will provide insight on whether the route home from school would be an effective environment to implement policies and interventions.

A further strength of this study is the use of routinely collected data. Many studies assessing the effect of the environment on BMI have used self-reported data. Self-reported values are usually available on a larger scale as it is less resource and skill intensive to record this data type. However, some studies have found it to be an unreliable data source, particularly with regard to children and adolescents [438–440]. The HSW dataset was collected by trained experts and so the data quality can be trusted.

The statistical analysis that was undertaken was robust. The HSW dataset was a representative sample across socioeconomic backgrounds and the BMI was normally distributed in the cohort. Therefore, the results that are presented in this chapter are for a population whereby there are no underlying characteristics that may have biased the results. The multilevel model (MLM) ensured that the hierarchical structure of the data was accounted for; adjusting the MLM for age, gender, deprivation, ethnic minority and mother BMI to account for any association between the confounders and BMI.

9.4.3 Limitations

A limitation of this work is how the participants of HSW were categorised as "walkers" and "non-walkers". Environmental characteristics were not calculated for the non-walking group. This would have required an excessive computational effort to model routes from all homes to all schools in south Wales. Therefore, assumptions had to be made to categorise the participants as data was not available to identify which pupils in the HSW dataset walked to and/or from school. The assumption was made that participants who lived less than 4.8km away from school would be likely to walk,

and participants that lived more than 4.8km away from their school would not walk. If an individual actively travels to and from school it is very likely to be walking as less than 1% of secondary school pupils cycle to school in Wales [441]. The assumption that participants would not walk if they lived further than 4.8km from their school was deemed reasonable for two reasons: 1) In the UK, children that live more than 4.8 km from their nearest school are entitled to free transport [413]; 2) the characteristics of the GPS data collected in Chapter 7 suggest that children rarely walk further than 4.8km to school (> 99% of the walking routes were shorter than 5km).

Due to the time constraints of this project, it was not possible to link information on the food outlet type with the health data. This should be explored in future work as Hobbs et al. found different sizes of association with exposure to different food outlet types and BMI [10]. Evaluating whether different outlet types are more or less associated with higher BMIs could be useful for policy makers. Equipping policy and decision makers with a more in depth knowledge of the RFE should aid them in developing more successful interventions and policies. For example, policies banning fast food outlets within 400m of school premises are becoming increasingly popular for LAs to implement in the UK, but the surveys I undertook (Chapter 5) suggest that children buy unhealthy snack food, rather than takeaway meals. This suggests that the type of food children are purchasing on their journey to school and home should be investigated more thoroughly, as it may be valuable to explore the role of convenience stores in the role of children's food purchasing habits.

9.4.4 Implications

a) For Research

Exposure to the RFE along the walk to school and the walk home have not been reported separately before. Differentiating between the walk to school and the walk home has shown that children are exposed to a greater extent to the RFE along the route home from school. Not only did the routes to school and home from school vary in shape (due to the weighted model) but outlet opening times were included. Children's routes to and from school have not been modelled to this complexity before, but the results suggest that high resolution modelling is required to be able to highlight complex patterns and behaviours.

Exposures to the RFE have not been linked to anonymised health data before. This is novel and contributes to the literature on linking cohorts of environment and health data to undertake research for the benefit of society.

Routinely collected, population level BMIs are not currently recorded in Wales. There are several surveys that collect BMI data on a large scale in Wales and the UK [442, 443], but this data is self-reported. BMI data is available in several datasets held in SAIL. However, BMI is not routinely recorded by GPs and therefore may not be representative. People who do have their BMI recorded will likely be for a health related reason e.g. they are overweight or underweight. The results presented in this chapter highlight that a population level record of BMIs would be a powerful tool in obesity research. The evidence presented here supports the need for the continuation of schemes like the CMP in the UK. In Wales, the Child Measurement Programme (CMP) has not been rolled out across age ranges 9-10 due to financial constraints. However, if the CMP were to record longitudinal BMI data, this would be extremely valuable. There are a lack of longitudinal obesity studies and in the UK, none have been published. Different aspects of obesity drivers could be investigated and of particular interest to me would be whether a change in the environment over time has an effect on obesity rates.

This chapter defined the home environment as an 800m network buffer around the home location [85, 444]. The 800m buffer is one of many ways that the home environment has been defined and measured in the literature [70, 71, 153]. Methodological differences have made it very difficult to interpret whether access to the RFE at home influences health outcomes and obesity. Standardised methods of defining the home environment should be agreed so that studies can be comparable. In order for a standardised method to be proposed and accepted by the research community, how people define their home environment needs to be investigated. To my knowledge, studies using GIS to investigate the influence of the environment on health, have not incorporated how residents define their home environments. There are studies that have used resident-defined neighbourhood mapping and it may be possible to draw from these studies, methods that could help health-geographers [349, 445]. This is important to understand because if researchers are not characterising the home environment as people relate to it in real life, our results are meaningless to inform policy makers and practitioners. How people relate to their home environment could

be investigated through qualitative research methods such as interviews and questionnaires and quantitative methods such as GPS trackers. For example, a similar approach could be taken to define the home environment as I have taken to define route choice along routes to and from school. Using GPS data would highlight characteristics of the home environment and show how people relate to their home environment. How people relate to their home environment may differ by age, gender, socioeconomic status and whether you live in an urban or rural area.

The same could be suggested for defining the school environment. This study has quantified the school RFE as a 400m network buffer around the school premises but 400m around the school premises was an arbitrary number chosen by policy makers. There is no strong evidence to support the concept that limiting food outlets within 400m of secondary schools will be beneficial to school children. 400m may not necessarily be characteristic of where children are influenced by the RFE. Research should be undertaken to understand children's purchasing habits and identify if there are particular aspects of the RFE that children particularly engage with. Studies have found that interventions inside the school premises are very effective. Researchers and policy makers should discuss the current evidence base to decide whether making changes to the RFE immediately surrounding schools is an effective use of resources.

The finding that exposure pm is positively associated with BMI is an important finding. Further research needs to be undertaken to investigate whether a higher exposure to the RFE along the route home is *causing* higher BMIs. Future studies should look to include purchasing data and eating habits and this would give insight in to the pattern that has been identified in this study.

The home locations of individuals were identified using the WDS. However, the duration for which the individual had lived at that premises was not accounted for. It may have been the case that an individual may have recently moved into the home they were identified as living in. If this was the case, the environmental exposures that an individual were subject to at their new home and potentially new school and walk to and from school would not have had as great an influence on the individual. Exploring the longitudinal element of exposure to the RFE should be explored in future research. A challenge of longitudinal obesity studies however, is the lack of longitudinal BMI data available.

b) For Policy

This study resonates with and supports the aim of the Well-being of Future Generations Act 2015 [446]. The Act is based on the knowledge that the environment, economy, society and culture are all connected and means that public bodies have to work together to achieve a healthier, sustainable and more equal Wales. The results from this study can be used to recommend future research and also to provoke evaluation of current policies.

The implications of this research on policy are closely related to the implications on research discussed above. This is because Public Health research looks to be informed by research and research looks to inform policy. Investing in policies that limit fast food outlets around school premises may not be the most effective policy to implement. Research has shown that access to unhealthy food does influence consumption and attitudes [434, 447] and in order to achieve a reduction, or at least curbing of obesity rates, healthy attitudes to food must be encouraged. Local Planning Authorities (LPAs) must take responsibility in the allocation of where food premises are positioned, particularly unhealthy food premises. It is unethical to close outlets based on the food that they sell but the question should be asked of whether we should be encouraging new businesses to be contributing to the culture of demand – when the demand is detrimental to people's health. The Scottish Government released a statement in September 2017 that said unpopular choices would have to be made if obesity was going to be cut. A doctor was also quoted as saying until the government is willing to take an economic hit, obesity rates will not change. It is important to empathise with business owners but a constructive way of encouraging healthy eating habits has to be found if we are to curb and reduce obesity rates. It may be painful for businesses to have an economic hit but the long term goals of healthy attitudes to food and reducing obesity are so important. Currently obesity is costing the NHS billions of pounds every year.

It is not known whether limiting new fast food outlets within 400m of school premises is an effective way to reduce obesity as not enough evaluations have been undertaken. However, if researchers can meaningfully define the "school environment" and understand buying habits of school children then this will provide insight in to where or when restrictions could be placed on food outlets selling unhealthy food. The policy of limiting fast food outlets within 400m of school premises has not been evaluated

and it would be worthwhile to evaluate whether this policy has had an effect through a natural experiment. Obesity is very complex and will require interventions at multiple scales in different environments that we experience throughout our day.

Reporting the result that exposure pm is positively associated with BMI should be balanced to support current policy that encourages active travel to school. Care needs to be taken in reporting these results as it would be detrimental for parents to remove the opportunity for physical activity in order to reduce a child's exposure to the RFE. Active travel encourages an active lifestyle and this could be more beneficial than the detrimental effect of purchasing an unhealthy snack on the route home. It will be important to understand what types of food children are purchasing on their route home and whether an exposure to the RFE influences their general attitude to food.

The Active Travel Wales Act 2013 [441] requires LAs to plan for and promote active travel journeys in Wales. The "Active Travel Promotion in Schools" has roots in this act. An evaluation of this program is planned to be undertaken in July 2018. The results of this study should be shared with the Welsh Government to inform them and give further insight for the evaluation.

Policy makers should look to invest in infrastructure such as footpaths and cycle paths. This would help to fulfil the aims of the "Active Travel Promotion in Schools" programme by promoting active travel to and from school and it gives children the confidence to actively travel to school as footpaths and cycle paths are safe from the dangers of traffic. Three times more traffic accidents involving children happen on the way home from school compared with children travelling to school. Investing in infrastructure would help children to safely commute by encouraging children away from main roads and locations where unhealthy food outlets are located [433].

9.4.5 Future work

There are many opportunities to develop the work presented in this chapter. Firstly, there is currently a travel-to-school survey that is being piloted by Neath Port Talbot council and the outputs of this survey will be imported in to SAIL. There are plans for this survey to be implemented in schools across Wales. The potential to know how a child travels to school would refine the model of predicting children's exposure to the RFE along their routes to and from school.

A further point of development would be to undertake the analysis on a larger population. The BMI data used in this study was from a single dataset held in SAIL. Scoping would need to be undertaken in SAIL to assess alternative datasets that contain information on BMI. BMI is not widely recorded for individuals in SAIL as it is not routinely collected by GPs for the UK population. Height and weight are measured when an individual first registers with their GP. For children, height and weight are measured for national programs such as the CMP. Routinising height and weight collections would be valuable in providing large cohorts of longitudinal BMI data for representative samples.

School level deprivation was used as an indicator of deprivation in this study. An individual level indices of deprivation is under construction at Welsh Government [448] and this should be explored as a confounder in the analysis. When analysis is being undertaken at such a high spatial resolution, deprivation indices for small area geographies are not necessarily helpful as information can be lost and it is not possible to understand sub-group characteristics [449].

The HSW dataset was very rich and this means that the dataset has the potential to be interrogated further and used in further research on this topic. Furthermore, other outcomes could be linked to environmental exposures such as educational outcomes or hospital admissions.

9.5 Conclusions

A higher BMI is associated with a greater exposure to the RFE along the walk home from school. This relationship has not been identified as causal, but the findings suggest that the walk home from school is a new environment in which to explore the possibility of developing interventions and policies. Investing in infrastructure such as footpaths and cycle paths will support current policies in place and could reduce children's exposure to the RFE. Further research needs to be undertaken to better understand children's purchasing habits. This will help to further inform policy makers and practitioners in developing successful interventions and policies that will result in a population level decrease in obesity.

Chapter 10

Conclusions

A novel methodology has been developed to model children's exposure to the 'retail food environment' (RFE) and these exposures have been linked in a unique data linkage environment with health data. Exposure to the RFE along children's walking routes to and from school were positively associated with BMI. To undertake this study, food outlet data was collected for Wales, an open source road network dataset was built and evaluated, and GPS data was collected to more fully understand route choice for children that walk to and from school.

Extensive data collection and data quality checks were undertaken to ensure that the exposures would be reliable. A longitudinal dataset of the RFE for Wales was collected from Local Authorities (LAs). LAs collect and store a goldmine of data and the potential to use this for secondary purposes, especially in research, is vast. However, because of limited resources, time and sometimes perspective, it can be challenging to work with LAs and obtain their data. Researchers need to promote the use of LA data, look to forge relationships with them and work synergistically. A local authority in England has embedded a researcher within the department, and this has been successful in strengthening relationships, improving data availability and collaboration.

This study reports that LA food outlet data is the most suitable for studies investigating exposure to the RFE. We can be confident that LAs accurately record food outlets as it is a legal requirement to register a food business with the LA at least 28 days before trading begins. However, the completeness of the spatial element of these datasets, varies across Wales. LAs were able to provide address level point data to varying extents. Despite this, addresses were provided for over 99% of outlets and therefore, the outlets could be geocoded. Although this study was cross-sectional, a longitudinal dataset was required as BMI data was available for 2009-2010. PoI and FSA were not appropriate data sources for this study. The PoI dataset has address level spatial locators and longitudinal data is available from 2015, however, the dataset contains

many duplicates and closed outlets. Furthermore, there is no way to identify outlets that have closed for trading. The FSA data was not appropriate as the data was postcode level, cross sectional data. Using postcode level data risks introducing ecological fallacy and cross sectional data does not capture the change in outlet location. Secondary datasets are being increasingly used in exposure studies, but the quality of these data sources is under reported. Early studies investigating exposure to the RFE initially collected data from LAs, but the data was difficult to obtain and required processing to be operationalised for spatial research. Secondary sources are being used as an alternative to LA data as they are free to download and are perceived to be research ready. Researchers should be upfront about the data that they are using in studies and the limitations associated with the data [10].

Further to collecting RFE data, an open source road network dataset was built and evaluated to assess whether it is a better alternative to a commercially produced road dataset. Using an open source road dataset was of interest so that the research could be reproducible elsewhere in the world. OpenStreetMap (OSM) was reported as a suitable alternative to the Ordnance Survey Integrated Transport Network (ITN) dataset in modelling teenager's exposure to the RFE. The four prominent findings of the evaluation were a) OSM is comparable with ITN for road and footpath coverage in Wales; b) OSM minor roads coverage is lacking in rural mid-Wales; c) Footpaths are an important part of the network when modelling teenager's walking routes to school d) OSM records more footpath data than ITN.

In addition to collecting data to generate exposures to the RFE, another primary aim of this study was to develop a new GIS based methodology that could routinely model exposures along children's walking routes to and from school that were positively associated with GPS exposures. In order to develop a method that would calculate exposures that were representative of GPS exposures, route data was collected from school children aged 13. Three rounds of data collection were reported, the second being informed by the limitations of the first and the third being informed by the limitations of the second. Self-reported routes were first collected for 11 teenagers aged 13-16. This was an important pilot study but: a) the sample was small (11) b) only routes *to* school were reported c) there were limitations with the data collection method. The second round of data collection involved working with a local secondary school and collecting GPS data using a mobile phone app and food diary data from the

participants. This was a well prepared, novel data collection method but it was extremely challenging to work with schools and participation rates were low (1/6 schools participated). This process was a steep learning curve but an excellent experience for working with participants. This data collection round did not yield a large sample and so a research team was approached to use data that had already been collected. This was a rich dataset of 995 students aged 13-14 from Bristol for the PEAR study. I developed an automated cleaning algorithm which selected out the walking routes from this data. The preparation of this GPS data provided my project with a robust dataset that could be used to inform and evaluate a weighted network GIS model.

Environmental characteristics of the GPS walking data were analysed, and the results informed the weightings used to build the road network model. Calculating routes using the shortest network algorithm along a weighted network resulted in more realistic route choice than the standard shortest network route. Future research could explore machine learning methods to optimise the weightings of the road network.

Exposures to the RFE were calculated along the GIS modelled routes to and from school, and for the home and school environments. These environmental exposures were linked to individual anonymised health data in the SAIL databank. Linking environment data and obesity data has not been undertaken in this way before. The results show that exposure to the RFE is positively associated with BMI. This study does not report causality, but it does suggest a new area to develop policies and interventions. More research is needed to understand children's purchasing habits, and this could inform the types of food children are buying and from which outlet types they are making purchases.

This thesis reports a new methodology for modelling children's walking routes to and from school which contributes to the current literature. This method has the potential to allow population level exposures to be calculated, something that is not possible with GPS data. Population level analysis is essential if we are to inform policies that will be able to drive a societal change resulting in large reductions in the prevalence of obesity. Collecting a longitudinal RFE data for the whole of Wales will prove to be a valuable data source for future studies. Researchers must continue to promote the use of LA data as it is a valuable source of administrative data which can be reused for

secondary purposes, such as research. However, LAs often lack time and resources to maintain the data that they collect in a way that is easily used in secondary data analysis. Continuing to work with LAs to demonstrate the usefulness and benefits of administrative data will contribute to a collaborative way of working between academia and LAs. This synergy will provide researchers with rich data sources and LAs with expertise, resources and research that may not be available to them without this professional relationship.

There are many ways in which this study can be developed for future research. These are discussed in detail through the thesis but the most prominent developments would be:

- Promote the use of LA data and encourage LAs to engage more with researchers;
- Investigating whether different data sources produce different exposures and how this influences the observed association with BMI. Hobbs et al reported that there was a small difference in the size of associations between food outlet exposure and BMI between two different food outlet datasets [10]. However, this study compared exposures in the home neighbourhood and they were defined as LSOA boundaries and Radial buffers – which could be described as over simplified definitions;
- Combining OSM and ITN footpath datasets to produce a hybrid network. This would provide a more thorough coverage, particularly for footpath data
- Explore machine learning techniques to find optimum weights for network;
- Undertake research to understand more fully the purchasing habits of teenagers during a school day;
- Explore the potential for using more BMI data that is currently stored in SAIL/Promoting the routinisation of recording BMI measurements.

The cost and impact of obesity is widespread and far reaching across the globe. Without change, the health costs associated with being overweight or obese will become insupportable [9]. A policy that is becoming increasingly popular is the limiting of fast food outlets within 400m around a school. The results from this study show that exposure to the RFE within 400m around the school premises is not associated with BMI. Further research should investigate the typology of outlets

surrounding school premises and whether children purchase unhealthy food from these outlets. Furthermore, preliminary results in this study suggest that fast food takeaways may not be as important in the childhood obesity pathway as policy makers think. Policies may be better placed focussing on encouraging convenience stores and mini-supermarkets to make healthy foods more appealing to children [450]. Interventions

Encouraging active travel to school is another major policy area in the UK [451] but for the most part, policy has not been successful in increasing the number of children that actively travel to school. We need to make it easy and safe for children to actively travel by investing in infrastructure that can facilitate an active lifestyle. The findings from this study support the concept that the environment is playing a role in childhood obesity, but we need to be looking at system change that together can drive large scale reduction in obesity. Interventions that promote active travel to school, without making modifications to the built environment have limited success rates [452]. Interventions such as speed limits and one way systems reduce the number of road traffic accidents. In turn, this creates safer environments where children are more likely to actively travel [453]. There is evidence to suggest that improving the built environment to create safer and more interesting environments to traverse through increases physical activity.

The route to and from school is not the reason for childhood obesity. However, if we want to encourage children to actively travel we need to provide them with an environment that will not undermine the positive action of active travel by providing abundant opportunity to purchase unhealthy food. Innovative policies and interventions that can support system level change are beginning to be implemented in the UK, for example, the sugar tax. But a greater understanding of the relationship between the built environment and health will help to develop interventions that can continue to reduce health burdens [454]. Furthermore, moving policy and research away from linear, causal models and facilitating system change resonates with the intension of the Future Generations Act in Wales. This act acknowledges that the environment society and culture are all connected and that people need to live in a community that has everything they need to thrive and prosper. This includes a safe and supportive environment that will promote physical activity and healthy lifestyle choices that will promote good health. Positive changes can and need to be made to the built environment. These changes should be informed by research that can infer

causality and have the potential to drive system level change that can reduce obesity levels on a population scale [103].

Appendix 1.

Initial email sent to Local Authorities

Dear XXXX

My name is Amy Mizen, and I am a PhD student at Swansea University, in the College of Medicine. I am investigating whether the location and potential change in density over time of fast food outlets found on home-to-school routes has an impact on childhood obesity.

I have been given your contact details by Mr Ronnie Alexander. I was wondering whether you have a historical database of food outlets for Blaenau Gwent, which I could import into a GIS to calculate outlet density around routes to schools. If not, are you able to suggest another contact who may be able to help me please? We have received a comprehensive dataset from Swansea City and County Council, so in order that the datasets are comparable, if possible could we have the following details:

- *Address/XY co ordinates*
- *UPRN*
- *Address*
- *Opening times*
- *Any outlet type categorisation*
- *When the premises opened*
- *When (if) the premises closed*

Thank you in advance for any help that you are able to give.

Kind Regards,

Amy

Appendix 2.

Local Authority Classifications of food outlets

Cardiff	
LA Category	Usage
	Fixed Pitch/Market
	Mobile Trader
	Public House
	Club (Profit Making)
	Registered Members Club
	Take Away Premises
	Restaurant/Cafe
	Canteen
	Hospital Kitchen
	Home Caterer/Outside Caterer
	Burger Van (SUBSIDIARY ONLY)
	Ice-Cream Van (SUBSIDIARY ONLY
	Nightclub
	School Production Kitchen
	School Breakfast Club
	After School Club
	School Clubs
	Play Group / Centre
	School Dining Kitchen
	Youth Centre/ Club
	Catering Other
	Health and Beauty
	Consumer Services Other
	Church Hall
	Amusement Arcade
	Sports Grd/Golf/Playing Field
	Theatre
	Sports Hall
	Bingo Hall
	Leisure Centre
	Community Hall
	Casino
	Special Events
	Children's Play Area
	Entertainment/Leisure Other
	Bakehouse/Bakery

Food Manuf. Other
Non Food Indus & Manuf. Other
Blg. and Common Areas
Company Offices
Other Offices
Educational Est.
Hospital
Central Govt. Est.
Council Est.
Nursery Sessional (LA)
Nursery Full Day (LA)
Nursery Full Day (HSE)
Nursery Sessional (HSE)
Church or Religious Est.
Other Public and Educational
New Car Sales
Haulage/Carriers/Removals
Grocers
Supermarket
Butchers
Fishmongers

Baker Shop & Flour Confect'ery
Sweet Confec/Tobacco/Newsagent
Chemist/drugstore
Minimarket
Off Licence
Ice Cream Retailer
Milk & Dairy Produce Retail
Deli/Health Food Shop
Other Retail Food
Sport Shop
Clothes shop
Book Shop
Charity Shop
Garden Centre
Florist
Department Store
Ironmongers
Non Food Retail Other
Hotel

Guest House/Boarding House
Nursing Home
Private Residential Home
LA and Housing Ass. Home
Small Homes (Res. Home)
Hostel

Domestic Premises
Day Centre
Child Minder
Accommodation Other

Greengrocer Wholesale/Warehous
food Wholesale/Warehouse Other
Non Food W/W

Carmarthen

LA
Category   Premises Usage
122 ICE CREAM
029 BUTCHERS
009 MEAT APPROVED PREMISES
267 Food Manufacturer
067 UNLICENSED CAFE/RESTAURAUNT
211 SUPERMARKET
066 CATERER
213 TAKEAWAY FOOD
187 RESIDENTIAL HOME
031 LICENSED CAFE/RESTAURAUNT
083 DIY/HARDWARE
276 Warehouse
239 OTHER FOOD
ZP6 RESTAURANTS
MCT Camping and Touring Caravan Site
168 Play Group
115 HOUSE FOOD PRODUCTION
146 MIXED PRODUCTS - RETAIL
019 BED & BREAKFAST
240 OTHER FOOD SERVICES
027 BUILDERS SUPPLIES
136 LEISURE
186 RELIGIOUS ESTABLISHMENTS
055 COMMUNITY HALL
605 SHARED HOUSING
097 FINANCE COMPANY
142 MOBILE SNACK ROUND
149 OUTSIDE CATERER
138 MEDICAL/HEALTH
051 CHEMIST
081 DOMESTIC
099 FARMING

012 AFTER SCHOOL CLUBS
191 SCHOOL/COLLEGE
092 FREEZER CENTRES
036 CLUB-LICENSED
264 Drinks/Smoothie Bar
145 MIXED FARM
095 FISH/SEAFOOD
251 Registered Gardeners and Handyperson
201 SPORTS/OUTDOOR LEISURE
141 MUSEUM
110 GARDEN SUPPLIES
053 CONF/TOBACCO/NEWSAGENT
C13 Food Vehicle - Catering
106 GROCER
246 OTHER SERVICES
207 STORAGE
248 Farm Shop
109 GREENGROCER
114 HEALTH/WHOLE FOOD
111 HOTEL/GUEST HOUSE-
UNLICENSED
254 Temporary Event Licence Location
104 GIFTS/FANCY GOODS
268 Delicatessen
148 NURSING HOME
073 DAIRY FARM

048 CINEMA/THEATRE
H01 Single Occupied House
270 General Store
046 CRECHE/NURSERY
273 Fitness Club
177 PUBLIC HOUSE
064 CHARITY SHOP
069 BEAUTY GOODS/TREATMENT
022 BAKER
242 OTHER GOODS
275 Internet Cafe
292 Food Producer
ZP7 SPORTS CLUBS
293 Care Facility
128 KITCHEN/BATHROOM SUPPLIES
107 GOAT FARM
165 PIG FARM
159 PUBLIC ENTERTAINMENT VENUE
238 WORKS/TOURIST ATTRACTION

118 HAULAGE CONTRACTOR
058 CONFECTIONERY

230 VEHICLE SALES/AUCTIONS
153 OFF LICENCE
266 Community Shop
065 CANTEEN
026 BREWERY
074 DAY CENTRE
IFG Meat Products Manufacturer
F99 Other Retail Food

ZP8 VILLAGE/COMMUINTY HALLS
070 CYCLES/MOTOR CYCLES
176 POULTRY FARM
198 SLAUGHTERING
170 POST OFFICE
216 VIDEO TAPE HIRE
MST Static and Touring Caravan Site
158 PETROL STATION
105 HOTEL/GUEST HOUSE-LICENSED
244 OTHER PERSONAL SERVICES
030 BOTTLED WATER PLANT
050 CLOTHING
140 MILK & DAIRY PRODUCTS
202 SPORTS HALL
108 GOVERNMENT OWNED
PROPERTY
233 VEHICLE SERVICE/REPAIR
161 PET FOOD
219 TOYS/GAMES
HMP Possible HMO's
H02 Single Occupied Flat
015 BABY/NURSERY EQUIPMENT

086 EGG PRODUCER
121 HOSPITAL
011 ARCHITECT
H03 HMO Bedsit
196 SHEEP FARM
135 CARMS. REGISTERED LOTTERY
SOC.
094 FARM SUPPLIES
120 HOLIDAY ACCOMADATION SELF
CATE
004 ANIMAL FEEDS

H06 Hostel/B&B
197 SHOW GROUND
103 GARAGE SHOP
H04 Shared House
016 BOATS/CHANDLERY

112 HAIRDRESSER
188 RIDING ESTABLISHMENT
241 OTHER HIRE SERVICES
020 BEEF FARM
127 JEWELLRY/CLOCKS
209 STREET COLLECTION SOCIETY
101 FURNITURE
010 ART/CRAFT
281 Complementary therapies
302 Large / Commercial water supply
HCP Housing Commercial Property
258 Etching and Glass Spraying
132 LAND
028 BETTING/GAMBLING
037 CARAVAN/CAMP SITE
232 VEHICLE PARTS
087 ELECTRICAL
AFB Fish Farm
166 AVERAGE QUANTITY PACKER
181 POTTERY
039 CARAVAN SITE - HOLIDAY
007 DAIRY APROVED PREMISES
008 FISH APPROVED PREMISES
199 Scrap Yard

RCT	
LA Category	PREMISES USAGE
	A01 Baker/Flour Confectionery
	A06 Grocers/Small/Med Supermarket
	A30 Ice Cream Parlour
	AAA Supermarket/Hypermarket
	B02 Conf/Tob/News/Stationer
	B13 Sweet Shop
	T00 Garage/Food Retail
	U01 Restaurant/Cafeteria
	U06 Takeaway Meals

Swansea		
LA Category	usage	sucode
	U19 Food Warehouse	U02 Caterers
	U01 Cafeteria & Restaurant (public)	
		U01 Cafeteria & Restaurant (public)
	C99 Other goods	A10 Delicatessen
	A05 Greengrocer	Z09 Casual Pitch Traders
	U12 Club (Members)	U06 Takeaway Meals
	N32 Watersports/Outdoor Sports	
		U20 Workplace Canteen
	A06 Supermarket	Z02 Market Trader
	A15 Mini Supermarket	A07 Off Licence
	A99 Other Food Premises	
	Z02 Market Trader	
	U17 Private House used for a Food Business	J70 Cooling Tower
	B11 Private House - Business	U21 Care home
	U26 Indian Takeaway	A04 Fishmonger
		B11 Private House - Business
	S28 Nursery	A99 Other Food Premises
		W05 Finance/Pawnbroker
	S34 Residential Home	N99 Other Leisure Premises
	U04 Hotel/Guest House	
	L01 Garage/Petrol	
	A10 Delicatessen	
	S29 Nursing Home	
	O08 Single Household	
	G03 Garden Centre/Fencing/Sheds	
	Z03 Mobile	
	U15 School "club" Food	
	P04 After School Club	
	U05 Pub / Nightclub	H04 Furniture
	U99 Other Catering/Accommodation	A01 Baker
	M01 Non-Food Factory	
	W06 Insurance	A02 Butcher
	B71 Charitable Organisation	
	U02 Caterers	
	U06 Takeaway Meals	

A02 Butcher
K99 Other Chemicals/Additives
A09 Chemist / Drug Store
C41 Tobac/News/Stationer
U03 Club (Proprietary)
A07 Off Licence
U27 Chinese Takeaway
P99 Other Public & Educational
C13 General Dept Store
U30 Fish & Chip Shop
O99 Temporary
N14 Casino
U20 Workplace Canteen   P14 School
A17 Cash & carry
A04 Fishmonger
U23 College/ University catering
A01 Baker   Q24 Spa
U22 Food factory
N99 Other Leisure Premises   Z03 Mobile
N22 Holiday Caravan Site   T16 Cars
P16 University
B91 Self Catering Holiday Accomodation
P03 School/College/University
A14 General Food Store
H03 Fancy Goods & Gifts   S28 Nursery
A13 Sweetshop
A19 Butcher Raw Meat Only
P14 School
A03 Dairy/Milk Depot
N25 Non-Licensed Club
P06 Other Council Establishment   A93 Alcohol
P11 School holiday clubs
U13 B&Bs (rooms in private houses)
S99 Other Service premises
U28 Turkish Takaway

C05 Clothes Shop
N16 Community Hall
N28 Sports Ground /
Premise
O12 Empty building
(non-dwelling)
N26 Outdoor Activity
Centre
P07 Post Office   Q44 CatII 4
U31 Fast Food Takeaway
U16 Residential/care
catering
A20 Butcher RTE/Raw
B12 Private house -
mobile trader

H05 Office Equipment
S44 Car wash
C14 Greeting Cards   B70 Lottery
N19 Entertainment /
Gambling   Q23 Pool
S04 Hospital
B99 Other miscellaneous
P09 Religious Meeting
Place
C10 Fabrics /
Haberdashery
S18 Ear Piercing -
Premises
Z05 Site of farmers/local
produce market
F11 Riding
Establishment
P15 College
H07 Electrical
Goods/Appliances &
Repairs
F42 Wholesaler/Retailer
Pet Food
A11 Health Food   Q36 CatI F
U29 Sandwich Shop
T40 Used Car Dealer
N31 Theatre
H04 Furniture
U07 Approved Premises
Fish
F29 Single Species Farm
U25 Day Centre   N31 Theatre
O07 Empty Dwelling   A91 Brewery

U10 Approved Premise
Meat Products
B20 Uncoded Premise - F
& S
A91 Brewery
Z07 Burger Van
F18 Smallholding
N30 Temporary Event
N23 Holiday Camp Site
S21 Music Shop
G02 D.I.Y. Inc
Equipment Hire
N15 Cinema
C32 Sports Shop
S24 Laundrette / Dry
Cleaners
B21 Proposed Food
Premises
C04 Book Shop
U24 Luncheon club
N13 Church Hall
T16 Cars
O02 HMO: CatB
Standard/Shared House   S04 Hospital
B90 Street Trading Pitch
F99 Other agricultural
C33 Shoe Shop
Z09 Casual Pitch Traders
B30 Private House - BCON
F01 Livestock Farm
C49 Confectioners
U14 School catering
C11 Flowers
S72 Childrens Home
S58 Storage Units
P08 Electricity & Gas Board
F28 Mixed Species Farm
T13 Bus / Train Station
T60 Pleasure Boat
G01 Builders Merchants
A92 Continental Gifts & Food
W99 Other Financial etc
X01 Temporary

Appendix 3.

Classifications used to extract unhealthy outlets from FSA and PoI

```
              |TYPE=="Restaurant/Cafe/Canteen"

              |TYPE=="Retailers - other"

              |TYPE=="Retailers - supermarkets/hypermarkets"

              |TYPE=="Takeaway/sandwich shop")
```

Write to csv file

```
write.csv(unhealthyRCT_FSA, file="unhealthyRCTFSA.csv")
```

```
###########################################################################
###########################################################################
```

Import Swansea FSA data

```
SwanseaFSA <- read.csv("Swansea_fsa.csv",header=TRUE)


unhealthySwansea_FSA <- subset(SwanseaFSA,TYPE=="Other catering premises"

              |TYPE=="Restaurant/Cafe/Canteen"

              |TYPE=="Retailers - other"

              |TYPE=="Retailers - supermarkets/hypermarkets"

              |TYPE=="Takeaway/sandwich shop")
```

Write to csv file

```
write.csv(unhealthySwansea_FSA, file="unhealthySwanseaFSA.csv")
```

```
###########################################################################
###########################################################################
```

Import Torfaen data

```
TorfaenFSA <- read.csv("Torfaen_fsa.csv",header=TRUE)


unhealthyTorfaen_FSA <- subset(TorfaenFSA,TYPE=="Other catering premises"

              |TYPE=="Restaurant/Cafe/Canteen"

              |TYPE=="Retailers - other"

              |TYPE=="Retailers - supermarkets/hypermarkets"

              |TYPE=="Takeaway/sandwich shop")
```

```
write.csv(unhealthyTorfaen_FSA, file="unhealthyTorfaenFSA.csv")
```

###
###

#Import Blaenau Gwent data

```
GwentFSA <- read.csv("BlaenauGwent_fsa.csv",header=TRUE)

unhealthyGwent_FSA <- subset(GwentFSA,TYPE=="Other catering premises"
                |TYPE=="Restaurant/Cafe/Canteen"
                |TYPE=="Retailers - other"
                |TYPE=="Retailers - supermarkets/hypermarkets"
                |TYPE=="Takeaway/sandwich shop")
```

Write to csv file

```
write.csv(unhealthyGwent_FSA, file="unhealthyGwentFSA.csv")
```

###
###

Potential unhealthy outlets were extracted from PoI data using the following R code:

##set working directory

```
setwd("A:/PHD/POI/poi_932950")
```

#Import POI data

```
POI <- read.csv("poi.csv",header=TRUE,sep="|")
```

#Subset Swansea data

```
Swansea <- subset(POI,Administrative.Boundary=="Abertawe - Swansea")

unhealthySwansea_POI <-
subset(Swansea,PointX.Classification.Code=="9470661" #Bakeries
                |PointX.Classification.Code=="9470663" #Confectioners
                |PointX.Classification.Code=="9470699" #Convenience
```

|PointX.Classification.Code=="9470665" **#Delicatessens**

|PointX.Classification.Code=="9470667" **#Frozen foods**

|PointX.Classification.Code=="9470705" **#Markets**

|PointX.Classification.Code=="9470819" **#Supermarket Chains**

|PointX.Classification.Code=="9470798" **#Tea & coffee**
merchants

|PointX.Classification.Code=="1020013" **#Cafes**

|PointX.Classification.Code=="1020018" **#Takeaways & FFO**

|PointX.Classification.Code=="1020019" **#**Fast food delivery
services

|PointX.Classification.Code=="1020020" **#Fish and chip shops**

|PointX.Classification.Code=="1020025") **#Internet cafes**

setwd("A:/PhD/Writing/Chapter 4 LA vs POI and FSA/Analysis")

write.csv(unhealthySwansea_POI, file="unhealthySwanseaPOI.csv")

##

#Subset Newport data

Newport <- subset(POI,Administrative.Boundary=="Casnewydd - Newport")

unhealthyNewport_POI <- subset(Newport,PointX.Classification.Code=="9470661"
#Bakeries

|PointX.Classification.Code=="9470663" **#Confectioners**

|PointX.Classification.Code=="9470699" **#Convenience**

|PointX.Classification.Code=="9470665" **#Delicatessens**

|PointX.Classification.Code=="9470667" **#Frozen foods**

|PointX.Classification.Code=="9470705" **#Markets**

|PointX.Classification.Code=="9470819" **#Supermarket Chains**

|PointX.Classification.Code=="9470798" **#Tea & coffee**
merchants

|PointX.Classification.Code=="1020013" **#Cafes**

|PointX.Classification.Code=="1020018" **#Takeaways & FFO**

|PointX.Classification.Code=="1020019" **#Fast food delivery**
services

```
                       |PointX.Classification.Code=="1020020" #Fish and chip sops
                       |PointX.Classification.Code=="1020025") #Internet cafes


setwd("~/First Year Report/Food Outlet Classifications/POI")

write.csv(unhealthyNewport_POI, file="unhealthyNewportPOI.csv")


###################################################################
##Subset RCT data


RCT <- subset(POI,Administrative.Boundary=="Rhondda Cynon Taf - Rhondda
Cynon Taf")

unhealthyRCT_POI <- subset(RCT,PointX.Classification.Code=="9470661"
#Bakeries
                       |PointX.Classification.Code=="9470663" #Confectioners
                       |PointX.Classification.Code=="9470699" #Convenience
                       |PointX.Classification.Code=="9470665" #Delicatessens
                       |PointX.Classification.Code=="9470667" #Frozen foods
                       |PointX.Classification.Code=="9470705" #Markets
                       |PointX.Classification.Code=="9470819" #Supermarket Chains
                       |PointX.Classification.Code=="9470798" #Tea & coffee
merchants
                       |PointX.Classification.Code=="1020013" #Cafes
                       |PointX.Classification.Code=="1020018" #Takeaways & FFO
                       |PointX.Classification.Code=="1020019" #Fast food delivery
services
                       |PointX.Classification.Code=="1020020" #Fish and chip sops
                       |PointX.Classification.Code=="1020025") #Internet cafes


setwd("~/First Year Report/Food Outlet Classifications/POI")

write.csv(unhealthyRCT_POI, file="unhealthyRCTPOI.csv")


###################################################################
##Subset Torfaen data
```

```r
Torfaen <- subset(POI,Administrative.Boundary=="Tor-faen - Torfaen")

unhealthyTorfaen_POI <- subset(Torfaen,PointX.Classification.Code=="9470661" #Bakeries
                |PointX.Classification.Code=="9470663" #Confectioners
                |PointX.Classification.Code=="9470699" #Convenience
                |PointX.Classification.Code=="9470665" #Delicatessens
                |PointX.Classification.Code=="9470667" #Frozen foods
                |PointX.Classification.Code=="9470705" #Markets
                |PointX.Classification.Code=="9470819" #Supermarket Chains
                |PointX.Classification.Code=="9470798" #Tea & coffee
merchants
                |PointX.Classification.Code=="1020013" #Cafes
                |PointX.Classification.Code=="1020018" #Takeaways & FFO
                |PointX.Classification.Code=="1020019" #Fast food delivery
services
                |PointX.Classification.Code=="1020020" #Fish and chip sops
                |PointX.Classification.Code=="1020025") #Internet cafes


setwd("~/First Year Report/Food Outlet Classifications/POI")

write.csv(unhealthyTorfaen_POI, file="unhealthyTorfaenPOI.csv")


###################################################################
##Subset Gwent data


Gwent <- subset(POI,Administrative.Boundary=="Blaenau Gwent - Blaenau Gwent")

unhealthyGwent_POI <- subset(Gwent,PointX.Classification.Code=="9470661" #Bakeries
                |PointX.Classification.Code=="9470663" #Confectioners
                |PointX.Classification.Code=="9470699" #Convenience
                |PointX.Classification.Code=="9470665" #Delicatessens
                |PointX.Classification.Code=="9470667" #Frozen foods
                |PointX.Classification.Code=="9470705" #Markets
                |PointX.Classification.Code=="9470819" #Supermarket Chains
```

```
                |PointX.Classification.Code=="9470798" #Tea & coffee
merchants

                |PointX.Classification.Code=="1020013" #Cafes

                |PointX.Classification.Code=="1020018" #Takeaways & FFO

                |PointX.Classification.Code=="1020019" #Fast food delivery
services

                |PointX.Classification.Code=="1020020" #Fish and chip sops

                |PointX.Classification.Code=="1020025") #Internet cafes


setwd("~/First Year Report/Food Outlet Classifications/POI")

write.csv(unhealthyGwent_POI, file="unhealthyGwentPOI.csv")


#####################################################################
#####################################################################
```

Appendix 4.

Survey instruments – Study information, travel diary, consent forms

DECIPHer
Interventions for Public Health
Improvement
A UKCRC Public Health Research
Centre of Excellence
Datblygu a Gwerthuso
Ymyriadau Cymhleth er
mwyn Gwella Iechyd. Un o
Ganolfannau Rhagoriaeth
Ymchwil Iechyd y Cyhoedd
UKCRC

Dear Head of Geography,

At Swansea University Medical School we are examining links between child health and the environment that children experience on their route to and from school. The results of this study will be important in providing evidence that is needed for local councils and the government to make changes to provide a healthy local environment for children and adults. Computer software is being used to predict the effect of the environment on children's health along their journeys to and from school. However, we would like to have some real journeys to evaluate and refine our predictions.

We would like to invite your school to participate in the study.

We would like to work with year 9 pupils that usually walk to school and have a smart phone. The study will give your students an introduction to GIS (Geographical Information Systems) which is a technique widely used in health and environment research and both public and private sectors. As well as contributing to the study, participation will also give your students hands on experience of data collection and an appreciation for GIS. It will be of particular use to your students if they are considering a health profession or any geographical related studies. Even if this is not the case, the sessions will give an introduction to data capture and analysis techniques that are valued by employers. In appreciation of participation, your students will be awarded vouchers to The LC Swansea and a healthy snack goodie bag. Furthermore, all children who complete the data collection will be entered in to a draw with a chance to win vouchers to Go Air in Cardiff and 360 café in Swansea.

We would like to record students' journeys to and from school for one week using an app that they are able to download for free on to their phone. This app is called 'My Tracks'. 'My Tracks' uses GPS to record the path, speed and distance of your journey when the app is turned on. In addition to this we would also like the students to complete a travel diary. The GPS routes will be compared with computer-predicted routes to check the accuracy of our estimates.

We suggest that three sessions during a lunchtime or registration time would be appropriate to meet with your students. The first occasion would be to introduce your students to the study and give out consent forms. A second visit to collect the consent forms, help your students install the app on their phone and explain the details of the data collection. The final time would be to collect the data from the students.

We understand that school life is extremely busy for teachers and students and so this is just a proposed method. We are happy to change this and be flexible in working together. Four schools have been approached in Swansea to participate in the study. On first occasion I can offer 20 places for 20 students to participate. However, should any schools decline, more places will become available.

Any personal details that are recorded in the data will be removed. This will ensure that the data cannot be linked to a student. The GPS data will be held on an encrypted computer in a restricted access building on the Swansea University campus. This building is specifically designed to securely hold sensitive data, in that researchers must adhere to strict security procedures. Only four assigned researchers will have access to the data.

After the study finishes, the data must be kept for several years in case the department is audited. All personal data collected will be held for specified and lawful purposes, in accordance with the Data Protection Act 1998. The Data Protection Act protects information from being accessed by unauthorised people. In exceptional and rare situations others could ask for these data, for example, to investigate a crime. However, GPS data require specialist software to reconstruct, and so mobile phone records are more likely to be used instead of the GPS data collected for such an instance.

In this envelope, a letter to parents, a consent form, a letter to students and an example of the travel diary have been included. Should you like more information or would like to meet to discuss the investigation further, please feel free to contact me; my details are below.

I look forward to hearing from you.

Yours Faithfully,

A**my Mizen**
DECIPHer PhD Student | Myfyriwr DECIPHer

_____

Data Science Building (Third Floor) | Yr Adeilad Gwyddor Data (Trudydd Llawr)

Swansea University Medical School | Ysgol Feddygaeth Prifysgol Abertawe

Singleton Park | Parc Singleton

SWANSEA SA2 8PP | ABERTAWE SA2 8PP

Wales, United Kingdom | Cymru, Y Deyrnas Unedig

Phone | Ffôn +44 (0) 1792 60 6279
Email | Ebost amy@chi.swan.ac.uk

DECIPHer

Interventions for Public Health Improvement
A UKCRC Public Health Research Centre of Excellence

Datblygu a Gwerthuso Ymyriadau Cymhleth er mwyn Gwella Iechyd. Un o Ganolfannau Rhagoriaeth Ymchwil Iechyd y Cyhoedd UKCRC

Dear Student,

I'm a researcher at Swansea University and I need your help with an investigation. I am investigating whether the environment young people walk through on their way to and from school affects their health. I have used computer software to make predictions about how the environment may be affecting young people. To see how good my predictions are, I need some real routes to compare to my results.

If you decide to join the experiment, you need to have a smart phone and walk to school. As part of the experiment, you will download a free app on to your phone. If you have an android phone the app is called 'Track My Trip' or if you have an iPhone, the app is called 'Easy Trails'. Just before you leave the house for school you'll need to open the app and press start. When you arrive at school you'll have to press stop. Whilst the app is running, your phone will record your journey to school. In the afternoon, before you leave the school grounds you'll have to remember to start recording your journey again and then stop the recording when you get home. If you are happy to provide your phone number, a text can be sent to you in the morning and afternoon to remind you to record your journey. After the experiment, any record of your phone number will be deleted. You will also have a travel diary to fill out at the end of every day - this will only take you five minutes.

Your name and address will be removed from your data so that nobody can see that you have participated in the study. The data of your journeys will be kept on an encrypted computer and your questionnaires will be kept in a locked cupboard. Once my project has finished we will have to keep your data for several years in case the department is reviewed. But your data will be kept safe and secure.

I hope that you will get involved in helping me with my project. The results of this study will be important in providing evidence to local councils and the government that is needed to make changes in order to provide a healthy environment for you. For participating in the study you will be given a healthy snack goodie bag. If you complete the data collection you will be entered in to a draw with a chance to win vouchers for Go Air in Cardiff.

If you have any questions before you sign up to the investigation, please feel free to contact me on the details below.

Kind Regards,

Amy Mizen

Data Science Building (Third Floor)
Swansea University Medical School
Singleton Park
SWANSEA SA2 8PP

Yr Adeilad Gwyddor (Trydydd Llawr)
Ysgol Feddygaeth Prifysgol Abertawe
Parc Singleton
ABERTAWE SA2 8PP

T +44(0)1792 60 6279
E amy@chi.swan.ac.uk
www.decipher.uk.net
Follow us on Dilynwch ni ar Twitter: @DECIPHerCentre

Dear Parent/Guardian,

At Swansea University Medical School we are examining links between child health and the environment that children experience on their route to and from school. Computer software is being used to predict the effect of the environment on children's health along their journeys to and from school. We would like to record some real journeys to see how good our predictions are.

We would like to record your child's journey to and from school for 5 days (June 6th – June 10th 2016) using an app that they are able to download for free on to their phone. In addition to this we would also like your child to complete a travel diary. All personal details will be removed from the data that we record on your child so that nobody can see that your child has participated in the study. For android phones, the app is called 'Track My Trip' and for Apple phones the app is called 'Easy Trails'. These apps use GPS to record the path, speed and distance of your journey when the app is turned on. There are no charges to use the app for the purpose of our experiment. However, please note that there are add-ins available to buy with the apps which would incur data charges if these are purchased. For the purpose of our investigation the app is free to use.

If you are happy for your child to provide their mobile number, a text can be sent out in the morning and afternoon to remind them to turn the app on. Once the experiment is over, all mobile phone numbers will be immediately deleted.

Your child's name and any personal details that are recorded in the data will be removed. This will ensure that the data cannot be linked to your child. The GPS data will be held on an encrypted computer in a restricted access building on the Swansea University campus. This building is specifically designed to hold sensitive data, in that researchers must adhere to strict security procedures. Only three assigned researchers will have access to the data and we are specially trained to deal with sensitive data.

Once all personal details have been removed from the GPS routes, they will be compared with computer-predicted routes to check the accuracy of the estimates. The results of this study will be important in providing evidence to the government that is needed to make changes in order to provide a healthy local environment for children and adults.

After the study finishes, the data must be kept for several years in case the department is audited. All personal data collected will be held for specified and lawful purposes, in accordance with the Data Protection Act 1998. The Data Protection Act protects information from being accessed by unauthorised people.

We hope that you will allow your child to participate in this survey. The sessions will give an introduction to data capture and analysis techniques that are valued by the private and public sectors. In appreciation of participation, your child will be awarded vouchers to The LC Swansea and a healthy snack goodie bag. Furthermore, all children who complete the data collection will be entered in to a draw with a chance to win vouchers for Go Air in Cardiff and 360 café in Swansea.

Yours Faithfully,

Amy Mizen

DECIPHer

Interventions for Public Health
Improvement
A UKCRC Public Health Research
Centre of Excellence

Datblygu a Gwerthuso
Ymyriadau Cymhleth er
mwyn Gwella Iechyd. Un o
Ganolfannau Rhagoriaeth
Ymchwil Iechyd y Cyhoedd
UKCRC

Dear Headteacher,

At Swansea University Medical School we are examining links between child health and the environment that children experience on their route to and from school.

We would like to invite your school to participate in the study.

We would like to work with year 9 pupils that usually walk to school and have a smart phone. The study will give your students an introduction to GIS (Geographical Information Systems) which is a technique widely used in health and environment research and both the public and private sectors. As well as contributing to the study, participation will give your students hands on experience of data collection and an appreciation for GIS which is on the Geography curriculum for KS3 and 4. Furthermore, the results of this study will be important in providing evidence that is needed to local councils and the government to make changes to provide a healthy local environment for children and adults.

We are happy to provide a report of your pupil's results that would be beneficial to your involvement with the Healthy Schools scheme that is promoted by Welsh Government.

I have left further information with your Head of Geography. If you would like to know more about the study or have any questions, please see my contact details below.

I look forward to your response.

Yours Faithfully,

Amy Mizen
DECIPHer PhD Student | Myfyriwr DECIPHer

Route To School Project Consent Form

THIS IS CONFIDENTIAL INFORMATION THAT WILL BE SECURELY STORED

PARTICIPANT'S NAME:…………………………………………………………………………………………….

PARTICIPANT'S ADDRESS:…………………………………………………………………………………………….

……………………………………………………………………………………………..

PARTICIPANT'S DATE OF BIRTH: ………………………………………………………….

PARTICIPANT'S SCHOOL………………………………………………………………….

To be completed by participant's **PARENT/GUARDIAN**:

Please answer the following questions if **you agree** to your child taking part in the project.

Please tick box as appropriate

I **AGREE** to my child taking part in the Route to School project ☐

My child and I understand that participation is entirely voluntary and that my child is free to withdraw at any time, and without giving a reason ☐

My child and I have read and understood the information sheet ☐

My child and I understand what this project is about ☐

My child and I have asked all the questions that we want to ☐

All of our questions have been answered in a way that we understand ☐

I am happy for my child to complete the travel diary ☐

I am happy for my child to provide their mobile number to receive reminders ☐

I am happy for my child to record their route to school for five days during June 2016 on the 'Track My Trip' app ☐

Signature…………………………………………………………………….(parent)

Date:………………………………………….

To be completed by the **participant** with the parent/guardian present

Participant to **circle** as appropriate

Have you read the information letter?	Yes / No
Has somebody explained the travel diary and app to you?	Yes / No
Do you understand what this project is about?	Yes / No
Have you asked the questions you want?	Yes / No
Have your questions been answered in a way that you understand?	Yes / No
Do you understand that it's okay to stop taking part at any time?	Yes / No
Are you happy to fill out the travel diary?	Yes / No
Are you happy to record your journey to school and home from school with the 'Track my Trip'/ 'Easy Trails' app?	Yes / No

If you are <u>happy</u> to take part in this project please write your name below

Full name:…………………………………….

Signature……………………………………………………………….(pupil)

Date:………………………………………….

PLEASE RETURN THIS FORM TO YOUR CHILD'S GEOGRAPHY TEACHER

Please check that you have answered **all** questions. A copy of these forms will be returned to you. **THANK YOU!**

Contact Us

If you would like to speak with us about the travel diary or any aspect of the study, please do not hesitate to contact:

[Street Address]
[City, ST ZIP Code]

Phone: [01792 606279]
Email: [Amy@chi.swan.ac.uk]

DECIPHer, the Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement, is one of five UKCRC Public Health Research Centres of Excellence coordinated by the Medical Research Council.

DECIPHer is a partnership between Cardiff, Bristol and Swansea Universies.

Swansea University

Data Science Building
Swansea University Medical School
Singleton Park
SWANSEA SA2 8PP

Swansea University
Prifysgol Abertawe

Swansea
Prifysgol

Travel to

Travel

Student ID:

Please complete this diary fo

13/ 06

Amy Mizen	Dr Sar Rodge
DECIPHER PhD Student	Associate Pro Epidemi
	Data Science
	Swansea University
	Singleton
	SWAN
	SA2 8
	Wale
	United Kin

Instructions for use

The purpose of this diary is to help us understand how to travel to and from school.

To keep your data confidential, please do not put your name on your booklet, instead put your pupil ID on the front page. If you don't know your pupil ID then you can ask your teacher.

On page 2, we would like to know about your travel to school in the last four weeks.

The rest of the diary is about the next four days. Please complete pages 4-10 of the diary for four days in a row, starting on the date we have written on the front cover. If you are ill and do not go to school, there is a box to tick and then you can leave the rest of the page blank.

For each school day we ask:

1. How you travelled to school
2. How long you spent travelling
3. How you travelled home from school
4. How long you spent getting home
5. Did you buy any food or drink on your way to school
6. Did you buy any food or drink on your way home from school
7. What kind of food or drink you bought.

When you have completed your diary, please bring it back to school with you.

Did you buy food/drink on your way
answer that applies)

Yes ☐   No ☐

Did you buy food/drink on your way
(Please **tick** the answer that applies)

Yes ☐   No ☐

If you did stop what kind of food/drin
the food and drink that applies)

Burger	**Energy Dr**
Cake	**Fizzy Drin**
Cereal/Fruit Bar	**Fried Chic**
Chinese/Indian takeaway	**Fruit jui**
Chips	**Fruit/ Veget**
Chocolate/Biscuits	**Ice Crea**
Coffee/Tea	**Milkshak**
Crisps	**Pasties/Sau Rolls**

DAY 4

Date: Day ____ Month____ Year____

☐ Tick this box if you did <u>not</u> go to school today and leave the rest blank

How did you travel **to school** today? (please tick all the boxes that apply)		How long did you spend travelling **to school**? (e.g. if you spent 10 minutes walking and 20 minutes on a bus, enter '10' on row (a) and '20' on row (c)
a. Walked		Minutes
b. Cycled		Minutes
c. Bus		Minutes
d. Train		Minutes
e. Car/Taxi		Minutes
f. Skateboard/Scooter		Minutes
g. Other		Minutes
How did you travel home **from school** today? (please tick all the boxes that apply)		How long did you spend travelling home **from school**? (e.g. if you spent 10 minutes walking and 20 minutes on a bus, enter '10' on row (a) and '20' on row (c)
a. Walked		Minutes
b. Cycled		Minutes
c. Bus		Minutes
d. Train		Minutes
e. Car/Taxi		Minutes
f. Skateboard/Scooter		Minutes
g. Other		Minutes

Travelling to and from s
weeks

1. What is the approximate d
   and your school?

2. In the last **four weeks**, how
   between your home and y
   box PER ROW)

	Always or nearly always	
a.Walked		
b.Cycled		
c.Bus		
d.Train		
e.Car/Taxi		
f.Skateboard/ Scooter		
g.Other		

3. What is the postcode for y
   postcode e.g. SA10 4QR, in

4. What is the postcode for y
   full postcode e.g. SA10 4Q

DAY 1

Date: Day _____ Month_____ Year_____

[ ] Tick this box if you did <u>not</u> go to school today and leave the rest blank

How did you travel **to school** today? (please tick all the boxes that apply)		How long did you spend travelling **to school**? (e.g. if you spent 10 minutes walking and 20 minutes on a bus, enter '10' on row (a) and '20' on row (c)
a. Walked		Minutes
b. Cycled		Minutes
c. Bus		Minutes
d. Train		Minutes
e. Car/Taxi		Minutes
f. Skateboard/Scooter		Minutes
g. Other		Minutes
How did you travel home **from school** today? (please tick all the boxes that apply)		How long did you spend travelling home **from school**? (e.g. if you spent 10 minutes walking and 20 minutes on a bus, enter '10' on row (a) and '20' on row (c)
a. Walked		Minutes
b. Cycled		Minutes
c. Bus		Minutes
d. Train		Minutes
e. Car/Taxi		Minutes
f. Skateboard/Scooter		Minutes
g. Other		Minutes

Did you buy food/drink on your way answer that applies)

Yes [ ]   No [ ]

Did you buy food/drink on your way (Please **tick** the answer that applies)

Yes [ ]   No [ ]

If you did stop what kind of food/drin *the food and drink that applies*)

Burger	Energy Dr
Cake	**Fizzy Drin**
Cereal/Fruit Bar	**Fried Chic**
Chinese/Indian takeaway	**Fruit jui**
Chips	**Fruit/ Veget**
Chocolate/Biscuits	**Ice Crea**
Coffee/Tea	**Milkshal**
Crisps	**Pasties/Sau Rolls**

DAY 3

Date: Day ____ Month____ Year____

☐ Tick this box if you did <u>not</u> go to school today and leave the rest blank

How did you travel **to school** today? (please tick all the boxes that apply)		How long did you spend travelling **to school**? (e.g. if you spent 10 minutes walking and 20 minutes on a bus, enter '10' on row (a) and '20' on row (c)
a.	Walked	Minutes
b.	Cycled	Minutes
c.	Bus	Minutes
d.	Train	Minutes
e.	Car/Taxi	Minutes
f.	Skateboard/Scooter	Minutes
g.	Other	Minutes
How did you travel home **from school** today? (please tick all the boxes that apply)		How long did you spend travelling home **from school**? (e.g. if you spent 10 minutes walking and 20 minutes on a bus, enter '10' on row (a) and '20' on row (c)
a.	Walked	Minutes
b.	Cycled	Minutes
c.	Bus	Minutes
d.	Train	Minutes
e.	Car/Taxi	Minutes
f.	Skateboard/Scooter	Minutes
g.	Other	Minutes

Did you buy food/drink on your way answer that applies)

Yes ☐   No ☐

Did you buy food/drink on your way (Please **tick** the answer that applies)

Yes ☐   No ☐

If you did stop what kind of food/drink *the food and drink that applies*)

Burger	**Energy Dr**
Cake	**Fizzy Drir**
Cereal/Fruit Bar	**Fried Chic**
Chinese/Indian takeaway	**Fruit jui**
Chips	**Fruit/ Veget**
Chocolate/Biscuits	**Ice Crea**
Coffee/Tea	**Milkshak**
Crisps	**Pasties/Sau Rolls**

226

DAY 2

Date: Day _____ Month_____ Year_____

☐ Tick this box if you did <u>not</u> go to school today and leave the rest blank

How did you travel **to school** today? (please tick all the boxes that apply)		How long did you spend travelling **to school**? (e.g. if you spent 10 minutes walking and 20 minutes on a bus, enter '10' on row (a) and '20' on row (c)
a. Walked		Minutes
b. Cycled		Minutes
c. Bus		Minutes
d. Train		Minutes
e. Car/Taxi		Minutes
f. Skateboard/Scooter		Minutes
g. Other		Minutes
How did you travel home **from school** today? (please tick all the boxes that apply)		How long did you spend travelling home **from school**? (e.g. if you spent 10 minutes walking and 20 minutes on a bus, enter '10' on row (a) and '20' on row (c)
a. Walked		Minutes
b. Cycled		Minutes
c. Bus		Minutes
d. Train		Minutes
e. Car/Taxi		Minutes
f. Skateboard/Scooter		Minutes
g. Other		Minutes

Did you buy food/drink on your way answer that applies)

Yes ☐   No ☐

Did you buy food/drink on your way (Please **tick** the answer that applies)

Yes ☐   No ☐

If you did stop what kind of food/drin *the food and drink that applies*)

Burger	**Energy Dr**
Cake	**Fizzy Drin**
Cereal/Fruit Bar	**Fried Chic**
Chinese/Indian takeaway	**Fruit jui**
Chips	**Fruit/ Vegetab**
Chocolate/Biscuits	**Ice Crea**
Coffee/Tea	**Milkshak**
Crisps	**Pasties/Sau Rolls**

Appendix 5.

The Physical Environment and Activity Relationships (in adolescents) (PEAR) Study

The aim of the PEAR project was "to investigate how different features of the physical environment relate to the time that 13-15 year olds spend engaged in physical activity" [403]. The project built on work that had been undertaken in the PEACH project [274, 277].

The study used objective measures to see how active children are in relation to where they are. These measures were: Objective measures of the environment (GIS); Objective measures of where people go (GPS); Accelerometers were used as an objective measure of physical activity. The study also used subjective measures of physical, social and local environmental determinants of physical activity and use of local environment [455].

The study examined physical activity and other lifestyle behaviours in relation to obesity and pubertal status and also in relation to urban/rural morphologies. Stress levels were measured on a weekday in relation to time spent outside.

The PEAR study was an MRC funded project. It is a cross-sectional study of year 9 students (aged 13-15 years) from Bristol, South Gloucestershire, North Somerset and Bath and North East Somerset. The study was run by Bristol University in partnership with University of East Anglia, Queen Mary University and Cardiff University.

Appendix 6.

Automated cleaning of GPS data

--Create a table just with weekday journeys

```
CREATE TABLE afternoon_routes.school_routes AS

SELECT * FROM all_routes.pear_routes WHERE day IN (1,2,3,4,5)
```

```
----------------------------STEP---------------------------------
------------------------------1-----------------------------------
------Remove gps points of participants with no home location--------
--------------------------------------------------------------------
```

--Cast "pupilid" column that is text to numeric type

```
ALTER TABLE afternoon_routes.school_routes ALTER COLUMN pupilid TYPE
numeric USING (pupilid::numeric);
```

```
DELETE FROM afternoon_routes.school_routes WHERE pupilid IN
(9051040,9011008,9051072,9011038,9061018);
```

--Make a copy for other methods

```
CREATE TABLE afternoon_routes.school_routes_original As

SELECT * FROM afternoon_routes.school_routes
```

```
----------------------------STEP---------------------------------
------------------------------2-----------------------------------
---------Remove gps points not during school commuting times---------
--------------------------------------------------------------------
```

```
DELETE FROM afternoon_routes.school_routes WHERE timetxt < '14:30:00'

DELETE FROM afternoon_routes.school_routes WHERE timetxt > '17:30:00'
```

```
----------------------------STEP---------------------------------
------------------------------2-----------------------------------
--------------Remove gps points inside (sumsnr <= 250)---------------
```

--DELETE FROM afternoon_routes.school_routes WHERE sumsnr <= 249;

-----------------------------STEP----------------------------------

-----------------------------3------------------------------------

--------------Remove gps points inside school footprint--------------

----

--give unique id to rows for next stage of indexing

ALTER TABLE afternoon_routes.school_routes ADD COLUMN unique_id SERIAL PRIMARY KEY;

--add geometry column to table

SELECT AddGeometryColumn('afternoon_routes','school_routes','geom',27700,'POINT',2);

UPDATE afternoon_routes.school_routes SET geom = ST_GeomFromText('POINT(' || school_routes.easting || ' ' || school_routes.northing || ')', 27700);

--Convert geom to xy coords for quicker processing

--gps route geom

ALTER TABLE afternoon_routes.school_routes ADD COLUMN xy_geom text

UPDATE afternoon_routes.school_routes SET xy_geom = st_astext(school_routes.geom)

--Import OS Sites Pear school data as 'all_routes.pear_schools_poly'

--Convert geom to xy coords for quicker processing

--gps route geom

CREATE TABLE afternoon_routes.school_gps_intersec AS

SELECT school_routes.unique_id,school_routes.geom, ST_Intersects(school_routes.geom,pear_schools_poly.geom) as intersect

FROM afternoon_routes.school_routes, clean_b.pear_schools_poly;

CREATE TABLE afternoon_routes.just_school_intersec_points AS

SELECT * FROM afternoon_routes.school_gps_intersec WHERE
school_gps_intersec.intersect = 't';

--ALTER TABLE afternoon_routes.just_school_intersec_points ADD COLUMN
xy_geom text

--UPDATE afternoon_routes.just_school_intersec_points SET xy_geom =
st_astext(just_school_intersec_points.geom)

DELETE FROM afternoon_routes.school_routes

USING afternoon_routes.just_school_intersec_points

WHERE school_routes.unique_id = just_school_intersec_points.unique_id;

-----------------------------STEP----------------------------------

-----------------------------4------------------------------------

-----------------Remove points within 40m of home--------------------

----

CREATE TABLE afternoon_routes.just_home_intersec_points AS

SELECT * FROM all_routes.just_home_intersec_points

CREATE TABLE afternoon_routes.home_gps_intersec AS

SELECT school_routes.unique_id, school_routes.pupilid, school_routes.geom,
ST_Intersects(school_routes.geom,pear_homes_buffer.geom) as intersect

FROM afternoon_routes.school_routes, all_routes.pear_homes_buffer

WHERE school_routes.pupilid = pear_homes_buffer.pear_id;

--Convert geom to xy coords for quicker processing

--gps route geom

ALTER TABLE afternoon_routes.just_home_intersec_points ADD COLUMN
xy_geom text

```
UPDATE afternoon_routes.just_home_intersec_points SET xy_geom =
st_astext(just_home_intersec_points.geom)


DELETE FROM afternoon_routes.school_routes

USING afternoon_routes.just_home_intersec_points

WHERE school_routes.xy_geom = just_home_intersec_points.xy_geom;
```

-----------------------------------ADD IN INDEXES--

```
--UPDATE unique id

ALTER TABLE afternoon_routes.school_routes DROP COLUMN unique_id

ALTER TABLE afternoon_routes.school_routes ADD COLUMN unique_id SERIAL
PRIMARY KEY;



--Create route_n index

ALTER TABLE afternoon_routes.school_routes ADD COLUMN route_n integer;


UPDATE afternoon_routes.school_routes SET route_n = v_table_name.rn

FROM

(

    SELECT dense_rank() OVER (order by pupilid,day) AS rn,unique_id as id

FROM afternoon_routes.school_routes


) AS v_table_name

WHERE school_routes.unique_id = v_table_name.id;


--Create route sequence index

ALTER TABLE afternoon_routes.school_routes ADD COLUMN route_seq integer;
```

```sql
UPDATE afternoon_routes.school_routes SET route_seq = v_table_name.rn

FROM

(

    SELECT row_number() OVER (partition by route_n order by timetxt) AS
rn,unique_id as id

FROM afternoon_routes.school_routes


) AS v_table_name

WHERE school_routes.unique_id = v_table_name.id


--Create copy of school_routes table

CREATE TABLE afternoon_routes.school_routes_basic_clean AS

SELECT * FROM afternoon_routes.school_routes
```

--Turn gps points in to line geom--

```sql
CREATE TABLE afternoon_routes.gps_lines AS

SELECT route_n,ST_MakeLine(school_routes.geom ORDER BY timetxt) As
gps_line

        FROM afternoon_routes.school_routes

        GROUP BY route_n;


--add geometry definition

ALTER TABLE afternoon_routes.gps_lines ALTER COLUMN gps_line  TYPE
geometry(linestring,27700) USING (gps_line::geometry);
```

---Clean start points of afternoon routes--

```sql
--Add school_id
ALTER TABLE afternoon_routes.school_routes ADD COLUMN school_id bigint;


UPDATE afternoon_routes.school_routes SET school_id = pear_homes.school_id
FROM all_routes.pear_homes
WHERE school_routes.pupilid = pear_homes.pear_id


--Select last gps point in school
CREATE TABLE afternoon_routes.school_intersec AS
SELECT
pupilid,datetxt,timetxt,day,weekend,speed,pdop,sumsnr,northing,easting,dist_home,
in_neigh,xy_geom,school_routes.geom,route_n,unique_id,route_seq,
        ST_DWithin(school_routes.geom,pear_schools_poly.geom,20) as intersect
FROM afternoon_routes.school_routes, clean_b.pear_schools_poly
WHERE school_routes.school_id = pear_schools_poly.school_id


CREATE TABLE afternoon_routes.school_intersec_t AS
SELECT * FROM afternoon_routes.school_intersec
WHERE "intersect" = 't'


CREATE TABLE afternoon_routes.afternoon_start_pt AS
SELECT route_n,max(timetxt) FROM
        (SELECT * FROM afternoon_routes.school_intersec_t) as foo
GROUP BY route_n
ORDER BY route_n


ALTER TABLE afternoon_routes.afternoon_start_pt ADD COLUMN geom geometry


UPDATE afternoon_routes.afternoon_start_pt SET geom = school_routes.geom
FROM afternoon_routes.school_routes
WHERE afternoon_start_pt.route_n = school_routes.route_n
AND afternoon_start_pt.max = school_routes.timetxt
```

ALTER TABLE afternoon_routes.afternoon_start_pt ALTER COLUMN geom TYPE geometry(point,27700) USING (geom::geometry);

--Delete everything before last point at school

CREATE TABLE afternoon_routes.school_gps_points AS

SELECT school_routes.pupilid,school_routes.route_n,school_routes.timetxt,school_routes.geom,school_routes.unique_id,school_routes.route_seq,school_routes.school_id

FROM afternoon_routes.school_routes, afternoon_routes.afternoon_start_pt

WHERE school_routes.timetxt < afternoon_start_pt.max

AND school_routes.route_n = afternoon_start_pt.route_n

ORDER BY route_n,timetxt

DELETE FROM afternoon_routes.school_routes

USING afternoon_routes.school_gps_points

WHERE school_routes.unique_id = school_gps_points.unique_id

--TEST and make lines

--Turn gps points in to line geom

CREATE TABLE afternoon_routes.gps_linesb AS

SELECT route_n,ST_MakeLine(school_routes.geom ORDER BY timetxt) As gps_line

      FROM afternoon_routes.school_routes

      GROUP BY route_n;

ALTER TABLE afternoon_routes.gps_linesb ALTER COLUMN gps_line TYPE geometry(linestring,27700) USING (gps_line::geometry);

-------------------------------Clean end points of afternoon route--------------------------------

----

CREATE TABLE afternoon_routes.school_routes_school_clean AS

```
SELECT * FROM afternoon_routes.school_routes


--Select last gps point within 20m of home
CREATE TABLE afternoon_routes.afternoon_end AS
SELECT school_routes.pupilid,school_routes.datetxt,school_routes.timetxt,
        school_routes.day,school_routes.weekend,school_routes.speed,school_rout
es.pdop,
        school_routes.sumsnr,school_routes.northing,school_routes.easting,school_
routes.dist_home,
        school_routes.in_neigh,school_routes.xy_geom,school_routes.geom,
        school_routes.route_n,school_routes.unique_id,school_routes.route_seq,
        ST_Intersects(school_routes.geom,pear_homes_buffer_20.geom) as
intersect
FROM afternoon_routes.school_routes, clean_c.pear_homes_buffer_20
WHERE school_routes.pupilid = pear_homes_buffer_20.pear_id


CREATE TABLE afternoon_routes.home_intersec_t AS
SELECT * FROM afternoon_routes.afternoon_end
WHERE "intersect" = 't'


CREATE TABLE afternoon_routes.afternoon_end_pt AS
SELECT route_n,min(timetxt) FROM
        (SELECT * FROM afternoon_routes.home_intersec_t) as foo
GROUP BY route_n
ORDER BY route_n


ALTER TABLE afternoon_routes.afternoon_end_pt ADD COLUMN geom geometry


UPDATE afternoon_routes.afternoon_end_pt SET geom = school_routes.geom
FROM afternoon_routes.school_routes
WHERE afternoon_end_pt.route_n = school_routes.route_n
AND afternoon_end_pt.min = school_routes.timetxt
```

ALTER TABLE afternoon_routes.afternoon_end_pt ALTER COLUMN geom TYPE geometry(point,27700) USING (geom::geometry);

--Delete everything before last point at home

CREATE TABLE afternoon_routes.home_gps_points AS

SELECT school_routes.pupilid,school_routes.route_n,school_routes.timetxt,school_routes.geom,school_routes.unique_id,school_routes.route_seq,school_routes.school_id

FROM afternoon_routes.school_routes, afternoon_routes.afternoon_end_pt

WHERE school_routes.timetxt > afternoon_end_pt.min

AND school_routes.route_n = afternoon_end_pt.route_n

ORDER BY route_n,timetxt

DELETE FROM afternoon_routes.school_routes

USING afternoon_routes.home_gps_points

WHERE school_routes.unique_id = home_gps_points.unique_id

--TEST and make lines

--Turn gps points in to line geom

CREATE TABLE afternoon_routes.gps_linesc AS

SELECT route_n,ST_MakeLine(school_routes.geom ORDER BY timetxt) As gps_line

        FROM afternoon_routes.school_routes

        GROUP BY route_n;

ALTER TABLE afternoon_routes.gps_linesc ALTER COLUMN gps_line TYPE geometry(linestring,27700) USING (gps_line::geometry);

CREATE TABLE afternoon_routes.school_routes_home_clean AS

SELECT * FROM afternoon_routes.school_routes


```sql
--Average speed
CREATE TABLE afternoon_routes.av_speed AS
SELECT route_n, avg(speed) FROM afternoon_routes.school_routes
GROUP BY route_n


CREATE TABLE afternoon_routes.driving_routes AS
SELECT * FROM afternoon_routes.av_speed
WHERE avg > 10
ORDER BY route_n


DELETE FROM afternoon_routes.school_routes
USING afternoon_routes.driving_routes
WHERE school_routes.route_n = driving_routes.route_n


--Allocate average col
--CREATE TABLE afternoon_routes.gps_linesb AS
--SELECT * FROM afternoon_routes.gps_lines
--ALTER TABLE afternoon_routes.gps_linesb ADD COLUMN av_speed numeric;
--UPDATE afternoon_routes.gps_linesb SET av_speed = av_speed.avg
--FROM afternoon_routes.av_speed
--WHERE gps_linesb.route_n = av_speed.route_n


--TEST and make lines
--Turn gps points in to line geom
CREATE TABLE afternoon_routes.gps_linesd AS
SELECT route_n,ST_MakeLine(school_routes.geom ORDER BY timetxt) As gps_line
        FROM afternoon_routes.school_routes
```

GROUP BY route_n;

ALTER TABLE afternoon_routes.gps_linesd ALTER COLUMN gps_line TYPE geometry(linestring,27700) USING (gps_line::geometry);

--Manual Cleaning--

--delete routes that have points > 40km/h

DELETE FROM afternoon_routes.school_routes WHERE route_n IN (9,17,18,20,36,33,34,49,56,54,71,77,82,86,113,109,119,121,122,126,129,138,139,140…

DELETE FROM afternoon_routes.school_routes WHERE unique_id IN (926746,926747,926748,926749,926750,926751,926752,926753,926754,926755,926756…

Appendix 7.

Creating a routable network.

The OSM road data was converted in to a network that could be used in network analysis using the pgRouting tools. Instructions of how to create a routable network were obtained from the pgRouting manual [456]. The vertices contained within the road network of OSM did not contain any topology information associated with them. The OSM road data was "noded" to allow network analysis. This means that where two or more vertices formed an intersection, the road vertices were broken at that intersection and node placed to represent a junction. Graph analysis functions were used to identify any topological problems in the data. The code below was used to node the OSM road data and remove

```
-- create table called network and populate it with all of the highways from the
planet_osm_line table.

CREATE TABLE network AS

SELECT * FROM planet_osm_line WHERE highway IS NOT NULL;


--Node the lines. This create network_noded

SELECT pgr_nodeNetwork('network', 0.00001,'osm_id','way');


--Create vertices table containing source and target columns

SELECT pgr_createTopology('network_noded',0.00001, 'way', 'id','source', 'target')


--Analyse new table with pgr_analyzeGraph

SELECT pgr_analyzeGraph('network_noded',0.00001,'way','id')


--Identify the isolated segments in the network

SELECT *

    FROM network_noded a, network_noded_vertices_pgr b,
network_noded_vertices_pgr c

    WHERE a.source=b.id AND b.cnt=1 AND a.target=c.id AND c.cnt=1;
```


-- So that the raw data is not changed, a new scheme was created called model_inputs and the road_network_noded and

-- road_network_noded_vertices_pgr were copied to the new schema

--Create a new schema with road edges and nodes from public schema

CREATE TABLE model_inputs.road_network_noded AS

SELECT * FROM public.network_noded;

CREATE TABLE model_inputs.road_network_noded_vertices_pgr AS

SELECT * FROM public.network_noded_vertices_pgr;

--Select isolated edges and compare with isolated edges from public schema

SELECT *

   FROM model_inputs.road_network_noded a,
model_inputs.road_network_noded_vertices_pgr b,
model_inputs.road_network_noded_vertices_pgr c

   WHERE a.source=b.id AND b.cnt=1 AND a.target=c.id AND c.cnt=1

-- Create a table of isolated edges of old_id from original line table, road_network

CREATE TABLE model_inputs.isolated AS

SELECT old_id FROM

(SELECT *

   FROM model_inputs.road_network_noded a,
model_inputs.road_network_noded_vertices_pgr b,
model_inputs.road_network_noded_vertices_pgr c

   WHERE a.source=b.id AND b.cnt=1 AND a.target=c.id AND c.cnt=1) AS foo;

--Delete isolated nodes from road_network_noded

--DELETE FROM model_inputs.road_network_noded

--WHERE road_network_noded.old_id = model_inputs.isolated.old_id

--Delete

DELETE FROM model_inputs.road_network_noded

WHERE road_network_noded.id IN (282273,283051…

-- The road_network_noded_vertices_pgr was then deleted so that the topology
could be recreated after the isolated edges were removed

--recreate topology

SELECT pgr_createTopology('model_inputs.road_network_noded',0.00001, 'way',
'id','source', 'target')

Appendix 8.

Example code of how environmental characteristics were calculated for GPS routes. The same code was implemented to calculate environmental characteristics for the GIS generated shortest network routes

```
---------------------------------------------------------------------------------------------------------
-----------------------------------------MORNING ROUTES-----------------------------------------
----------------------------------------------------------------------------------------------------------

-----------------------------Outcome--------------------------
--------------------------------1---------------------------------
----------------------Calculate length of route---------------

--------------------------------------------------------------------
--Create the lines for each route
CREATE TABLE morning_routes.route_lines AS
SELECT * FROM morning_routes.cleaned_morning_routes_w_home


--Calculate length
CREATE TABLE morning_routes.route_length AS
SELECT route_n,st_length(route_lines.geom) as length
FROM morning_routes.route_lines


--Delete * routes that do not have a length
DELETE FROM morning_routes.route_length
WHERE length = 0


--Create a route characteristics table
CREATE TABLE morning_routes.route_characteristics AS
SELECT * FROM morning_routes.route_length


SELECT * FROM morning_routes.greenspace
-----------------------------Outcome--------------------------
--------------------------------2---------------------------------
----Calculate percentage greenspace along route-----
```

CREATE TABLE logistic_regression_outcomes_with_fp.green_buffer_union AS

SELECT st_union(green_buffer.geom)

FROM logistic_regression_outcomes_with_fp.green_buffer

ALTER TABLE logistic_regression_outcomes_with_fp.green_buffer_union ALTER COLUMN st_union TYPE geometry(MultiPolygon,27700) USING (st_union::geometry);

CREATE TABLE morning_routes.green_buffer AS

SELECT * FROM logistic_regression_outcomes_with_fp.green_buffer_union

--Create table that contains the length of the the route that is within 25m of greenspace

CREATE TABLE morning_routes.greenspace AS

SELECT route_n,st_length(st_intersection(route_lines.geom,green_buffer.st_union)) as greenspace_length

FROM morning_routes.route_lines,morning_routes.green_buffer

--Add column to route_characteristics called greenspace(%)

ALTER TABLE morning_routes.greenspace ADD greenspace_pct double precision;

UPDATE morning_routes.greenspace SET greenspace_pct = ((greenspace.greenspace_length::double precision / route_characteristics.length::double precision) * 100)

FROM morning_routes.route_characteristics

WHERE greenspace.route_n = route_characteristics.route_n

------------------------------Outcome-------------------------------

--------------------------------3--------------------------------------

-----------Calculate percentage bluespace along route-----

CREATE TABLE morning_routes.bluespace_buffer AS

```sql
SELECT * FROM logistic_regression_outcomes_with_fp.bluespace_unioned


--Create table that contains the length of the the route that is within 25m of
bluespace

CREATE TABLE morning_routes.bluespace_along_route AS

SELECT
route_n,st_length(st_intersection(route_lines.geom,bluespace_buffer.st_union)) as
bluespacespace_length

FROM morning_routes.route_lines,morning_routes.bluespace_buffer




--Add column to route_characteristics called bluespace(%)

ALTER TABLE morning_routes.bluespace_along_route ADD bluespace_pct double
precision;


UPDATE morning_routes.bluespace_along_route SET bluespace_pct =
((bluespace_along_route.bluespacespace_length::double precision /
route_characteristics.length::double precision) * 100)

FROM morning_routes.route_characteristics

WHERE bluespace_along_route.route_n=route_characteristics.route_n




----------------------------Outcome------------------------------

--------------------------------4----------------------------------

-----Calculate number of traffic lights along route-----------

---------------------------------------------------------------------

--Add column to route_characteristics table called traffic_lights


CREATE TABLE morning_routes.traffic_light_count AS

SELECT route_lines.route_n,count(traffic_lights_buffer.*) AS traffic_light_count
FROM
morning_routes.route_lines,logistic_regression_outcomes_with_fp.traffic_lights_buff
er

   WHERE ST_DWithin(route_lines.geom,traffic_lights_buffer.geom, 1) GROUP BY
route_lines.route_n ORDER BY route_n
```

---------------------------Outcome------------------------------

-------------------------------5----------------------------------

--------Calculate number of pedestrian crossings along route---------

----

--Add column to route_characteristics table called

CREATE TABLE morning_routes.crossings_count AS

SELECT route_lines.route_n,count(crossing_buffer.*) AS crossing_count

FROM morning_routes.route_lines,morning_route_lr_outcomes.crossing_buffer

WHERE ST_DWithin(route_lines.geom,crossing_buffer.geom, 1) GROUP BY route_lines.route_n ORDER BY route_n

----------------------------Outcome------------------------------

-------------------------------6----------------------------------

-------------Calculate number of accidents along route--------------

----

CREATE TABLE morning_routes.accident_count AS

SELECT route_lines.route_n AS route_number,count(traffic_accident_buffer.*) AS accident_count

FROM morning_routes.route_lines,logistic_regression_outcomes_with_fp.traffic_accident_buffer

   WHERE ST_DWithin(route_lines.geom,traffic_accident_buffer.geom, 1) GROUP BY route_lines.route_n ORDER BY route_n

----------------------------Outcome------------------------------

-------------------------------7----------------------------------

-------------Calculate percentage main road along route--------------

----

CREATE TABLE morning_routes.bristol_roads AS

SELECT * FROM planet_osm_line

WHERE highway IS NOT NULL;

--Get srid of raw data (900913)

SELECT st_srid(way) FROM morning_routes.bristol_roads LIMIT 50;

--Transform to WGS 84 4326

ALTER TABLE morning_routes.bristol_roads ALTER COLUMN way TYPE geometry (Linestring,4326) USING ST_Transform(way,4326);

SELECT st_srid(way) FROM morning_routes.bristol_roads LIMIT 50;

--Transform to unprojected OSGB 1936

ALTER TABLE morning_routes.bristol_roads ALTER COLUMN way TYPE geometry (Linestring,4277) USING ST_Transform(way,4277);

SELECT st_srid(way) FROM morning_routes.bristol_roads LIMIT 50

--Transform to BNG

ALTER TABLE morning_routes.bristol_roads ALTER COLUMN way TYPE geometry (Linestring,27700) USING ST_Transform(way,27700);

SELECT st_srid(way) FROM morning_routes.bristol_roads LIMIT 50

CREATE TABLE morning_routes.bristol_main_roads AS

SELECT bristol_roads.way

FROM morning_routes.bristol_roads

WHERE highway IN ('primary','secondary','motorway','trunk',

'tertiary','tertiary link','motorway link',

'secondary link', 'primary link', 'trunk link')

CREATE TABLE morning_routes.main_rd_buffer AS

SELECT st_union(st_buffer(bristol_main_roads.way,15))

FROM morning_routes.bristol_main_roads;

CREATE TABLE morning_routes.main_rd_length AS

SELECT route_n,st_length(st_intersection(route_lines.geom, main_rd_buffer.st_union)) as main_rd_length

FROM morning_routes.route_lines,morning_route_lr_outcomes.main_rd_buffer;

--Add column to route_characteristics called greenspace(%)

ALTER TABLE morning_routes.main_rd_length ADD main_rd_pct double precision;

UPDATE morning_routes.main_rd_length SET main_rd_pct =
((main_rd_length.main_rd_length::double precision /
route_characteristics.length::double precision) * 100)

FROM morning_routes.route_characteristics

WHERE main_rd_length.route_n=route_characteristics.route_n

-----------------------------Outcome------------------------------

--------------------------------8---------------------------------

-----------Calculate percentage residential road along route---------

----

CREATE TABLE morning_routes.bristol_residential_roads AS

SELECT bristol_roads.way

FROM morning_routes.bristol_roads

WHERE highway IN ('living_street','residential')

CREATE TABLE morning_routes.res_rd_buffer AS

SELECT st_union(st_buffer(bristol_residential_roads.way,15))

FROM morning_routes.bristol_residential_roads

CREATE TABLE morning_routes.res_rd_length AS

SELECT route_n,st_length(st_intersection(route_lines.geom,
res_rd_buffer.st_union)) as res_length

FROM morning_routes.route_lines,morning_route_lr_outcomes.res_rd_buffer

--Add column to route_characteristics called greenspace(%)

ALTER TABLE morning_routes.res_rd_length ADD res_rd_pct double precision;

UPDATE morning_routes.res_rd_length SET res_rd_pct =
((res_rd_length.res_length::double precision / route_characteristics.length::double
precision) * 100)

FROM morning_routes.route_characteristics

WHERE res_rd_length.route_n=route_characteristics.route_n

-----------------------------Outcome------------------------------

--------------------------------9----------------------------------

------------Calculate percentage footpath road along route-----------

----

CREATE TABLE morning_routes.bristol_footpaths AS

SELECT bristol_roads.way

FROM morning_routes.bristol_roads

WHERE highway IN ('footway','steps','pedestrian','cycleway','path')

CREATE TABLE morning_routes.footpath_buffer AS

SELECT st_union(st_buffer(bristol_footpaths.way,15))

FROM morning_routes.bristol_footpaths

CREATE TABLE morning_routes.footpath_buffer_b AS

SELECT st_buffer(bristol_footpaths.way,15)

FROM morning_routes.bristol_footpaths

CREATE TABLE morning_routes.footpath_buffer_d AS

SELECT st_union(footpath_buffer_c.geom)

FROM morning_routes.footpath_buffer_c

CREATE TABLE morning_routes.footpath_length AS

SELECT route_n,st_length(st_intersection(route_lines.geom,
footpath_buffer_d.st_union)) as f_length

FROM morning_routes.route_lines,morning_route_lr_outcomes.footpath_buffer_d

--Add column to route_characteristics called greenspace(%)

ALTER TABLE morning_routes.footpath_length ADD footpath_pct double precision;

UPDATE morning_routes.footpath_length SET footpath_pct = ((footpath_length.f_length::double precision / route_characteristics.length::double precision) * 100)

FROM morning_routes.route_characteristics

WHERE footpath_length.route_n=route_characteristics.route_n

-----------------------------Outcome-----------------------------

--------------------------------11---------------------------------

---------------Calculate percentage minor rd along route-------------

----

CREATE TABLE morning_routes.bristol_minor AS

SELECT bristol_roads.way

FROM morning_routes.bristol_roads

WHERE highway IN ('unsurfaced','track','road','unclassified')

CREATE TABLE morning_routes.minor_buffer AS

SELECT st_union(st_buffer(bristol_minor.way,15))

FROM morning_routes.bristol_minor

CREATE TABLE morning_routes.minor_buffer_b AS

SELECT st_union(minor_rd_buffer.geom)

FROM morning_routes.minor_rd_buffer

CREATE TABLE morning_routes.minor_length AS

SELECT route_n,st_length(st_intersection(route_lines.geom, minor_buffer_b.st_union)) as minor_length

FROM morning_routes.route_lines,morning_route_lr_outcomes.minor_buffer_b

--Add column to route_characteristics called greenspace(%)

ALTER TABLE morning_routes.minor_length ADD minor_pct double precision;

UPDATE morning_routes.minor_length SET minor_pct =
((minor_length.minor_length::double precision / route_characteristics.length::double precision) * 100)

FROM morning_routes.route_characteristics

WHERE minor_length.route_n=route_characteristics.route_n

-----------------------------Outcome------------------------------

------------------------------11---------------------------------

---------------Calculate percentage woodland along route-------------

----

CREATE TABLE morning_routes.woodland_unioned AS

SELECT * FROM logistic_regression_outcomes.woodland_unioned

--Create table that contains the length of the the route that is within 25m of bluespace

CREATE TABLE morning_routes.woodland_along_route AS

SELECT
route_n,st_length(st_intersection(route_lines.geom,woodland_unioned.st_union)) as woodland_length

FROM morning_routes.route_lines,morning_route_lr_outcomes.woodland_unioned

--Add column to route_characteristics called woodland(%)

ALTER TABLE morning_routes.woodland_along_route ADD woodland_pct double precision;

UPDATE morning_routes.woodland_along_route SET woodland_pct =
((woodland_along_route.woodland_length::double precision /
route_characteristics.length::double precision) * 100)

FROM morning_routes.route_characteristics

WHERE woodland_along_route.route_n=route_characteristics.route_n

----------------------------Outcome------------------------------

------------------------------12---------------------------------

---------------Calculate number food outlets along route-------------

----

-- Create food outlets table

CREATE TABLE morning_routes.food_outlets AS

SELECT * FROM food_outlets.unhealthy_bristol_outlets

-- Count food outets along routes

CREATE TABLE morning_routes.exposure AS

SELECT route_lines.route_n,count(food_outlets.*) AS outlet_count FROM morning_routes.route_lines, morning_routes.food_outlets

   WHERE ST_DWithin(route_lines.geom,food_outlets.geom, 100) GROUP BY route_lines.route_n

----------------------------Outcome------------------------------

------------------------------11---------------------------------

------------------------Create meta data---------------------------

----

--Add in greenspace data

ALTER TABLE morning_routes.route_characteristics ADD greenspace_pct double precision;

UPDATE morning_routes.route_characteristics SET greenspace_pct = greenspace.greenspace_pct

FROM morning_routes.greenspace

WHERE route_characteristics.route_n = greenspace.route_n;

--Add in bluespace data

ALTER TABLE morning_routes.route_characteristics ADD bluespace_pct double precision;

```sql
UPDATE morning_routes.route_characteristics SET bluespace_pct =
bluespace_along_route.bluespace_pct

FROM morning_routes.bluespace_along_route

WHERE route_characteristics.route_n = bluespace_along_route.route_n;


--Add in traffic_light data

ALTER TABLE morning_routes.route_characteristics ADD traffic_light_count bigint;


UPDATE morning_routes.route_characteristics SET traffic_light_count =
traffic_light_count.traffic_light_count

FROM morning_routes.traffic_light_count

WHERE route_characteristics.route_n = traffic_light_count.route_n;


UPDATE morning_routes.route_characteristics

SET traffic_light_count = 0

WHERE route_characteristics.traffic_light_count IS NULL;


--Add in pedestrian_crossing data

ALTER TABLE morning_routes.route_characteristics ADD pedestrian_crossings
bigint;


UPDATE morning_routes.route_characteristics SET pedestrian_crossings_count =
crossings_count.crossing_count

FROM morning_routes.crossings_count

WHERE route_characteristics.route_n = crossings_count.route_n;


UPDATE morning_routes.route_characteristics

SET pedestrian_crossings_count = 0

WHERE route_characteristics.pedestrian_crossings_count IS NULL;


--Add in accident data

ALTER TABLE morning_routes.route_characteristics ADD accidents_count bigint;
```

```sql
UPDATE morning_routes.route_characteristics SET accidents_count =
accident_count.accident_count

FROM morning_routes.accident_count

WHERE route_characteristics.route_n = accident_count.route_number;


UPDATE morning_routes.route_characteristics

SET accidents_count = 0

WHERE route_characteristics.accidents_count IS NULL;


--Add in main_road data

ALTER TABLE morning_routes.route_characteristics ADD main_road double
precision;


UPDATE morning_routes.route_characteristics SET main_road =
main_rd_length.main_rd_pct

FROM morning_routes.main_rd_length

WHERE route_characteristics.route_n = main_rd_length.route_n;


UPDATE  morning_routes.route_characteristics

SET main_road = 0

WHERE route_characteristics.main_road IS NULL;


--Add in residential_road data

ALTER TABLE morning_routes.route_characteristics ADD residential_road double
precision;


UPDATE morning_routes.route_characteristics SET residential_road =
res_rd_length.res_rd_pct

FROM morning_routes.res_rd_length

WHERE route_characteristics.route_n = res_rd_length.route_n;


UPDATE  morning_routes.route_characteristics

SET residential_road = 0

WHERE route_characteristics.residential_road IS NULL;
```

--Add in footpath data

ALTER TABLE morning_routes.route_characteristics ADD footpath double precision;

UPDATE morning_routes.route_characteristics SET footpath = footpath_length.footpath_pct

FROM morning_routes.footpath_length

WHERE route_characteristics.route_n = footpath_length.route_n;

UPDATE morning_routes.route_characteristics

SET footpath = 0

WHERE route_characteristics.footpath IS NULL;

--Add in minor data

ALTER TABLE morning_routes.route_characteristics ADD minor_road double precision;

UPDATE morning_routes.route_characteristics SET minor_road = minor_length.minor_pct

FROM morning_routes.minor_length

WHERE route_characteristics.route_n = minor_length.route_n;

UPDATE morning_routes.route_characteristics

SET minor_road = 0

WHERE route_characteristics.minor_road IS NULL;

--Add in woodland data

ALTER TABLE morning_routes.route_characteristics ADD woodland double precision;

UPDATE morning_routes.route_characteristics

SET woodland = woodland_along_route.woodland_pct

FROM morning_routes.woodland_along_route

WHERE route_characteristics.route_n = woodland_along_route.route_n;

SELECT * FROM morning_routes.route_characteristics

ORDER BY route_n

----------------------------Exposure-----------------------------

-----------------------------Scores------------------------------

----

ALTER TABLE morning_routes.route_characteristics ADD exposure bigint;

UPDATE morning_routes.route_characteristics SET exposure =
exposure.outlet_count

FROM morning_routes.exposure

WHERE route_characteristics.route_n = exposure.route_n;

UPDATE morning_routes.route_characteristics

SET exposure = 0

WHERE route_characteristics.exposure IS NULL

----------------------------Add in--------------------------------

----------------------------pear_id-------------------------------

----

ALTER TABLE morning_routes.route_characteristics ADD COLUMN "pear_id"
numeric

UPDATE morning_routes.route_characteristics SET pear_id = pupilid

FROM morning_routes.school_routes

WHERE route_characteristics.route_n = school_routes.route_n;

Appendix 9.

Typology of opening times created for outlets classified by typology based on Lake. Opening and closing times based on opening and closing times provided by Local Authorities

Outlet Type	Closing time
Café/Coffee Shop	09:00 - 17:00
Bakery	08:00 - 17:00
Takeaway	17:00 - 23:00
Supermarket	07:00 - 23:00
Convenience	07:00 - 23:00
Non Food Outlet	09:00 - 17:00
Specialist	09:00 - 17:00

Appendix 10.

Example of SQL code used to generate modelled routes to school along weighted network in PostGIS:

```
CREATE TABLESPACE newstore LOCATION 'D:\postgresql\data';

SET default_tablespace = fastpace
```

```
--Once osm is imported in to postGIS through osm2pgsql, create a table with just
road data. Pull out all entries in highway column.

--Raw OSM data is imported into public schema

CREATE TABLE road_network.osm_highway AS

SELECT * FROM planet_osm_line

WHERE highway IS NOT NULL
```

```
--Now reproject this data


--Get SRID

SELECT st_SRID(way) FROM road_network.osm_highway;


--Transform to WGS 84 4326

ALTER TABLE road_network.osm_highway ALTER COLUMN way TYPE geometry
(Linestring,4326) USING ST_Transform(way,4326);

--Check that transformation has been done

SELECT st_srid(way) FROM road_network.osm_highway LIMIT 50;


--Transform to unprojected OSGB 1936

ALTER TABLE road_network.osm_highway ALTER COLUMN way TYPE geometry
(Linestring,4277) USING ST_Transform(way,4277);

SELECT st_srid(way) FROM road_network.osm_highway LIMIT 50;
```

```
--Transform to BNG

ALTER TABLE road_network.osm_highway ALTER COLUMN way TYPE geometry
(Linestring,27700) USING ST_Transform(way,27700);

SELECT st_srid(way) FROM road_network.osm_highway LIMIT 50;
```

```
--Creating a routable network

CREATE TABLE road_network.network AS

SELECT * FROM road_network.osm_highway;


--Node the lines. This creates the table network_noded

SELECT pgr_nodeNetwork('road_network.network', 0.00001,'osm_id','way');


--Create vertices table containing source and target columns

SELECT pgr_createTopology('road_network.network_noded',0.00001, 'way',
'id','source', 'target');


--Analyse new table with pgr_analyzeGraph

SELECT pgr_analyzeGraph('road_network.network_noded',0.00001,'way','id');


--Identify the isolated segments in the network

SELECT *

   FROM road_network.network_noded a,
road_network.network_noded_vertices_pgr b,
road_network.network_noded_vertices_pgr c

   WHERE a.source=b.id AND b.cnt=1 AND a.target=c.id AND c.cnt=1;



-- Create a table of isolated edges of old_id from original line table, road_network

CREATE TABLE road_network.isolated AS

SELECT old_id FROM

(SELECT *
```

```sql
    FROM road_network.network_noded a,
road_network.network_noded_vertices_pgr b,
road_network.network_noded_vertices_pgr c

    WHERE a.source=b.id AND b.cnt=1 AND a.target=c.id AND c.cnt=1) AS foo;
```

--Delete isolated nodes from road_network_noded

```sql
DELETE FROM road_network.network_noded

USING road_network.isolated

WHERE network_noded.old_id = isolated.old_id
```

-- The road_network_noded_vertices_pgr was then deleted so that the topology could be recreated after the isolated edges were removed

```sql
DROP TABLE road_network.network_noded_vertices_pgr
```

--recreate topology

```sql
SELECT pgr_createTopology('road_network.network_noded',0.00001, 'way', 'id','source', 'target')
```

--Analyse new network

```sql
SELECT pgr_analyzeGraph('road_network.network_noded',0.00001,'way','id')
```

---HOMES--

```sql
CREATE TABLE home_sch.homes AS

SELECT * FROM home_sch.homes_within_10km_hsw
```

--FIND NEAREST NODE ALONG NETWORK FOR HOME AND SCHOOL LOCATIONS

--add node id column

```sql
ALTER TABLE home_sch.homes ADD COLUMN "nearest_node_id" bigint;
```

--get closest node on the road network for each home

```sql
UPDATE home_sch.homes SET nearest_node_id=

(SELECT id FROM road_network.network_noded_vertices_pgr

        ORDER BY the_geom <->
ST_GeomFromText(ST_AsText(homes.geom),27700)

        LIMIT 1);
--
--confirm that the nodes_id column has been populated

select * from home_sch.homes LIMIT 100




ALTER TABLE home_sch.homes ADD COLUMN "school_id" bigint;




UPDATE home_sch.homes SET school_id =1
WHERE (st_dwithin(




--Add in nearest node of school access point
ALTER TABLE home_sch.homes ADD COLUMN "nearest_school_node" bigint;




--get closest node on the road network for each home
UPDATE home_sch.homes SET nearest_school_node=

(SELECT nearest_node_id FROM home_sch.school_access_points

        ORDER BY geom <->
ST_GeomFromText(ST_AsText(homes.geom),27700)

        LIMIT 1);
--confirm that the nodes_id column has been populated

select * from test_data.school_access_points LIMIT 100




--Add in 5 column address from ABP
```

---SCHOOL--

--SELECT st_srid(geom) FROM test_data.school_access_points

--add node id column

ALTER TABLE home_sch.school_access_points ADD COLUMN "nearest_node_id" bigint;

--get closest node on the road network for each home

UPDATE home_sch.school_access_points SET nearest_node_id=

(SELECT id FROM road_network.network_noded_vertices_pgr

      ORDER BY the_geom <->
ST_GeomFromText(ST_AsText(school_access_points.geom),27700)

      LIMIT 1);

--confirm that the nodes_id column has been populated

select * from test_data.school_access_points LIMIT 100

--OUTLET---

--LOCATIONS--

--Select out outlets where LA provided uprn but they did not match with ABP uprns

CREATE TABLE food_outlets.outlets_without_coords AS

SELECT * FROM food_outlets.unhealthy_food_outlets

WHERE easting IS NULL

AND match_code=1

OR easting = 'NA'

AND match_code=1

--Create a table where NULL coords are removed

CREATE TABLE food_outlets.outlets_with_coords AS

SELECT * FROM food_outlets.unhealthy_food_outlets

WHERE easting IS NOT NULL

DELETE FROM food_outlets.outlets_with_coords WHERE easting = 'NA'

--parse coord columns from character varying to numeric

ALTER TABLE food_outlets.outlets_with_coords ALTER COLUMN easting TYPE numeric USING (easting::numeric)

ALTER TABLE food_outlets.outlets_with_coords ALTER COLUMN northing TYPE numeric USING (northing::numeric)

--ALTER TABLE food_outlets.unhealthy_food_outlets RENAME COLUMN field_20 TO y

-- Create geometry column

SELECT AddGeometryColumn('food_outlets','outlets_with_coords','geom',27700,'POINT',2);

UPDATE food_outlets.outlets_with_coords SET geom = ST_GeomFromText('POINT(' || outlets_with_coords.easting || ' ' || outlets_with_coords.northing|| ')', 27700);

---Network---
---Preparation---

-- First we must weight the road network add column to road network line table to store road weighting

ALTER TABLE road_network.network_noded ADD COLUMN "costing" double precision;

--Assign osm road classification to noded network

ALTER TABLE road_network.network_noded ADD COLUMN highway character varying (50)

UPDATE road_network.network_noded SET highway = osm_highway.highway

FROM road_network.osm_highway

WHERE network_noded.old_id = osm_highway.osm_id

--Three statistically significant environmental characteristics traffic lights, accidents, exposure

--Traffic Signals

--Undertake this in osm database

CREATE TABLE wales_traffic_signals AS

SELECT * FROM planet_osm_point

WHERE highway = 'traffic_signals'

SELECT st_SRID(way) FROM wales_traffic_signals;

--Transform to WGS 84 4326

ALTER TABLE wales_traffic_signals ALTER COLUMN way TYPE geometry (POINT,4326) USING ST_Transform(way,4326);

--Check that transformation has been done

SELECT st_srid(way) FROM wales_traffic_signals LIMIT 50;

--Transform to unprojected OSGB 1936

ALTER TABLE wales_traffic_signals ALTER COLUMN way TYPE geometry (POINT,4277) USING ST_Transform(way,4277);

SELECT st_srid(way) FROM wales_traffic_signals LIMIT 50;

--Transform to BNG

ALTER TABLE wales_traffic_signals ALTER COLUMN way TYPE geometry (POINT,27700) USING ST_Transform(way,27700);

SELECT st_srid(way) FROM wales_traffic_signals LIMIT 50;

CREATE TABLE road_network.ts_union AS

SELECT st_union(traffic_signals.geom)

FROM road_network.traffic_signals

CREATE TABLE road_network.roads_with_traffic_signals AS

```sql
SELECT network_noded.id,old_id,source,target,way,network_noded.highway
FROM road_network.network_noded,road_network.ts_union

WHERE ST_DWithin(network_noded.way,ts_union.st_union, 5)


ALTER TABLE road_network.network_noded ADD COLUMN "traffic_signals"
double precision;


UPDATE road_network.network_noded

SET traffic_signals = 1

FROM road_network.roads_with_traffic_signals

WHERE network_noded.id = roads_with_traffic_signals.id


UPDATE road_network.network_noded

SET traffic_signals = 0

WHERE network_noded.traffic_signals IS NULL



--Exposures
CREATE TABLE food_outlets.outlets_union AS

SELECT st_union(outlets_with_coords.geom)

FROM food_outlets.outlets_with_coords


CREATE TABLE road_network.roads_with_outlets AS

SELECT network_noded.id,old_id,source,target,way,network_noded.highway
FROM road_network.network_noded,food_outlets.outlets_union

WHERE ST_DWithin(network_noded.way,outlets_union.st_union, 25);


ALTER TABLE road_network.network_noded ADD COLUMN "outlets" double
precision;


UPDATE road_network.network_noded

SET outlets = 1

FROM road_network.roads_with_outlets

WHERE network_noded.id = roads_with_outlets.id;
```

```sql
UPDATE road_network.network_noded

SET outlets = 0

WHERE network_noded.outlets IS NULL;


SELECT * FROM exposure_models.network_noded

WHERE outlets =1


--Accidents
CREATE TABLE road_network.accidents AS

SELECT * FROM road_network."DfTRoadSafety_Accidents_2014"

UNION

SELECT * FROM road_network."DfTRoadSafety_Accidents_2013"


CREATE TABLE road_network.accidents_union AS

SELECT st_union(accidents.geom)

FROM road_network.accidents


CREATE TABLE road_network.roads_with_accidents AS

SELECT network_noded.id,old_id,source,target,way,network_noded.highway
FROM road_network.network_noded,road_network.accidents_union

WHERE ST_DWithin(network_noded.way,accidents_union.st_union, 1);


ALTER TABLE road_network.network_noded ADD COLUMN "accidents" double
precision;


UPDATE road_network.network_noded

SET accidents = 1

FROM road_network.roads_with_accidents

WHERE network_noded.id = roads_with_accidents.id;


UPDATE road_network.network_noded

SET accidents = 0
```

WHERE network_noded.accidents IS NULL;

SELECT * FROM exposure_models.roads_with_accidents

--MORNING--

--ROUTES---

--set costings in line table

UPDATE road_network.network_noded SET costing= 2 ;

UPDATE road_network.network_noded SET costing= 0.8 WHERE highway IN ('primary','secondary','motorway','trunk','tertiary','tertiary link','motorway link','secondary link', 'primary link', 'trunk link'); --MAIN ROAD

UPDATE road_network.network_noded SET costing= 0.5 WHERE highway IN ('living_street','residential'); --RESIDENTIAL

UPDATE road_network.network_noded SET costing= 0.8 WHERE highway IN ('footway','steps','pedestrian','cycleway','path'); --FOOTPATH

UPDATE road_network.network_noded SET costing= 1.3 WHERE highway IN ('unsurfaced','track','road','unclassified','service','proposed'); -- MINOR ROAD

UPDATE road_network.network_noded SET costing= 1000000 WHERE highway IN ('proposed'); -- PROPOSED ROAD

UPDATE road_network.network_noded SET costing= 1 WHERE traffic_signals = 1; -- ROAD EDGES WITH TRAFFIC SIGNALS

UPDATE road_network.network_noded SET costing= 0.8 WHERE accidents = 1; -- ROADS THAT HAVE HAD ACCIDENTS ON THEM

--UPDATE road_network.network_noded SET costing= 1.6 WHERE outlets = 1; -- ROADS THAT HAVE HAD OUTLETS ON THEM

CREATE INDEX ways_class_idx ON road_noded (id);

CREATE INDEX classes_idx ON road (id);

--

UPDATE exposure_models.routing_locations

SET school_node = nearest_node_id

FROM test_data.regression_end_route

WHERE routing_locations.school_id = regression_end_route.school_cod

----

---School 1---

---------------------------------------Bro Myrddin---

--shortest routes with geom

CREATE TABLE routes.bro_myrddin AS

SELECT *

FROM pgr_dijkstra(

   'SELECT id::integer as id,

      source::integer,

      target::integer,

      st_length(way)*costing::double precision AS cost FROM road_network.network_noded',

   ARRAY(SELECT homes.nearest_node_id FROM home_sch.homes, home_sch.hsw_schools WHERE st_dwithin(homes.geom,(SELECT hsw_schools.geom WHERE hsw_schools.tid = 1),4828)),

   (SELECT nearest_node FROM home_sch.hsw_schools WHERE tid = 1),false)

JOIN road_network.network_noded ON pgr_dijkstra.edge = network_noded.id;

--index routes

ALTER TABLE routes.bro_myrddin ADD COLUMN route_n INTEGER;

UPDATE routes.bro_myrddin SET route_n = v_table_name.rn

FROM

(

   SELECT dense_rank() over (order by start_vid) AS rn, seq

   FROM routes.bro_myrddin

) AS v_table_name

WHERE routes.bro_myrddin.seq = v_table_name.seq;

```
ALTER TABLE routes.bro_myrddin ADD COLUMN school INTEGER;


UPDATE routes.bro_myrddin

SET school = 9;



SELECT * FROM routes.bro_myrddin ORDER BY seq LIMIT 50


--Union the geometries for each route

CREATE TABLE routes.union_bro_myrddin AS

SELECT route_n, ST_Union(way) AS geom

FROM routes.bro_myrddin

GROUP BY route_n

ORDER BY route_n;
```

--School 2--

---Pentrehafod--

```
--shortest routes with geom

CREATE TABLE routes.pentrehafod AS

SELECT *

FROM pgr_dijkstra(

    'SELECT id::integer as id,

        source::integer,

        target::integer,

        st_length(way)*costing::double precision AS cost FROM
road_network.network_noded',

    ARRAY(SELECT homes.nearest_node_id FROM home_sch.homes,
home_sch.hsw_schools WHERE st_dwithin(homes.geom,(SELECT
hsw_schools.geom WHERE hsw_schools.tid = 2),4828)),

    (SELECT nearest_node FROM home_sch.hsw_schools WHERE tid = 2),false)

JOIN road_network.network_noded ON pgr_dijkstra.edge = network_noded.id;
```

```
--index routes

ALTER TABLE routes.pentrehafod ADD COLUMN route_n INTEGER;


UPDATE routes.pentrehafod SET route_n = v_table_name.rn
FROM
(
    SELECT dense_rank() over (order by start_vid) AS rn, seq
    FROM routes.pentrehafod
) AS v_table_name
WHERE routes.pentrehafod.seq = v_table_name.seq;


ALTER TABLE routes.pentrehafod ADD COLUMN school INTEGER;


UPDATE routes.pentrehafod
SET school = 3;



SELECT * FROM routes.pentrehafod ORDER BY seq LIMIT 50
```

---School 3---

---Dylan Thomas--

```
--shortest routes with geom
CREATE TABLE routes.dylan_thomas AS
SELECT *
FROM pgr_dijkstra(
    'SELECT id::integer as id,
        source::integer,
```

```
        target::integer,

        st_length(way)*costing::double precision AS cost FROM
road_network.network_noded',

    ARRAY(SELECT homes.nearest_node_id FROM home_sch.homes,
home_sch.hsw_schools WHERE st_dwithin(homes.geom,(SELECT
hsw_schools.geom WHERE hsw_schools.tid = 3),4828)),

  (SELECT nearest_node FROM home_sch.hsw_schools WHERE tid = 3),false)

JOIN road_network.network_noded ON pgr_dijkstra.edge = network_noded.id;




--index routes
ALTER TABLE routes.dylan_thomas ADD COLUMN route_n INTEGER;


UPDATE routes.dylan_thomas SET route_n = v_table_name.rn
FROM
(
    SELECT dense_rank() over (order by start_vid) AS rn, seq
    FROM routes.dylan_thomas
) AS v_table_name
WHERE routes.dylan_thomas.seq = v_table_name.seq;


ALTER TABLE routes.dylan_thomas ADD COLUMN school INTEGER;


UPDATE routes.dylan_thomas
SET school = 5;



SELECT * FROM routes.dylan_thomas ORDER BY seq LIMIT 50
```

--School 4---

--Willows--

--shortest routes with geom

CREATE TABLE routes.willows AS

SELECT *

FROM pgr_dijkstra(

   'SELECT id::integer as id,

      source::integer,

      target::integer,

      st_length(way)*costing::double precision AS cost FROM road_network.network_noded',

   ARRAY(SELECT homes.nearest_node_id FROM home_sch.homes, home_sch.hsw_schools WHERE st_dwithin(homes.geom,(SELECT hsw_schools.geom WHERE hsw_schools.tid = 4),4828)),

  (SELECT nearest_node FROM home_sch.hsw_schools WHERE tid = 4),false)

JOIN road_network.network_noded ON pgr_dijkstra.edge = network_noded.id;

--index routes

ALTER TABLE routes.willows ADD COLUMN route_n INTEGER;

UPDATE routes.willows SET route_n = v_table_name.rn

FROM

(

  SELECT dense_rank() over (order by start_vid) AS rn, seq

  FROM routes.willows

) AS v_table_name

WHERE routes.willows.seq = v_table_name.seq;

ALTER TABLE routes.willows ADD COLUMN school INTEGER;

UPDATE routes.willows

SET school = 6;

SELECT * FROM routes.willows ORDER BY seq LIMIT 50

--School 5--

--Bishop Gore---

--shortest routes with geom

CREATE TABLE routes.bishop_gore AS

SELECT *

FROM pgr_dijkstra(

   'SELECT id::integer as id,

      source::integer,

      target::integer,

      st_length(way)*costing::double precision AS cost FROM road_network.network_noded',

   ARRAY(SELECT homes.nearest_node_id FROM home_sch.homes, home_sch.hsw_schools WHERE st_dwithin(homes.geom,(SELECT hsw_schools.geom WHERE hsw_schools.tid = 5),4828)),

   (SELECT nearest_node FROM home_sch.hsw_schools WHERE tid = 5),false)

JOIN road_network.network_noded ON pgr_dijkstra.edge = network_noded.id;

--index routes

ALTER TABLE routes.bishop_gore ADD COLUMN route_n INTEGER;

UPDATE routes.bishop_gore SET route_n = v_table_name.rn

FROM

(

```
    SELECT dense_rank() over (order by start_vid) AS rn, seq

    FROM routes.bishop_gore

) AS v_table_name

WHERE routes.bishop_gore.seq = v_table_name.seq;


ALTER TABLE routes.bishop_gore ADD COLUMN school INTEGER;


UPDATE routes.bishop_gore

SET school = 1;



SELECT * FROM routes.bishop_gore ORDER BY seq LIMIT 50
```

--School 6--

--Porth--

```
--shortest routes with geom

CREATE TABLE routes.porth AS

SELECT *

FROM pgr_dijkstra(

    'SELECT id::integer as id,

        source::integer,

        target::integer,

        st_length(way)*costing::double precision AS cost FROM
road_network.network_noded',

    ARRAY(SELECT homes.nearest_node_id FROM home_sch.homes,
home_sch.hsw_schools WHERE st_dwithin(homes.geom,(SELECT
hsw_schools.geom WHERE hsw_schools.tid = 6),4828)),

    (SELECT nearest_node FROM home_sch.hsw_schools WHERE tid = 6),false)

JOIN road_network.network_noded ON pgr_dijkstra.edge = network_noded.id;
```

```
--index routes
ALTER TABLE routes.porth ADD COLUMN route_n INTEGER;


UPDATE routes.porth SET route_n = v_table_name.rn
FROM
(
    SELECT dense_rank() over (order by start_vid) AS rn, seq
    FROM routes.porth
) AS v_table_name
WHERE routes.porth.seq = v_table_name.seq;


ALTER TABLE routes.porth ADD COLUMN school INTEGER;


UPDATE routes.porth
SET school = 7;



SELECT * FROM routes.porth ORDER BY seq LIMIT 50




--------------------------------------------------------------------------------------------------
-------------------------------------------School 7-----------------------------------------------
---------------------------------------------Yr Yrfa---------------------------------------------
--------------------------------------------------------------------------------------------------
--shortest routes with geom
CREATE TABLE routes.yr_yrfa AS
SELECT *
FROM pgr_dijkstra(
    'SELECT id::integer as id,
            source::integer,
            target::integer,
```

st_length(way)*costing::double precision AS cost FROM road_network.network_noded',

  ARRAY(SELECT homes.nearest_node_id FROM home_sch.homes, home_sch.hsw_schools WHERE st_dwithin(homes.geom,(SELECT hsw_schools.geom WHERE hsw_schools.tid = 7),4828)),

  (SELECT nearest_node FROM home_sch.hsw_schools WHERE tid = 7),false)

JOIN road_network.network_noded ON pgr_dijkstra.edge = network_noded.id;

--index routes

ALTER TABLE routes.yr_yrfa ADD COLUMN route_n INTEGER;

UPDATE routes.yr_yrfa SET route_n = v_table_name.rn

FROM

(

  SELECT dense_rank() over (order by start_vid) AS rn, seq

  FROM routes.yr_yrfa

) AS v_table_name

WHERE routes.yr_yrfa.seq = v_table_name.seq;

ALTER TABLE routes.yr_yrfa ADD COLUMN school INTEGER;

UPDATE routes.yr_yrfa

SET school = 8;

SELECT * FROM routes.yr_yrfa ORDER BY seq LIMIT 50

---School 8---

---Olchfa---

---Entrance A---

--Update some homes with front entrance

```sql
UPDATE home_sch.homes

SET olchfa_entrance_node = 12118

FROM home_sch.change_school_entrance

WHERE homes.uprn = change_school_entrance.uprn


--shortest routes with geom

CREATE TABLE routes.olchfa_a AS

SELECT *

FROM pgr_dijkstra(

    'SELECT id::integer as id,

        source::integer,

        target::integer,

        st_length(way)*costing::double precision AS cost FROM
road_network.network_noded',

    ARRAY(SELECT homes.nearest_node_id FROM home_sch.homes,
home_sch.hsw_schools WHERE st_dwithin((SELECT homes.geom WHERE
olchfa_entrance_node = 12118),(SELECT hsw_schools.geom WHERE
hsw_schools.tid = 8),4828)), 12118,false)

JOIN road_network.network_noded ON pgr_dijkstra.edge = network_noded.id;



--index routes

ALTER TABLE routes.olchfa_a ADD COLUMN route_n INTEGER;


UPDATE routes.olchfa_a SET route_n = v_table_name.rn

FROM

(

    SELECT dense_rank() over (order by start_vid) AS rn, seq

    FROM routes.olchfa_a

) AS v_table_name

WHERE routes.olchfa_a.seq = v_table_name.seq;


ALTER TABLE routes.olchfa_a ADD COLUMN school INTEGER;
```

```
UPDATE routes.olchfa_a

SET school = 21;



SELECT * FROM routes.olchfa_a ORDER BY seq LIMIT 50
```

---School 8--

---Olchfa--

--Entrance B---

```
--shortest routes with geom

CREATE TABLE routes.olchfa_b AS

SELECT *

FROM pgr_dijkstra(

    'SELECT id::integer as id,

        source::integer,

        target::integer,

        st_length(way)*costing::double precision AS cost FROM
road_network.network_noded',

    ARRAY(SELECT homes.nearest_node_id FROM home_sch.homes,
home_sch.hsw_schools WHERE st_dwithin((SELECT homes.geom WHERE
olchfa_entrance_node = 11835),(SELECT hsw_schools.geom WHERE
hsw_schools.tid = 8),4828)), 11835,false)

JOIN road_network.network_noded ON pgr_dijkstra.edge = network_noded.id;



--index routes

ALTER TABLE routes.olchfa_b ADD COLUMN route_n INTEGER;


UPDATE routes.olchfa_b SET route_n = v_table_name.rn

FROM

(

    SELECT dense_rank() over (order by start_vid) AS rn, seq

    FROM routes.olchfa_b
```

) AS v_table_name

WHERE routes.olchfa_b.seq = v_table_name.seq;

ALTER TABLE routes.olchfa_b ADD COLUMN school INTEGER;

UPDATE routes.olchfa_b

SET school = 22;

SELECT * FROM routes.olchfa_b ORDER BY seq LIMIT 50

---School 8---

---Olchfa---

--Entrance C--

--shortest routes with geom

CREATE TABLE routes.olchfa_c AS

SELECT *

FROM pgr_dijkstra(

   'SELECT id::integer as id,

      source::integer,

      target::integer,

      st_length(way)*costing::double precision AS cost FROM road_network.network_noded',

   ARRAY(SELECT homes.nearest_node_id FROM home_sch.homes, home_sch.hsw_schools WHERE st_dwithin((SELECT homes.geom WHERE olchfa_entrance_node = 291082),(SELECT hsw_schools.geom WHERE hsw_schools.tid = 8),4828)), 291082,false)

JOIN road_network.network_noded ON pgr_dijkstra.edge = network_noded.id;

--index routes

ALTER TABLE routes.olchfa_c ADD COLUMN route_n INTEGER;

```
UPDATE routes.olchfa_c SET route_n = v_table_name.rn

FROM

(

    SELECT dense_rank() over (order by start_vid) AS rn, seq

    FROM routes.olchfa_c

) AS v_table_name

WHERE routes.olchfa_c.seq = v_table_name.seq;


ALTER TABLE routes.olchfa_c ADD COLUMN school INTEGER;


UPDATE routes.olchfa_c

SET school = 23;


--Union routes for three entrances

CREATE TABLE routes.olchfa AS

SELECT * FROM routes.olchfa_a

UNION ALL

SELECT * FROM routes.olchfa_b

UNION ALL

SELECT * FROM routes.olchfa_c;


ALTER TABLE routes.olchfa ADD COLUMN unique_id SERIAL PRIMARY KEY;


--Add in new route index for whole table

ALTER TABLE routes.olchfa ADD COLUMN route_idx INTEGER;


UPDATE routes.olchfa SET route_idx = v_table_name.rn

FROM

(

    SELECT dense_rank() over (order by start_vid,route_n,school) AS rn, unique_id

    FROM routes.olchfa
```

) AS v_table_name

WHERE routes.olchfa.unique_id = v_table_name.unique_id;

SELECT * FROM routes.olchfa

ORDER BY route_idx

LIMIT 1000

---School 9---

---Strade---

UPDATE home_sch.hsw_schools SET nearest_node = 197461 WHERE
school_name = 'Y Strade'

--shortest routes with geom

CREATE TABLE routes.strade AS

SELECT *

FROM pgr_dijkstra(

   'SELECT id::integer as id,

       source::integer,

       target::integer,

       st_length(way)*costing::double precision AS cost FROM
road_network.network_noded',

   ARRAY(SELECT homes.nearest_node_id FROM home_sch.homes,
home_sch.hsw_schools WHERE st_dwithin(homes.geom,(SELECT
hsw_schools.geom WHERE hsw_schools.tid = 9),4828)),

   (SELECT nearest_node FROM home_sch.hsw_schools WHERE tid = 9),false)

JOIN road_network.network_noded ON pgr_dijkstra.edge = network_noded.id;

--index routes

ALTER TABLE routes.strade ADD COLUMN route_n INTEGER;

UPDATE routes.strade SET route_n = v_table_name.rn

FROM

```
(
    SELECT dense_rank() over (order by start_vid) AS rn, seq
    FROM routes.strade
) AS v_table_name
WHERE routes.strade.seq = v_table_name.seq;


ALTER TABLE routes.strade ADD COLUMN school INTEGER;


UPDATE routes.strade
SET school = 10;



SELECT * FROM routes.strade ORDER BY seq LIMIT 50
```

---School 10---
---Fitzalan--
---Entrance A--

```
--shortest routes with geom
CREATE TABLE routes.fitzalan_a AS
SELECT *
FROM pgr_dijkstra(
    'SELECT id::integer as id,
            source::integer,
            target::integer,
            st_length(way)*costing::double precision AS cost FROM
road_network.network_noded',
    ARRAY(SELECT homes.nearest_node_id FROM home_sch.homes,
home_sch.hsw_schools WHERE st_dwithin((SELECT homes.geom WHERE
fitz_entrance_node = 39855),(SELECT hsw_schools.geom WHERE
hsw_schools.tid = 10),4828)), 39855,false)
JOIN road_network.network_noded ON pgr_dijkstra.edge = network_noded.id;
```

```
--index routes

ALTER TABLE routes.fitzalan_a ADD COLUMN route_n INTEGER;


UPDATE routes.fitzalan_a SET route_n = v_table_name.rn

FROM

(

    SELECT dense_rank() over (order by start_vid) AS rn, seq

    FROM routes.fitzalan_a

) AS v_table_name

WHERE routes.fitzalan_a.seq = v_table_name.seq;


ALTER TABLE routes.fitzalan_a ADD COLUMN school INTEGER;


UPDATE routes.fitzalan_a

SET school = 41;



SELECT * FROM routes.fitzalan_a ORDER BY seq LIMIT 50
```

---School 10--

---Fitzalan---

---Entrance B--

```
--shortest routes with geom

CREATE TABLE routes.fitzalan_b AS

SELECT *

FROM pgr_dijkstra(

    'SELECT id::integer as id,
```

```
        source::integer,

        target::integer,

        st_length(way)*costing::double precision AS cost FROM
road_network.network_noded',

    ARRAY(SELECT homes.nearest_node_id FROM home_sch.homes,
home_sch.hsw_schools WHERE st_dwithin((SELECT homes.geom WHERE
fitz_entrance_node = 323134),(SELECT hsw_schools.geom WHERE
hsw_schools.tid = 10),4828)), 323134,false)

JOIN road_network.network_noded ON pgr_dijkstra.edge = network_noded.id;
```

--index routes

```
ALTER TABLE routes.fitzalan_b ADD COLUMN route_n INTEGER;


UPDATE routes.fitzalan_b SET route_n = v_table_name.rn

FROM

(

    SELECT dense_rank() over (order by start_vid) AS rn, seq

    FROM routes.fitzalan_b

) AS v_table_name

WHERE routes.fitzalan_b.seq = v_table_name.seq;


ALTER TABLE routes.fitzalan_b ADD COLUMN school INTEGER;


UPDATE routes.fitzalan_b

SET school = 42;



SELECT * FROM routes.fitzalan_b ORDER BY seq LIMIT 50
```

---School 10---

---Fitzalan---

---Entrance C---

285

--shortest routes with geom

CREATE TABLE routes.fitzalan_c AS

SELECT *

FROM pgr_dijkstra(

   'SELECT id::integer as id,

      source::integer,

      target::integer,

      st_length(way)*costing::double precision AS cost FROM road_network.network_noded',

  ARRAY(SELECT homes.nearest_node_id FROM home_sch.homes, home_sch.hsw_schools WHERE st_dwithin((SELECT homes.geom WHERE fitz_entrance_node = 323122),(SELECT hsw_schools.geom WHERE hsw_schools.tid = 10),4828)), 323122,false)

JOIN road_network.network_noded ON pgr_dijkstra.edge = network_noded.id;

--index routes

ALTER TABLE routes.fitzalan_c ADD COLUMN route_n INTEGER;

UPDATE routes.fitzalan_c SET route_n = v_table_name.rn

FROM

(

  SELECT dense_rank() over (order by start_vid) AS rn, seq

  FROM routes.fitzalan_c

) AS v_table_name

WHERE routes.fitzalan_c.seq = v_table_name.seq;

ALTER TABLE routes.fitzalan_c ADD COLUMN school INTEGER;

UPDATE routes.fitzalan_c

SET school = 43;

```sql
CREATE TABLE routes.fitzalan AS

SELECT * FROM routes.fitzalan_a

UNION ALL

SELECT * FROM routes.fitzalan_b

UNION ALL

SELECT * FROM routes.fitzalan_c;


ALTER TABLE routes.fitzalan ADD COLUMN unique_id SERIAL PRIMARY KEY;


--Add in new route index for whole table

ALTER TABLE routes.fitzalan ADD COLUMN route_idx INTEGER;


UPDATE routes.fitzalan SET route_idx = v_table_name.rn

FROM

(

    SELECT dense_rank() over (order by start_vid,route_n,school) AS rn, unique_id

    FROM routes.fitzalan

) AS v_table_name

WHERE routes.fitzalan.unique_id = v_table_name.unique_id;
```

------------------------------------Merge each school in to one table----------------------------
--

```sql
olchfa

UNION

SELECT * FROM routes.willows

UNION

SELECT * FROM routes.bishop_gore

UNION

SELECT * FROM routes.porth

UNION
```

287

```sql
SELECT * FROM routes.yr_yrfa

UNION

SELECT * FROM routes.olchfa_a

UNION

SELECT * FROM routes.olchfa_b

UNION

SELECT * FROM routes.olchfa_c

UNION

SELECT * FROM routes.strade

UNION

SELECT * FROM routes.fitzalan_a

UNION

SELECT * FROM routes.fitzalan_b

UNION

SELECT * FROM routes.fitzalan_b


CREATE TABLE routes.morning_routes AS

SELECT * FROM routes.pentrehafod

UNION

SELECT * FROM routes.olchfa_a

UNION

SELECT * FROM routes.olchfa_b


--SELECT * FROM exposure_models.model_1_route ORDER BY
route_n,start_vid,path_seq;


ALTER TABLE routes.morning_routes ADD COLUMN unique_id SERIAL PRIMARY
KEY;


--Add in new route index for whole table

ALTER TABLE routes.morning_routes ADD COLUMN route_idx INTEGER;


UPDATE routes.morning_routes SET route_idx = v_table_name.rn
```

```sql
FROM
(

    SELECT dense_rank() over (order by start_vid,route_n,school) AS rn, unique_id

    FROM routes.morning_routes
) AS v_table_name
WHERE routes.morning_routes.unique_id = v_table_name.unique_id;

---------------------------------------------------------------------

--Union the geometries for each route
CREATE TABLE routes.union_geoms AS
SELECT route_idx, ST_Union(way) AS geom
FROM routes.morning_routes
GROUP BY route_idx
ORDER BY route_idx


SELECT * FROM routes.morning_routes
ORDER BY route_idx, path_seq
LIMIT 400


SELECT COUNT (*) FROM routes.union_geoms
WHERE route_idx= 10
-- Count food outets along routes
CREATE TABLE routes.food_outlet_count AS
SELECT union_geoms.route_idx AS route_number,count(outlets_with_coords.*) AS
outlet_count FROM routes.union_geoms, food_outlets.outlets_with_coords

    WHERE ST_DWithin(union_geoms.geom,outlets_with_coords.geom, 100)
GROUP BY union_geoms.route_idx


SELECT * FROM routes.food_outlet_count ORDER BY route_number LIMIT 10
```

--create a table that contains starting node for each route number

CREATE TABLE routes.start_node AS

SELECT * FROM routes.morning_routes WHERE path_seq=1 ORDER BY route_idx

-- Create a table that contains start node for each route

--First create new table with all household data in it from 'homes_within_three_miles' table

CREATE TABLE routes.home_exposures AS

SELECT * FROM home_sch.homes

--Cast nearest_node_id column from bigint to integer

--Create new column where nearest_node_id can be copied to as an integer

ALTER TABLE routes.home_exposures ADD COLUMN nearest_node_id_int integer;

UPDATE routes.home_exposures

SET nearest_node_id_int = nearest_node_id;

--Change the column name back to nearest_node_id

ALTER TABLE routes.home_exposures DROP COLUMN nearest_node_id;

ALTER TABLE routes.home_exposures ADD COLUMN nearest_node_id integer;

UPDATE routes.home_exposures

SET nearest_node_id = nearest_node_id_int;

ALTER TABLE routes.home_exposures DROP COLUMN nearest_node_id_int;

--SELECT * FROM sail_rerun.home_exposures LIMIT 100

-- The add in column that will contain the route_n

ALTER TABLE routes.home_exposures ADD COLUMN route_n integer

--Update the route_n column to be equal to route_n number in start table

UPDATE routes.home_exposures

SET route_n = start_node.route_idx

FROM routes.start_node

WHERE nearest_node_id = start_node.start_vid;

--Add column for exposure scores

ALTER TABLE routes.home_exposures ADD COLUMN exposures integer;

UPDATE routes.home_exposures

SET exposures = food_outlet_count.outlet_count

FROM routes.food_outlet_count

WHERE routes.home_exposures.route_n = food_outlet_count.route_number

SELECT * FROM sail_rerun.homes WHERE nearest_node_id = 126138

SELECT COUNT (*) FROM routes.morning_routes

WHERE school_id IS NULL

--CREATE SAIL META DATA

CREATE TABLE routes.sail_meta AS

SELECT tid AS SYSTEM_ID,uprn AS UPRN, x_coordinate AS OS_X, y_coordinate AS OS_Y, route_n,exposures As ENV_METRIC_1

FROM routes.home_exposures;

SELECT COUNT (*) FROM routes.home_exposures

SELECT * FROM routes.home_exposures

ORDER BY tid

LIMIT 10;

```sql
SELECT * FROM routes.sail_meta LIMIT 10


SELECT * FROM food_outlets.outlets_with_coordinates

WHERE opened < '01/01/2015'

OR closed > '01/01/2015'

OR opened IS NULL


--Get mastermap for Wales

CREATE TABLE wales_la.wales_mm AS



--Get mastermap for Swansea

CREATE TABLE wales_la.swansea_mm AS

SELECT * FROM osmm_topo.topographicarea,wales_la.swansea

WHERE st_dwithin(topographicarea.geom,swansea.geom,2000)

DELETE FROM logistic_regression_outcomes_with_fp.traffic_accident_buffer
WHERE field_12 < '07:30'

DELETE FROM logistic_regression_outcomes_with_fp.traffic_accident_buffer
WHERE field_12 > '09:30' AND field_12 < '14:30:00'

DELETE FROM logistic_regression_outcomes_with_fp.traffic_accident_buffer
WHERE field_12 > '16:30'


CREATE TABLE logistic_regression_outcomes_with_fp.accident_count AS

SELECT route_lines.route_idx AS route_number,count(traffic_accident_buffer.*) AS
accident_count

FROM
all_routes.route_lines,logistic_regression_outcomes_with_fp.traffic_accident_buffer

   WHERE ST_DWithin(route_lines.geom,traffic_accident_buffer.geom, 1) GROUP
BY route_lines.route_idx ORDER BY route_idx
```

CREATE TABLE wales_la.swansea_mm AS

SELECT * FROM osmm_topo.topographicarea,wales_la.swansea

WHERE st_dwithin(topographicarea.geom,swansea.geomet,5)

SELECT * FROM wales_la.swansea_mm

CREATE TABLE wales_la.swansea_mm AS

SELECT * FROM osmm_topo.topographicarea,wales_la.swansea

WHERE st_dwithin(topographicarea.geom,swansea.geomet,5)

CREATE TABLE wales_la.test AS

SELECT * FROM osmm_topo.topographicarea,wales_la.swansea

WHERE st_dwithin(topographicarea.geom,swansea.geomet,5)

LIMIT 10

CREATE TABLE wales_la.swansea_mm AS

SELECT ogc_fid,fid,featurecode,version,
versiondate,theme,broken,calculatedareavalue,changedate,reasonforchange,descri
ptivegroup,descriptiveterm,make,physicallevel,physicalpresence,style_description,st
yle_code,geom

FROM osmm_topo.topographicarea,wales_la.swansea

WHERE st_dwithin(topographicarea.geom,swansea.geomet,5)

SELECT st_SRID(geomet) FROM wales_la.swansea

LIMIT 10;

Bibliography

1. National Obesity Observatory: **The economic burden of obesity**. 2010(October):1–13.

2. Tremmel M, Gerdtham U-G, Nilsson P, Saha S: **Economic Burden of Obesity: A Systematic Literature Review**. *Int J Environ Res Public Health* 2017, **14**:435.

3. Biro FM, Wien M: **Childhood obesity and adult morbidities 1 – 4**. 2010, **91**:1499–1505.

4. Serdula M, Ivery D, Coates R, Freedman D, Williamson D, Byers T: **Do obese children become obese adults**. *Preventive* 1993, **22**:167–177.

5. Gómez Puente JM, Martínez-Marcos M: **Overweight and obesity: Effectiveness of interventions in adults**. *Enfermería Clínica (English Ed* 2018, **28**:65–74.

6. Roskam AR, Sc M, Schaap MM, Sc M: **Socioeconomic Inequalities in Health in 22 European Countries**. *N Engl J Med* 2008, **359**:1290–1291.

7. Bailey L: *Measurement Programme for Wales*. Cardiff; 2015.

8. Rutter H, Savona N, Glonti K, Bibby J, Cummins S, Finegood DT, Greaves F, Harper L, Hawe P, Moore L, Petticrew M, Rehfuess E, Shiell A, Thomas J, White M: **The need for a complex systems model of evidence for public health**. *Lancet* 2017, **6736**:9–11.

9. Butland B, Jebb S, Kopelman P, McPherson K, Thomas S, Mardell J, Parry V: **Tackling Obesities : Future Choices – Project report**. *Gov Off Sci* 2007:1–161.

10. Hobbs M, Green M, Griffiths C, Jordan H, Saunders J, McKenna J: **How different data sources and definitions of neighbourhood influence the association between food outlet availability and body mass index: a cross-sectional study**. *Perspect Public Health* 2017, **137**:158–161.

11. Harrison F, Burgoine T, Corder K, van Sluijs EMF, Jones A: **How well do modelled routes to school record the environments children are exposed to?: A cross-sectional comparison of GIS-modelled and GPS-measured routes to school**. *Int J Health Geogr* 2014, **13**:5.

12. Lobstein T, Baur L, Uauy R, Obesity I: **Obesity in children and young people : a crisis in**. 2004, **5**:4–85.

13. Han JC, Lawlor D a, Kimm SYS: **Childhood obesity.** *Lancet* 2010, **375**:1737–48.

14. Williams J, Scarborough P, Matthews a., Cowburn G, Foster C, Roberts N, Rayner M: **A systematic review of the influence of the retail food environment around schools on obesity-related outcomes.** *Obes Rev* 2014, **15**:1–16.

15. McLellan F: **Obesity rising to alarming levels around the world**. *Lancet* 2002, **359**:1412.

16. Choices N: **Obesity - NHS Choices**. .

17. Himes JH: **Challenges of Accurately Measuring and Using BMI and Other Indicators of Obesity in Children**. *Pediatrics* 2009, **124**(Supplement 1):s3–s22.

18. Johansson E, Böckerman P, Kiiskinen U, Heliövaara M: **Obesity and labour market success in Finland: The difference between having a high BMI and being fat**. *Econ Hum Biol* 2009, **7**:36–45.

19. Krakauer NY, Krakauer JC: **A new body shape index predicts mortality hazard independently of body mass index**. *PLoS One* 2012, **7**.

20. WHO: **Waist Circumference and Waist-Hip Ratio: Report of a WHO Expert Consultation**. *World Heal Organ* 2008(December):8–11.

21. Elobeid MA, Desmond RA, Thomas O, Keith SW, Allison DB: **Waist circumference values are increasing beyond those expected from BMI increases**. *Obesity* 2007, **15**:2380–2383.

22. Ford E, Maynard L, Li C: **Trends in Mean Waist Circumference and Abdominal Obesity Among US Adults, 1999-2012**. *J Am Med Assoc* 2014, **312**:1151–1153.

23. Burkhauser R V., Cawley J: **Beyond BMI: The value of more accurate measures of fatness and obesity in social science research**. *J Health Econ* 2008, **27**:519–529.

24. Gorber SC, Tremblay M, Moher D, Gorber B: **A comparison of direct vs. self-report measures for assessing height, weight and body mass index: A systematic review**. *Obes Rev* 2007, **8**:307–326.

25. Blackburn H, Jacobs D: **Commentary: Origins and evolution of Body Mass Index (BMI): Continuing saga**. *Int J Epidemiol* 2014, **43**:665–669.

26. O'Neill D: **Measuring obesity in the absence of a gold standard**. *Econ Hum Biol* 2015, **17**(June 2013):116–128.

27. Kelley GA, Kelley KS: **Exercise and BMI z-score in overweight and obese children and adolescents: protocol for a systematic review and network meta-analysis of randomised trials**. *BMJ Open* 2016, **6**:e011258.

28. Hagström H, Tynelius P, Rasmussen F: **High BMI in late adolescence predicts future severe liver disease and hepatocellular carcinoma: a national, population-based cohort study in 1.2 million men**. *Gut* 2017, **0**:1–7.

29. Rho YH, Lu N, Peloquin CE, Man A, Zhu Y, Zhang Y, Choi HK: **Independent impact of gout on the risk of diabetes mellitus among women and men: a population-based, BMI-matched cohort study**. *Ann Rheum Dis* 2016, **75**:91–95.

30. Wong WW, Strizich G, Heo M, Heymsfield SB, Himes JH, Rock CL, Gellman MD, Siega-Riz AM, Sotres-Alvarez D, Davis SM, Arredondo EM, Van Horn L, Wylie-Rosett J, Sanchez-Johnsen L, Kaplan RC, Mossavar-Rahmani Y: **Relationship between body fat and BMI in a US hispanic population-based cohort study: Results from HCHS/SOL**. *Obesity* 2016, **24**:1561–1571.

31. Conrad D, Capewell S: **Associations between deprivation and rates of childhood overweight and obesity in England, 2007-2010: an ecological study.** *BMJ Open* 2012, **2**:e000463.

32. Cetateanu A, Jones A: **Understanding The Relationship Between Food Environments, Deprivation And Childhood Overweight And Obesity: Evidence From A Cross Sectional England-Wide Study**. *Health Place* 2014.

33. Finegood DT, Merth TDN, Rutter H: **Implications of the Foresight Obesity System Map for Solutions to Childhood Obesity**. *Obesity* 2010, **18**:S13–S16.

34. Dahlgren G, Whitehead M: *Policies and Strategies to Promote Social Equity in Health Background Document to WHO – Strategy Paper. Volume 14*; 2007(September 1991).

35. Karra E, O'Daly OG, Choudhury AI, Yousseif A, Millership S, Neary MT, Scott WR,

Chandarana K, Manning S, Hess ME, Iwakura H, Akamizu T, Millet Q, Gelegen C, Drew ME, Rahman S, Emmanuel JJ, Williams SCR, R??ther UU, Br??ning JC, Withers DJ, Zelaya FO, Batterham RL: **A link between FTO, ghrelin, and impaired brain food-cue responsivity**. *J Clin Invest* 2013, **123**:3539–3551.

36. Malone M: **Medications associated with weight gain.** *Ann Pharmacother* 2005, **39**:2046–55.

37. Vila L, Roglans N, Alegret M, Sanchez RM, Vazquez-Carrera M, Laguna JC: **Suppressor of cytokine signaling-3 (SOCS-3) and a deficit of serine/threonine (Ser/Thr) phosphoproteins involved in leptin transduction mediate the effect of Fructose on rat liver lipid metabolism**. *Hepatology* 2008, **48**:1506–1516.

38. Fisher JO, Rolls BJ, Birch LL: **Children's bite size and intake of an entrée are greater with large portions than with age-appropriate or self-selected portions**. *Am J Clin Nutr* 2003, **77**:1164–1170.

39. Johnson-Taylor WL, Everhart JE: **Modifiable environmental and behavioral determinants of overweight among children and adolescents: report of a workshop.** *Obesity (Silver Spring)* 2006, **14**:929–966.

40. Sallis JE, Glanz K: **Environments in Activity , Eating , Childhood**. *Futur Child* 2006, **16**:89–108.

41. Sallis JF, Spoon C, Cavill N, Engelberg JK, Gebel K, Parker M, Thornton CM, Lou D, Wilson AL, Cutter CL, Ding D: **Co-benefits of designing communities for active living: an exploration of literature.** *Int J Behav Nutr Phys Act* 2015, **12**:30.

42. Egger G, Swinburn B: **An "ecological" approach to the obesity pandemic.** *BMJ* 1997, **315**:477–80.

43. Shier V, An R, Sturm R: **Is there a robust relationship between neighbourhood food environment and childhood obesity in the USA?** *Public Health* 2012, **126**:723–30.

44. Bava CM, Jaeger SR, Park J: **Constraints upon food provisioning practices in "busy" women's lives: Trade-offs which demand convenience**. *Appetite* 2008, **50**:486–498.

45. Drewnowski A: **Obesity and the food environment: dietary energy density and diet costs.** *Am J Prev Med* 2004, **27**(3 Suppl):154–62.

46. Coombes E, Jones A: **Gamification of active travel to school: A pilot evaluation of the Beat the Street physical activity intervention**. *Health Place* 2016, **39**:62–69.

47. van Sluijs EMF, Skidmore PML, Mwanza K, Jones AP, Callaghan AM, Ekelund U, Harrison F, Harvey I, Panter J, Wareham NJ, Cassidy A, Griffin SJ: **Physical activity and dietary behaviour in a population-based sample of British 10-year old children: the SPEEDY study (Sport, Physical activity and Eating behaviour: environmental Determinants in Young people).** *BMC Public Health* 2008, **8**:388.

48. Van Der Horst K, Oenema a., Ferreira I, Wendel-Vos W, Giskes K, Van Lenthe F, Brug J: **A systematic review of environmental correlates of obesity-related dietary behaviors in youth**. *Health Educ Res* 2007, **22**:203–226.

49. Story M, Neumark-Sztainer D, French S: **Individual and environmental influences on adolescent eating behaviors.** *J Am Diet Assoc* 2002(3 Suppl):S40–S51.

50. Crockett SJ, Sims LS: **Environmental influences on children's health**. *Lancet* 1995, **367**:224–369.

51. Maziak W, Ward KD, Stockton MB: **Childhood obesity: are we missing the big picture?** *Obes Rev* 2008, **9**:35–42.

52. Wiecha JL, Peterson KE, Ludwig DS, Kim J, Sobol A, Gortmaker SL: **When Children Eat What They Watch**. *Arch Pediatr Adolesc Med* 2013, **160**:436–442.

53. Utter J, Scragg R, Schaaf D: **Associations between television viewing and consumption of commonly advertised foods among New Zealand children and young adolescents.** *Public Health Nutr* 2006, **9**:606–612.

54. Aktaş Arnas Y: **The effects of television food advertisement on children's food purchasing requests**. *Pediatr Int* 2006, **48**:138–145.

55. Papas M a, Alberg AJ, Ewing R, Helzlsouer KJ, Gary TL, Klassen AC: **The built environment and obesity.** *Epidemiol Rev* 2007, **29**:129–43.

56. Harrison F, Jones AP, van Sluijs EMF, Cassidy A, Bentham G, Griffin SJ: **Environmental correlates of adiposity in 9-10 year old children: Considering home and school neighbourhoods and routes to school**. *Soc Sci Med* 2011, **72**:1411–1419.

57. Day PL, Pearce J: **Obesity-promoting food environments and the spatial clustering of food outlets around schools.** *Am J Prev Med* 2011, **40**:113–21.

58. Beaulac J, Kristjansson E, Cummins S: **A systematic review of food deserts, 1966-2007.** *Prev Chronic Dis* 2009, **6**:A105.

59. Smoyer-Tomic KEJCSKDRCANCVYNCEHJH: **The association between neighbourhood socioeconomic statusand exposure to supermarkets and fast food outlets**. *Health Place* 2008, **14**:740–754.

60. Lovasi GS, Hutson M a, Guerra M, Neckerman KM: **Built environments and obesity in disadvantaged populations.** *Epidemiol Rev* 2009, **31**:7–20.

61. Burdette H: **Neighborhood playgrounds, fast food restaurants, and crime: relationships to overweight in low-income preschool children**. *Prev Med (Baltim)* 2004, **38**:57–63.

62. Lake A a, Townshend T, Alvanides S: *Obesogenic Environments: Complexities, Perceptions and Objective Measures.* Chichester, UK: Wiley-Blackwell; 2010.

63. Lake A, Townshend T: **Obesogenic environments: exploring the built and food environments**. *J R Soc Promot Health* 2006, **126**:262–267.

64. Hendrickson D, Smith C, Eikenberry N: **Fruit and vegetable access in four low-income food deserts communities in Minnesota**. *Agric Human Values* 2006, **23**:371–383.

65. Cummins S, Macintyre S: **"Food deserts"--evidence and assumption in health policy making.** *BMJ* 2002, **325**:436–8.

66. Pinstrup-Andersen P: **Food security: definition and measurement**. *Food Secur* 2009, **1**:5–7.

67. Davies G, Frausin G, Parry L: **Are There Food Deserts in Rainforest Cities?** *Ann Am Assoc Geogr* 2017, **107**:794–811.

68. Smith D, Cummins S, Clark C, Stansfeld S: **Does the local food environment around schools affect diet? Longitudinal associations in adolescents attending secondary schools in East London.** *BMC Public Health* 2013, **13**:70.

69. Ni Mhurchu C, Vandevijvere S, Waterlander W, Thornton LE, Kelly B, Cameron a J, Snowdon W, Swinburn B: **Monitoring the availability of healthy and unhealthy foods and non-alcoholic beverages in community and consumer retail food environments globally.** *Obes Rev* 2013, **14 Suppl 1**(October):108–19.

70. Wilkins EL, Morris MA, Radley D, Griffiths C: **Using Geographic Information Systems to measure retail food environments: Discussion of methodological considerations and a proposed reporting checklist (Geo-FERN)**. *Heal Place* 2017, **44**(October 2016):110–117.

71. Cobb LK, Appel LJ, Franco M, Jones-Smith JC, Nur A, Anderson C a. M: **The relationship of the local food environment with obesity: A systematic review of methods, study quality, and results**. *Obesity* 2015, **23**:1331–1344.

72. Maillot M, Darmon N, Vieux F, Drewnowski A: **Low energy density and high nutritional quality are each associated with higher diet costs in French adults**. *Am J Clin Nutr* 2007, **86**:690–696.

73. Drewnowski A, Darmon N: **Food Choices and Diet Costs: an Economic Analysis**. *Am Soc Nutr Sci* 2005, **135**:900–904.

74. Drewnowski A, Darmon N: **The economics of obesity: dietary energy density and energy cost.** *Am J Clin Nutr* 2005, **82**:265–273.

75. Epstein LH, Dearing KK, Paluch RA, Roemmich JN, Cho D: **Price and maternal obesity influence purchasing of low and high energy-dense foods**. *Am J Clin Nutr* 2007, **86**:914–922.

76. Capacci S, Mazzocchi M, Shankar B: **The regional price of junk foods relative to healthy foods in the UK: indirect estimation of a time series, 1997-2009**. In *86th Annu Conf Agric Econ Soc Univ Warwick, United Kingdom*; 2012.

77. Jones NR V, Conklin AI, Suhrcke M, Monsivais P: **The growing price gap between more and less healthy foods: Analysis of a novel longitudinal UK dataset**. *PLoS One* 2014, **9**.

78. Lloyd S, Lawton J, Caraher M, Singh G, Horsley K, Mussa F: **A tale of two localities: Healthy eating on a restricted income**. *Health Educ J* 2010, **70**:48–56.

79. Monsivais P, Drewnowski A: **The Rising Cost of Low-Energy-Density Foods**. *J Am Diet Assoc* 2007, **107**:2071–2076.

80. Morris MA, Hulme C, Clarke GP, Edwards KL, Cade JE: **What is the cost of a healthy diet? Using diet data from the UK Women's Cohort Study**. *J Epidemiol Community Heal* 2014, **68**:1043=1049.

81. Newby PK: **Examining Energy Density: Comments on Diet Quality, Dietary Advice, and the Cost of Healthful Eating**. *J Am Diet Assoc* 2006, **106**:1166–1169.

82. Rush E, Puniani N, Snowling N, Paterson J: **Food security, selection, and healthy eating in a Pacific Community in Auckland New Zealand**. *Asia Pac J Clin Nutr* 2007, **16**:448–454.

83. Barratt J: **The cost and availability of healthy food choices in southern Derbyshire**. *J Hum Nutr Diet* 1997, **10**:63–69.

84. Williams AJ, Wyatt KM, Hurst AJ, Williams CA: **A systematic review of associations between the primary school built environment and childhood overweight and obesity**. *Health Place* 2012, **18**:504–514.

85. Austin SB, Melly SJ, Sanchez BN, Patel A, Buka S, Gortmaker SL: **Clustering of fast-food**

restaurants around schools: a novel application of spatial statistics to the study of food environments. *Am J Public Health* 2005, **95**:1575–81.

86. Dunton GF, Ph D, Kaplan J, Wolch J, Ph D, Jerrett M, Ph D, Reynolds KD, Ph D: **Systematic Review**. 2013, **10**.

87. Galvez MP, Pearl M, Yen IH: **Childhood Obesity and the Built Environment: A review of the Literature from 2008-2009**. *Curr Opin Pediatr* 2010, **22**:202–207.

88. Fraser LK, Edwards KL, Cade J, Clarke GP: **The geography of Fast Food outlets: a review.** *Int J Environ Res Public Health* 2010, **7**:2290–308.

89. Davis B, Carpenter C: **Proximity of fast-food restaurants to schools and adolescent obesity.** *Am J Public Health* 2009, **99**:505–10.

90. Pearson T, Russell J, Campbell MJ, Barker ME: **Do "food deserts" influence fruit and vegetable consumption?--A cross-sectional study.** *Appetite* 2005, **45**:195–7.

91. Sturm R, Datar a: **Body mass index in elementary school children, metropolitan area food prices and food outlet density.** *Public Health* 2005, **119**:1059–68.

92. An R, Sturm R: **School and residential neighborhood food environment and diet among California youth.** *Am J Prev Med* 2012, **42**:129–35.

93. Sallis JF, Floyd MF, Rodriguez DA, Saelens BE: **The Role of Built Environments in Physical Activity, Obesity, and CVD**. *Circulation* 2012, **125**:729–737.

94. Pearson AL, Bentham G, Day P, Kingham S: **Associations between neighbourhood environmental characteristics and obesity and related behaviours among adult New Zealanders.** *BMC Public Health* 2014, **14**:553.

95. Sarkar C, Gallacher J, Webster C: **Built environment configuration and change in body mass index: The Caerphilly Prospective Study (CaPS)**. *Heal Place* 2013, **19**:33–44.

96. Bedimo-Rung AL: **The Significance of Parks to Physical Activity and**. *Am J Prev Med* 2005, **28**(2S2):159–168.

97. Coombes E, Jones AP, Hillsdon M: **The relationship of physical activity and overweight to objectively measured green space accessibility and use**. *Soc Sci Med* 2010, **70**:816–822.

98. Frumkin H: **Health, equity, and the built environment**. *Environ Health Perspect* 2005, **113**:2001–2002.

99. Comber A, Brunsdon C, Green E: **Using a GIS-based network analysis to determine urban greenspace accessibility for different ethnic and religious groups**. *Landsc Urban Plan* 2008, **86**:103–114.

100. Jones A, Hillsdon M, Coombes E: **Greenspace access, use, and physical activity: Understanding the effects of area deprivation**. *Prev Med (Baltim)* 2009, **49**:500–505.

101. Lachowycz K, Jones a. P: **Greenspace and obesity: A systematic review of the evidence**. *Obes Rev* 2011, **12**:183–189.

102. Cummins S, Fagg J: **Does greener mean thinner? Associations between neighbourhood greenspace and weight status among adults in England.** *Int J Obes* 2012, **36**:1108–1113.

103. Sarkar C, Webster C, Gallacher J: **Association between adiposity outcomes and residential density: a full-data, cross-sectional analysis of 419 562 UK Biobank adult**

participants. *Lancet Planet Heal* 2017, **1**:e277–e288.

104. Levine JA, Vander Weg MW, Hill JO, Klesges RC: **Non-exercise activity thermogenesis: The crouching tiger hidden dragon of societal weight gain**. *Arterioscler Thromb Vasc Biol* 2006, **26**:729–736.

105. Healy G, Dunstan DW, Salmon J, Cerin E, Shaw J, Zimmet P, Owen N: **Beneficial associations with metabolic risk**. *Diabetes Care* 2008, **31**:661–666.

106. De Bourdeaudhuij I, Van Cauwenberghe E, Spittaels H, Oppert JM, Rostami C, Brug J, Van Lenthe F, Lobstein T, Maes L: **School-based interventions promoting both physical activity and healthy eating in Europe: A systematic review within the HOPE project**. *Obes Rev* 2011, **12**:205–216.

107. Kriemler S, Meyer U, Martin E, van Sluijs E, Andersen L, Martin B: **Effect of school-based interventions on physical activity and fitness in children and adolescents: a review of reviews adn systematic update**. *Br J Sports Med* 2011, **45**:923–30.

108. Design Council: *Active by Design: Designing Places for Healthy Lives - a Short Guide*. 2014.

109. Pucher J, Dijkstra L: **Promoting Safe Walking and Cycling to Improve Public Health Walking and Cycling : the MOST sustainable transport modes**. *Am J Public Health* 2003, **93**:1509–1516.

110. Pucher J, Buehler R: **Making cycling irresistible: Lessons from the Netherlands, Denmark and Germany**. *Transp Rev* 2008, **28**:495–528.

111. Merom D, Tudor-Locke C, Bauman A, Rissel C: **Active commuting to school among NSW primary school children: implications for public health.** *Health Place* 2006, **12**:678–687.

112. Sirard JR, Slater ME: **Walking and Bicycling to School: A Review**. *Am J Lifestyle Med* 2008, **2**:372–396.

113. Panter JR, Jones AP, Van Sluijs EMF, Griffin SJ: **Neighborhood, Route, and School Environments and Children's Active Commuting**. *Am J Prev Med* 2010, **38**:268–278.

114. Panter JR, Jones AP, van Sluijs EM: **Environmental determinants of active travel in youth: a review and framework for future research.** *Int J Behav Nutr Phys Act* 2008, **5**:34.

115. Black C, Collins A, Snell M: **Encouraging Walking: The Case of Journey-to-school Trips in Compact Urban Areas**. *Urban Stud* 2001, **38**:1121–1141.

116. Mitra R: **Independent Mobility and Mode Choice for School Transportation: A Review and Framework for Future Research**. *Transp Rev* 2013, **33**:1–23.

117. Ding D, Sallis JF, Kerr J, Lee S, Rosenberg DE: **Neighborhood environment and physical activity among youth: A review**. *Am J Prev Med* 2011, **41**:442–455.

118. Feng J, Glass T a., Curriero FC, Stewart WF, Schwartz BS: **The built environment and obesity: A systematic review of the epidemiologic evidence**. *Heal Place* 2010, **16**:175–190.

119. Durand CP, Andalib M, Dunton GF, Wolch J, Pentz MA: **A Systematic Review of Built Environment Factors Related to Physical Activity and Obesity Risk: Implications for Smart Growth Urban Planning**. *Obes Rev* 2012, **12**:1–15.

120. Forsyth A, Hearst M, Oakes JM, Schmitz KH: **Design and destinations: Factors**

influencing walking and total physical activity. *Urban Stud* 2008, **45**:1973–1996.

121. Reilly M, Landis J: **The Influence of Built-Form and Land Use on Mode Choice**. *Development* 2002, **4**:1–51.

122. **Environmental Barriers to Activity | Obesity Prevention Source | Harvard T.H. Chan School of Public Health** [https://www.hsph.harvard.edu/obesity-prevention-source/obesity-causes/physical-activity-environment/#references]

123. Harrison RA, Gemmell I, Heller RF: **The population effect of crime and neighbourhood on physical activity: An analysis of 15 461 adults**. *J Epidemiol Community Health* 2007, **61**:34–39.

124. Gómez JE, Johnson BA, Selva M, Sallis JF: **Violent crime and outdoor physical activity among inner-city youth**. *Prev Med (Baltim)* 2004, **39**:876–881.

125. Ortega FB, Ruiz JR, Sjostrom M: **Physical activity, overweight and central adiposity in Swedish children and adolescents: the European Youth Heart Study**. *Int J Behav Nutr Phys Act* 2007, **4**:61–71.

126. McGinn AP, Evenson KR, Herring AH, Huston SL, Rodriguez DA: **Exploring associations between physical activity and perceived and objective measures of the built environment**. *J Urban Heal* 2007, **84**:162–184.

127. Boarnet MG, Anderson CL, Day K, Mcmillan T: **Evaluation of the California Safe Routes to School Legislation. Urban Form Changes and Children ' s Active Transportation to School**. *Am J Prev Med* 2005, **28**.

128. Moudon Vernez A, Stewart O: *Moving Forward: Safe Routes to School Progress in Five States*. 2012(July).

129. Sadler RC, Gilliland J a, Arku G: **An application of the edge effect in measuring accessibility to multiple food retailer types in Southwestern Ontario, Canada.** *Int J Health Geogr* 2011, **10**(May):34.

130. Rodgers SE, Demmler JC, Dsilva R, Lyons RA: **Protecting health data privacy while using residence-based environment and demographic data.** *Heal place* 2012, **18**:209–17.

131. Reilly JJ, Methven E, McDowell ZC, Hacking B, Alexander D, Stewart L, Kelnar CJH: **Health consequences of obesity**. *Arch Dis Child* 2003, **88**:748–752.

132. Kaplan KM, Wadden TA: **Childhood obesity and self-esteem**. *J Pediatr* 1986, **109**:367–370.

133. Phillips RG, Hill a J: **Fat, plain, but not friendless: self-esteem and peer acceptance of obese pre-adolescent girls.** *Int J Obes Relat Metab Disord* 1998, **22**:287–293.

134. Chu NF, Wang DJ, Shieh SM, Rimm EB: **Plasma leptin concentrations and obesity in relation to insulin resistance syndrome components among school children in Taiwan--The Taipei Children Heart Study.** *Int J Obes Relat Metab Disord* 2000, **24**:1265–1271.

135. Chinn S, Rona RJ: **Can the increase in body mass index explain the rising trend in asthma in children?** *Thorax* 2001, **56**:845–50.

136. Figueroa-Muñoz JI, Chinn S, Rona RJ: **Association between obesity and asthma in 4-11 year old children in the UK.** *Thorax* 2001, **56**:133–7.

137. Dietz WH: **Health consequences of obesity in youth: childhood predictors of adult**

disease. *Pediatrics* 1998, **101**(3 Pt 2):518–525.

138. Glavin K, Roelants M, Strand BH, Júlíusson PB, Lie KK, Helseth S, Hovengen R: **Important periods of weight development in childhood: a population-based longitudinal study.** *BMC Public Health* 2014, **14**:160.

139. Wang YC, McPherson K, Marsh T, Gortmaker SL, Brown M: **Health and economic burden of the projected obesity trends in the USA and the UK**. *Lancet* 2011, **378**:815–825.

140. Wee CC, Phillips RS, Legedza ATR, Davis RB, Soukup JR, Colditz GA, Hamel MB: **Health care expenditures associated with overweight and obesity among us adults: Importance of age and race**. *Am J Public Health* 2005, **95**:159–165.

141. Withrow D, Alter DA: **The economic burden of obesity worldwide: A systematic review of the direct costs of obesity**. *Obes Rev* 2011, **12**:131–141.

142. Townshend T, Lake A: **Obesogenic environments: current evidence of the built and food environments**. *Perspect Public Health* 2017, **137**:38–44.

143. Craig P, Dieppe P, Macintyre S, Health P, Unit S, Michie S, Nazareth I, Petticrew M: **Developing and evaluating complex interventions : new guidance MRC**. 2008:39.

144. Frieden TR: **A framework for public health action: The health impact pyramid**. *Am J Public Health* 2010, **100**:590–595.

145. Whitlock EP, Orleans CT, Pender N, Allan J: **Evaluating primary care behavioral counseling interventions. An evidence-based approach**. *Am J Prev Med* 2002, **22**:267–284.

146. Moher M, Yudkin P, Wright L, Turner R, Fuller a, Schofield T, Mant D: **Cluster randomised controlled trial to compare three methods of promoting secondary prevention of coronary heart disease in primary care.** *BMJ* 2001, **322**:1338.

147. Gerhardus A, Becher H, Groenewegen P, Mansmann U, Meyer T, Pfaff H, Puhan M, Razum O, Rehfuess E, Sauerborn R, Strech D, Wissing F, Zeeb H, Hummers-Pradier E: **Applying for, reviewing and funding public health research in Germany and beyond**. *Heal Res Policy Syst* 2016, **14**:43.

148. Musa GJ, Chiang P-H, Sylk T, Bavley R, Keating W, Lakew B, Tsou H-C, Hoven CW: **Use of GIS Mapping as a Public Health Tool - From Cholera to Cancer**. *Heal Serv Insights* 2013, **6**:111–116.

149. Light RU: **The Progress of Medical Geography**. *Am Geogr Soc* 1944, **34**:636–641.

150. Monmonier MS: **Jacques M. May**. *Prof Geogr* 1976, **28**:93.

151. **Change in alcohol outlet density and alcohol-related harm to population health (CHALICE): A comprehensive record-linked database study in Wales** [http://www.nets.nihr.ac.uk/projects/phr/09300702]

152. Fone D, Dunstan F, White J, Webster C, Rodgers S, Lee S, Shiode N, Orford S, Weightman A, Brennan I, Sivarajasingam V, Morgan J, Fry R, Lyons R: **Change in alcohol outlet density and alcohol-related harm to population health (CHALICE).** *BMC Public Health* 2012, **12**:428.

153. Burgoine T, Alvanides S, Lake A a: **Creating "obesogenic realities"; do our methodological choices make a difference when measuring the food environment?** *Int J Health Geogr* 2013, **12**:33.

154. Blakey K, Feltbower RG, Parslow RC, James PW, Pozo BG, Stiller C, Vincent TJ, Norman P, Patricia A McKinney, Murphy MF, Craft AW, McNally RJQ: **Is fluoride a risk factor for bone cancer? Small area analysis of osteosarcoma and ewing sarcoma diagnosed among 0-49-year-olds in Great Britain, 1980-2005**. *Int J Epidemiol* 2014, **43**:224–234.

155. Richardson S, Abellan JJ, Best N: **Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire (UK)**. *Stat Methods Med Res* 2006, **15**:385–407.

156. Wheeler BW, Kothencz G, Pollard AS: **Geography of non-melanoma skin cancer and ecological associations with environmental risk factors in England.** *Br J Cancer* 2013, **109**:235–41.

157. Kuchuk AA, Krzyzanowski M, Huysmans K: **The application of WHO's health and environment geographic information system (HEGIS) in mapping environmental health risks for the European region**. *J Hazard Mater* 1998, **61**:287–290.

158. Jonathan A. Patz, Diarmid Campbell-Lendrum TH& JAF: **Impact of regional climate change on human health**. *Nature* 2005, **438**:310–317.

159. Briggs DJ, Hoogh1 C de, Gulliverb J, Wills J, Elliotta P, Kinghamc S, Smallboned K: **A regression-based method for mapping traffic-related air pollution : application and testing in four contrasting urban environments**. 2000:151–167.

160. Nykiforuk CIJ, Flaman LM: **Geographic information systems (GIS) for Health Promotion and Public Health: a review.** *Health Promot Pract* 2011, **12**:63–73.

161. Plescia M, Koontz S, Laurent S: **Community assessment in a vertically integrated health care system**. *Am J Public Health* 2001, **91**:811–814.

162. Schlundt DG, Mushi C, Larson CO, Marrs M: **Use of innovative technologies in the evaluation of Nashville's REACH 2010 community action plan: reducing disparities in cardiovascular disease and diabetes in the African American community.** *J Ambul Care Manage* 2001, **24**:51–60.

163. Apparicio P, Abdelmajid M, Riva M, Shearmur R: **Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues.** *Int J Health Geogr* 2008, **7**:7.

164. Carver A, Timperio A, Crawford D: **Playing it safe: The influence of neighbourhood safety on children's physical activity-A review**. *Heal Place* 2008, **14**:217–227.

165. Nutsford D, Pearson AL, Kingham S: **An ecological study investigating the association between access to urban green space and mental health**. *Public Health* 2013, **127**:1005–1011.

166. Charreire H, Casey R, Salze P, Simon C, Chaix B, Banos A, Badariotti D, Weber C, Oppert J-M: **Measuring the food environment using geographical information systems: a methodological review.** *Public Health Nutr* 2010, **13**:1773–1785.

167. Forsyth A, Lytle L, Riper D Van: **Finding food: Issues and challenges in using Geographic Information Systems to measure food access.** *J Transp Land Use* 2010, **3**:43–65.

168. Burgoine T, Forouhi NG, Griffin SJ: **Associations between exposure to takeaway food outlets , takeaway food consumption , and body weight in Cambridgeshire , UK : population based , cross**. *BMJ* 2014, **348**.

169. Griffiths C, Frearson A, Taylor A, Radley D, Cooke C: **A cross sectional study investigating the association between exposure to food outlets and childhood obesity in Leeds, UK.** *Int J Behav Nutr Phys Act* 2014, **11**:138.

170. **Yell.com - the UK's leading online business directory** [https://www.yell.com/]

171. **InfoUSA | Mailing Lists | Email Marketing Lists | Business | Sales Leads | Consumer** [https://www.infousa.com/]

172. Charreire H, Mackenbach JD, Ouasti M, Lakerveld J, Compernolle S, Ben-Rebah M, McKee M, Brug J, Rutter H, Oppert JM: **Using remote sensing to define environmental characteristics related to physical activity and dietary behaviours: A systematic review (the SPOTLIGHT project)**. *Heal Place* 2014, **25**:1–9.

173. Morland KB, Evenson KR: **Obesity prevalence and the local food environment.** *Health Place* 2009, **15**:491–5.

174. Timperio A, Ball K, Roberts R, Campbell K, Andrianopoulos N, Crawford D: **Children's fruit and vegetable intake: Associations with the neighbourhood food environment**. *Prev Med (Baltim)* 2008, **46**:331–335.

175. Smoyer-Tomic KE, Spence JC, Raine KD, Amrhein C, Cameron N, Yasenovskiy V, Cutumisu N, Hemphill E, Healy J: **The association between neighborhood socioeconomic status and exposure to supermarkets and fast food outlets**. *Heal Place* 2008, **14**:740–754.

176. Inagami S, Cohen DA, Brown AF, Asch SM: **Body mass index, neighborhood fast food and restaurant concentration, and car ownership**. *J Urban Heal* 2009, **86**:683–695.

177. Kwate NOA, Yau CY, Loh JM, Williams D: **Inequality in obesigenic environments: Fast food density in New York City**. *Heal Place* 2009, **15**:364–373.

178. Pearce J, Hiscock R, Blakely T, Witten K: **A national study of the association between neighbourhood access to fast-food outlets and the diet and weight of local residents**. *Heal Place* 2009, **15**:193–197.

179. Simon PA, Kwan D, Angelescu A, Shih M, Fielding JE: **Proximity of fast food restaurants to schools: Do neighborhood income and type of school matter?** *Prev Med (Baltim)* 2008, **47**:284–288.

180. Turrell G, Giskes K: **Socioeconomic disadvantage and the purchase of takeaway food: A multilevel analysis**. *Appetite* 2008, **51**:69–81.

181. Burgoine T, Monsivais P: **Characterising food environment exposure at home, at work, and along commuting journeys using data on adults in the UK.** *Int J Behav Nutr Phys Act* 2013, **10**:85.

182. Fone D, Morgan J, Fry R, Rodgers S, Orford S, Farewell D, Dunstan F, White J, Sivarajasingam V, Trefan L, Brennan I, Lee S, Shiode N, Weightman A, Webster C, Lyons R: **Change in alcohol outlet density and alcohol-related harm to population health (CHALICE): a comprehensive record-linked database study in Wales**. *Chang alcohol outlet density alcohol-related harm to Popul Heal a Compr Rec database study Wales* 2016, **4**.

183. Macintyre S, McKay L, Cummins S, Burns C: **Out-of-home food outlets and area deprivation: case study in Glasgow, UK.** *Int J Behav Nutr Phys Act* 2005, **2**:16.

184. **The Food Standards Act 1999 | Food Standards Agency** [https://www.food.gov.uk/enforcement/regulation/foodstandardsact]

185. **Food Hygiene Rating (Wales) Act 2013 and Food Hygiene Rating (Wales) Regulations 2013: guidance for food authorities | Food Standards Agency** [https://www.food.gov.uk/enforcement/enforcework/hygienescoresresources/fhrs-act-wales]

186. **Local Authority Enforcement Monitoring System (LAEMS) Last reviewed : March 2017**. 2017(March).

187. **Starting a food business | Food Standards Agency** [https://www.food.gov.uk/business-industry/startingup]

188. Burgoine T, Harrison F: **Comparing the accuracy of two secondary food environment data sources in the UK across socio-economic and urban/rural divides.** *Int J Health Geogr* 2013, **12**:2.

189. Lake A, Burgoine T, Stamp E, Grieve R: **The foodscape : classification and field validation of secondary data sources across urban/rural and socio-economic classifications in England.** *Int J Behav Nutr Phys Act* 2012, **9**:37.

190. Burgoine T: **Collecting accurate secondary foodscape data. A reflection on the trials and tribulations**. *Appetite* 2010, **55**:522–527.

191. Li Q, Zhang T, Wang H, Zeng Z: **Dynamic accessibility mapping using floating car data: A network-constrained density estimation approach**. *J Transp Geogr* 2011, **19**:379–393.

192. Devillers R, Bédard Y, Jeansoulin R, Moulin B: **Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data**. *Int J Geogr Inf Sci* 2007, **21**:261–282.

193. Wang Y, Lobstein T: **Worldwide trends in childhood overweight and obesity.** *Int J Pediatr Obes* 2006, **1**:11–25.

194. Paquet C, Daniel M, Kestens Y, Léger K, Gauvin L: **Field validation of listings of food stores and commercial physical activity establishments from secondary data.** *Int J Behav Nutr Phys Act* 2008, **5**:58.

195. Janse AJ, Gemke RJBJ, Uiterwaal CSPM, Van Der Tweel I, Kimpen JLL, Sinnema G: **Quality of life: Patients and doctors don't always agree: A meta-analysis**. *J Clin Epidemiol* 2004, **57**:653–661.

196. Oort van PAJ, Oort P Van: **Spatial data quality: from description to application**. Wageningen Universiteit; 2005.

197. Mendez DD, Kim KH, Hardaway CR, Fabio A: **Neighborhood racial and socioeconomic disparities in the food and alcohol environment: Are there differences by commercial data sources?** *J Racial Ethn Heal Disparities* 2015, **3**:108–116.

198. Lucan SC: **Concerning limitations of food-environment research: A narrative review and commentary framed around obesity and diet-related diseases in youth**. *J Acad Nutr Diet* 2015, **115**:205–212.

199. Guptill S, Morrison J: *Elements of Spatial Data Quality*. 1st edition. Oxford: Pergamon; 1995.

200. Kresse W, Fadaie K: *ISO Standards for Geographic Information*. Berlin: Springer-Verlag; 2003.

201. Zielstra D, Hochmair HH, Neis P: **Assessing the effect of data imports on the**

completeness of openstreetmap - A United States case study. *Trans GIS* 2013, **17**:315–334.

202. Hochmair HH, Zielstra D, Neis P: **Assessing the completeness of bicycle trail and lane features in OpenStreetMap for the United States**. *Trans GIS* 2015, **19**:63–81.

203. Girres JF, Touya G: **Quality Assessment of the French OpenStreetMap Dataset**. *Trans GIS* 2010, **14**:435–459.

204. Haklay M: **How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets**. *Environ Plan B Plan Des* 2010, **37**:682–703.

205. Helbich M, Amelunxen C, Neis P: **Comparative Spatial Analysis of Positional Accuracy of OpenStreetMap and Proprietary Geodata**. *Proc GI_Forum 2012 Geovisualization, Soc Learn* 2012:24–33.

206. Koukoletsos T, Haklay M, Ellul C: **An automated method to assess Data Completeness and Positional Accuracy of OpenStreetMap**. *Int Conf GeoComputation* 2011, **3**:236–241.

207. Neis P, Zielstra D, Zipf A: **The Street Network Evolution of Crowdsourced Maps: OpenStreetMap in Germany 2007–2011**. *Futur Internet* 2011, **4**:1–21.

208. Kounadi O: **Assessing the quality of OpenStreetMap data**. *Geogr Inf Sci Univ Coll* 2009(August):0–80.

209. Keßler C, Theodore R, Groot A De: **Geographic Information Science at the Heart of Europe**. 2013:21–37.

210. Bader MDM, Ailshire JA, Morenoff JD, House JS: **Measurement of the local food environment: A comparison of existing data sources**. *Am J Epidemiol* 2010, **171**:609–617.

211. Boone-Heinonen J, Diez-Roux A V., Goff DC, Loria CM, Kiefe CI, Popkin BM, Gordon-Larsen P: **The neighborhood energy balance equation: Does neighborhood food retail environment + physical activity environment = obesity? The CARDIA study**. *PLoS One* 2013, **8**:1–11.

212. Carroll-Scott A, Gilstad-Hayden K, Rosenthal L, Peters SM, McCaslin C, Joyce R, Ickovics JR: **Disentangling neighborhood contextual associations with child body mass index, diet, and physical activity: The role of built, socioeconomic, and social environments**. *Soc Sci Med* 2013, **95**:106–114.

213. Jennings A, Welch A, Jones AP, Harrison F, Bentham G, Van Sluijs EMF, Griffin SJ, Cassidy A: **Local food outlets, weight status, and dietary intake: Associations in children aged 9-10 years**. *Am J Prev Med* 2011, **40**:405–410.

214. Thornton LE, Pearce JR, Macdonald L, Lamb KE, Ellaway A: **Does the choice of neighbourhood supermarket access measure influence associations with individual-level fruit and vegetable consumption? A case study from Glasgow.** *Int J Health Geogr* 2012, **11**:29.

215. Block JP, Scribner RA, Desalvo KB: **Fast food, race/ethnicity, and income: A geographic analysis**. *Am J Prev Med* 2004, **27**:211–217.

216. Farley TA, Baker ET, Futrell L, Rice JC: **The ubiquity of energy-dense snack foods: A national multicity study**. *Am J Public Health* 2010, **100**:306–311.

217. Lake A a., Burgoine T, Greenhalgh F, Stamp E, Tyrrell R: **The foodscape: Classification**

and field validation of secondary data sources. *Heal Place* 2010, **16**:666–673.

218. Macdonald L, Ellaway A, Macintyre S: **The food retail environment and area deprivation in Glasgow City, UK**. *Int J Behav Nutr Phys Act* 2009, **6**:52.

219. Burgoine T, Lake AA, Stamp E, Alvanides S, Mathers JC, Adamson AJ: **Changing foodscapes 1980-2000, using the ASH30 Study**. *Appetite* 2009, **53**:157–165.

220. Wang MC, Cubbin C, Ahn D, Winkleby MA: **Changes in neighbourhood food store environment, food behaviour and body mass index, 1981–1990**. *Public Health Nutr* 2008, **11**:963–970.

221. White M, Bunting J, Williams E, Raybould S, Adamson A, Mathers JC: **Do "food deserts" exist? A multi-level, geographical analysis of the relationship between retail food access, socio-economic position and dietary intake**. 2004(February).

222. Cummins SCJ, McKay L, MacIntyre S: **McDonald's restaurants and neighborhood deprivation in Scotland and England**. *Am J Prev Med* 2005, **29**:308–310.

223. Fraser LK, Edwards KL: **The association between the geography of fast food outlets and childhood obesity rates in Leeds, UK.** *Health Place* 2010, **16**:1124–8.

224. Fraser LK, Clarke GP, Cade JE, Edwards KL: **Fast food and obesity: a spatial analysis in a large United Kingdom population of children aged 13-15.** *Am J Prev Med* 2012, **42**:e77-85.

225. **Change of use** [https://www.planningportal.co.uk/info/200130/common_projects/9/change_of_use.]

226. **Policy statement – UPRN** [https://www.ordnancesurvey.co.uk/about/governance/policies/addressbase-uprn.html]

227. Singh SK: **Evaluating two freely available geocoding tools for geographical inconsistencies and geocoding errors**. *Open Geospatial Data, Softw Stand* 2017, **2**:11.

228. Faure E, Danjou AMN, Clavel-Chapelon F, Boutron-Ruault M-C, Dossus L, Fervers B: **Accuracy of two geocoding methods for geographic information system-based exposure assessment in epidemiological studies**. *Environ Heal* 2017, **16**:15.

229. Kestens Y, Lebel A, Chaix B, Clary C, Daniel M, Pampalon R, Theriault M, Subramanian S V: **Association between Activity Space Exposure to Food Establishments and Individual Risk of Overweight**. *PLoS One* 2012, **7**.

230. Wen LM, Fry D, Rissel C, Dirkis H, Balafas A, Merom D: **Factors associated with children being driven to school: Implications for walk to school programs**. *Health Educ Res* 2008, **23**:325–334.

231. Hobbs M, Griffiths C, Green M, Jordan H, McKenna J: **Associations between the physical activity and food environment and obesity: A cross sectional study of UK Adults**. *J Epidemiol Community Heal* 2016, **70**.

232. Timperio AF, Ball K, Roberts R, Andrianopoulos N, Crawford DA: **Children ' s takeaway and fast-food intakes : associations with the neighbourhood food environment**. *Public Health Nutr* 2009, **12**:1960–1964.

233. Kestens Y, Lebel A, Daniel M, Thériault M, Pampalon R: **Using experienced activity spaces to measure foodscape exposure.** *Health Place* 2010, **16**:1094–103.

234. Buck C, Börnhorst C, Pohlabeln H, Huybrechts I, Pala V, Reisch L, Pigeot I: **Clustering of unhealthy food around German schools and its influence on dietary behavior in school children: a pilot study.** *Int J Behav Nutr Phys Act* 2013, **10**:65.

235. Howard PH, Fitzpatrick M, Fulfrost B: **Proximity of food retailers to schools and rates of overweight ninth grade students: an ecological study in California.** *BMC Public Health* 2011, **11**:68.

236. Higgs G, Fry R, Langford M: **Investigating the implications of using alternative GIS-based techniques to measure accessibility to green space**. *Environ Plan B Plan Des* 2012, **39**:326–343.

237. Mizen A, Fry R, Grinnell D, E. Rodgers S: **Quantifying the Error Associated with Alternative GIS-based Techniques to Measure Access to Health Care Services**. *Health Serv Res* 2015, **2**:439–445.

238. Dessing D, de Vries SI, Hegeman G, Verhagen E, van Mechelen W, Pierik FH: **Children's route choice during active transportation to school: difference between shortest and actual route**. *Int J Behav Nutr Phys Act* 2016, **13**:48.

239. Lin W: **Revealing the making of OpenStreetMap: A limited account**. *Can Geogr / Le Géographe Can* 2015, **59**:69–81.

240. Goodchild MF: **Citizens as sensors: The world of volunteered geography**. *GeoJournal* 2007, **69**:211–221.

241. Tapscott D, Williams AD, Herman D: *Government 2.0 : Transforming Government and Gover Nance for the Twenty-First Century*. 2008(January).

242. Goodchild MF, Li L: **Assuring the quality of volunteered geographic information**. *Spat Stat* 2012, **1**:110–120.

243. Chilton S, Building F, Burroughs T: **Crowdsourcing Is Radically Changing the Geodata Landscape : Case Study of Openstreetmap**. In *Proc Twenty-fourth Int Cartogr Conf*. Santiago, Chile, Chile; 2009:1–7.

244. Ramm F, Topf J, Chilton S: *OpenStreetMap: Using and Enhancing the Free Map of the World*. Cambridge, UK, UK: UIT Cambridge; 2010.

245. **Stats - OpenStreetMap Wiki** [http://wiki.openstreetmap.org/wiki/Stats]

246. Oort van PAJ: **Spatial data quality: from description to application**. 2006.

247. **ISO 19157:2013 - Geographic information -- Data quality** [http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32575]

248. Mooney P, Corcoran P, Winstanley AC: **Towards quality metrics for OpenStreetMap**. *Proc 18th SIGSPATIAL Int Conf Adv Geogr Inf Syst GIS 10* 2010:514–517.

249. Neis P, Zipf A: **Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap**. *ISPRS Int J Geo-Information* 2012, **1**:146–165.

250. Arsanjani JJ, Barron C, Bakillah M, Helbich M: **Assessing the Quality of OpenStreetMap Contributors together with their Contributions**. *16th Agil Conf* 2013.

251. Barron C, Neis P, Zipf A: **A Comprehensive Framework for Intrinsic OpenStreetMap**

Quality Analysis. *Trans GIS* 2014, **18**:877–895.

252. van Exel M, Dias E, Fruijtier S: **The impact of crowdsourcing on spatial data quality indicators**. 2010(July 2015):1–4.

253. Trame J, Keßler C: **Exploring the lineage of volunteered geographic information with heat maps**. *GeoViz, Hamburg, Ger* 2011:6–7.

254. Mooney P, Corcoran P: **Characteristics of Heavily Edited Objects in OpenStreetMap**. *Futur Internet* 2012, **4**:285–305.

255. Mooney P, Corcoran P: **The Annotation Process in OpenStreetMap**. *Trans GIS* 2012, **16**:561–579.

256. Jean-françois G: **A model to estimate length measurements uncertainty in vector databases**. In *Proc Seventh Int Symp Spat Data Qual*. Coimbra, Portugal; 2011:1–6.

257. Haklay M, Antoniou V, Basiouka S: *Crowdsourced Geographic Information Use in Government. Global Facility for Disaster Reduction & Recovery (GFDRR)*. 2014.

258. Antoniou V, Skopeliti A, Military H, Service G, Engineer S: **Measures and Indicators of VGI Quality: An Overview**. *Remote Sens Spat Inf Sci* 2015, **2**.

259. Sehra SS, Singh J, Rai HS: **A Systematic Study of OpenStreetMap Data Quality Assessment**. In *2014 11th Int Conf Inf Technol New Gener*; 2014:377–381.

260. Ather A: **A quality analysis of openstreetmap data**. 2009(May).

261. Zielstra D, Zipf A: **A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany**. *Methods* 2010, **1**:1–15.

262. Forghani M, Delavar M: **A Quality Study of the OpenStreetMap Dataset for Tehran**. *ISPRS Int J Geo-Information* 2014, **3**:750–763.

263. Hayakawa T, Imi Y, Ito T: **Analysis of Quality of Data in OpenStreetMap**. *2012 IEEE 14th Int Conf Commer Enterp Comput* 2012:131–134.

264. Pourabdollah A, Morley J, Feldman S, Jackson M: **Towards an Authoritative OpenStreetMap: Conflating OSM and OS OpenData National Maps' Road Network**. *ISPRS Int J Geo-Information* 2013, **2**:704–728.

265. Stopher PR, Shen L, Liu W, Ahmed A: **The challenge of obtaining ground truth for GPS processing**. *Transp Res Procedia* 2015, **11**:206–217.

266. Paz-Soldan VA, Reiner RC, Morrison AC, Stoddard ST, Kitron U, Scott TW, Elder JP, Halsey ES, Kochel TJ, Astete H, Vazquez-Prokopec GM: **Strengths and Weaknesses of Global Positioning System (GPS) Data-Loggers and Semi-structured Interviews for Capturing Fine-scale Human Mobility: Findings from Iquitos, Peru**. *PLoS Negl Trop Dis* 2014, **8**.

267. Chaix B, Méline J, Duncan S, Merrien C, Karusisi N, Perchoux C, Lewin A, Labadi K, Kestens Y: **GPS tracking in neighborhood and health studies: a step forward for environmental exposure assessment, a step backward for causal inference?** *Health Place* 2013, **21**:46–51.

268. Burgoine T, Jones AP, Brouwer RJN, Neelon SEB: **Associations between BMI and home , school and route environmental exposures estimated using GPS and GIS : do we see evidence of selective daily mobility bias in children ?** 2015, **14**:1–12.

269. Gesler W: **The uses of spatial analysis in medical geography: A review**. *Soc Sci Med* 1986, **23**:963–973.

270. Sherman JE, Spencer J, Preisser JS, Gesler WM, Arcury TA: **A suite of methods for representing activity space in a healthcare accessibility study**. *Int J Health Geogr* 2005, **4**:24.

271. Lipperman-Kreda S, Morrison C, Grube JW, Gaidus A: **Youth activity spaces and daily exposure to tobacco outlets**. *Health Place* 2015, **34**:30–33.

272. Villanueva K, Giles-Corti B, Bulsara M, McCormack GR, Timperio A, Middleton N, Beesley B, Trapp G: **How far do children travel from their homes? Exploring children's activity spaces in their neighborhood**. *Heal Place* 2012, **18**:263–273.

273. Page AS, Cooper AR, Griew P, Davis L, Hillsdon M: **Independent mobility in relation to weekday and weekend physical activity in children aged 10-11 years: The PEACH Project.** *Int J Behav Nutr Phys Act* 2009, **6**:2.

274. Page AS, Cooper AR, Griew P, Jago R: **Independent mobility, perceptions of the built environment and children's participation in play, active travel and structured exercise and sport: the PEACH Project.** *Int J Behav Nutr Phys Act* 2010, **7**:17.

275. Wheeler BW, Cooper AR, Page AS, Jago R: **Greenspace and children's physical activity: A GPS/GIS analysis of the PEACH project**. *Prev Med (Baltim)* 2010, **51**:148–152.

276. Bohte W, Maat K: **Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands**. *Transp Res Part C Emerg Technol* 2009, **17**:285–297.

277. Page AS, Cooper AR, Griew P, Davis L, Hillsdon M: **Independent mobility in relation to weekday and weekend physical activity in children aged 10–11 years: The PEACH Project**. *Int J Behav Nutr Phys Act* 2009, **6**:2.

278. Cetateanu A, Jones A: **How can GPS technology help us better understand exposure to the food environment? A systematic review**. *SSM - Popul Heal* 2016, **2**:196–205.

279. O'Connor A, Zerger A, Itami B: **Geo-temporal tracking and analysis of tourist movement**. *Math Comput Simul* 2005, **69**:135–150.

280. Wolf ID, Hagenloh G, Croft DB: **Visitor monitoring along roads and hiking trails: How to determine usage levels in tourist sites**. *Tour Manag* 2012, **33**:16–28.

281. Cetateanu A, Luca B-A, Popescu AA, Page A, Cooper A, Jones A: **A novel methodology for identifying environmental exposures using GPS data**. *Int J Geogr Inf Sci* 2016, **8816**(February):1–17.

282. Lachowycz K, Jones AP, Page AS, Wheeler BW, Cooper AR: **What can global positioning systems tell us about the contribution of different types of urban greenspace to children's physical activity?** *Health Place* 2012, **18**:586–94.

283. Zheng Y, Liu L, Wang L, Xie X: **Learning transportation mode from raw gps data for geographic applications on the web**. *Proceeding 17th Int Conf World Wide Web - WWW '08* 2008:247.

284. Auld J, Williams C, Mohammadian A, Nelson P: **An automated GPS-based prompted recall survey with learning algorithms**. *Transp Lett* 2009, **1**:59–79.

285. Xu C, Ji M, Chen W, Zhang Z: **Identifying travel mode from GPS trajectories through**

fuzzy pattern recognition. *Proc - 2010 7th Int Conf Fuzzy Syst Knowl Discov FSKD 2010* 2010, **2**:889–893.

286. Feng T, Timmermans HJP: **Transportation mode recognition using GPS and accelerometer data**. *Transp Res Part C Emerg Technol* 2013, **37**:118–130.

287. Lin M, Hsu WJ, Lee ZQ: **Detecting modes of transport from unlabelled positioning sensor data**. *J Locat Based Serv* 2013:272–290.

288. Cho GH, Rodríguez DA, Evenson KR: **Identifying walking trips using GPS data**. *Med Sci Sports Exerc* 2011, **43**:365–372.

289. Chung EH, Shalaby A: **A trip reconstruction tool for GPS-based personal travel surveys**. *Transp Plan Technol* 2005, **28**:381–401.

290. Ploeg M Ver, Breneman V, Farrigan T, Hamrick K, Hopkins D, Kaufman P, Lin B, Nord M, Smith T, Williams R, Kinnison K, Olander C, Singh A, Tuckermanty E, Krantz-kent R, Polen C, Mcgowan H, Kim S: *Access to Affordable and Nutritious Food : Measuring and Understanding Food Deserts and Their Consequences Report to Congress*. 2009(June).

291. Kelman CW, Bass AJ, Holman CD: **Research use of linked health data--a best practice protocol**. *Aust N Z J Public Health* 2002, **26**:251–255.

292. Ford D V., Jones KH, Verplancke J-P, Lyons RA, John G, Brown G, Brooks CJ, Thompson S, Bodger O, Couch T, Leake K: **The SAIL Databank: building a national architecture for e-health research and evaluation**. *BMC Health Serv Res* 2009, **9**:157.

293. Lyons R a, Jones KH, John G, Brooks CJ, Verplancke J-P, Ford D V, Brown G, Leake K: **The SAIL databank: linking multiple health and social care datasets.** *BMC Med Inform Decis Mak* 2009, **9**:3.

294. Black N: **High-quality clinical databases: Breaking down barriers**. *Lancet* 1999, **353**:1205–1206.

295. Harron K, Gilbert R, Cromwell D, Van Der Meulen J: **Linking data for mothers and babies in de-identified electronic health data**. *PLoS One* 2016, **11**:1–18.

296. Ford I, Murray H, Packard CJ, Shepherd J, Macfarlane PW, Cobbe SM: **Long-Term Follow-up of the West of Scotland Coronary Prevention Study**. *N Engl J Med* 2007, **357**:1477–1486.

297. Exeter DJ, Rodgers S, Sabel CE: **"Whose data is it anyway?" The implications of putting small area-level health and social data online.** *Health Policy* 2014, **114**:88–96.

298. **Unethical to restrict linkage of health data | ScienceNordic**. .

299. Wellcome Trust: **Summary Report of Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data**. 2013(July 2013):1–15.

300. British Heart Foundation: *Policy Statement: Patient Data in Medical Research*. 2012.

301. **What is impact? - Economic and Social Research Council** [http://www.esrc.ac.uk/research/impact-toolkit/what-is-impact/]

302. Parliamentary Office of Science & Technology: **Big Data and Public Health**. *POSTnote* 2014:1–4.

303. Helbich M, Emmichoven MJZ van, Dijst MJ, Kwan M-P, Pierik FH, Vries SI de: **Natural and built environmental exposures on children's active school travel: A Dutch global**

positioning system-based cross-sectional study. *Health Place* 2016, **39**:101–109.

304. Curtis AJ, Mills JW, Leitner M: **Spatial confidentiality and GIS: re-engineering mortality locations from published maps about Hurricane Katrina.** *Int J Health Geogr* 2006, **5**:44.

305. Zimmerman DL, Pavlik C: **Quantifying the effects of mask metadata disclosure and multiple releases on the confidentiality of geographically masked health data**. *Geogr Anal* 2008, **40**:52–76.

306. Kamel Boulos MN, Cai Q, Padget JA, Rushton G: **Using software agents to preserve individual health data confidentiality in micro-scale geographical analyses**. *J Biomed Inform* 2006, **39**:160–170.

307. Zandbergen P a.: **Ensuring Confidentiality of Geocoded Health Data: Assessing Geographic Masking Strategies for Individual-Level Data**. *Adv Med* 2014, **2014**:1–14.

308. Hampton KH, Fitch MK, Allshouse WB, Doherty I a., Gesink DC, Leone P a., Serre ML, Miller WC: **Mapping health data: Improved privacy protection with donut method geomasking**. *Am J Epidemiol* 2010, **172**:1062–1069.

309. Allshouse WB, Fitch MK, Hampton KH, Gesink DC, Doherty I a, Leone P a, Serre ML, Miller WC: **Geomasking sensitive health data and privacy protection: an evaluation using an E911 database.** *Geocarto Int* 2010, **25**:443–452.

310. Huckett JC: **Synthetic data methods for disclosure limitation**. 2008.

311. Young C, Martin D, Skinner C: **Geographically intelligent disclosure control for flexible aggregation of census data**. *Int J Geogr Inf Sci* 2009, **23**:457–482.

312. **Involving People in Research** [http://www.involvingpeopleinresearch.org.au/]

313. Trust W: **Enabling data linkage to maximise the value of public health research data**. 2015(March).

314. Jones KH, Laurie G, Stevens L, Dobbs C, Ford D V., Lea N: **The other side of the coin: Harm due to the non-use of health-related data**. *Int J Med Inform* 2017, **97**:43–51.

315. Wartenberg D, Thompson WD: **Privacy versus public health: the impact of current confidentiality rules.** *Am J Public Health* 2010, **100**:407–12.

316. Armstrong MP, Rushton G, Zimmerman DL: **Geographically masking health data to preserve confidentiality.** *Stat Med* 1999, **18**:497–525.

317. O'Brien DG, Yasnoff W a: **Privacy, confidentiality, and security in information systems of state health agencies.** *Am J Prev Med* 1999, **16**:351–8.

318. Buckovich S a, Rippen HE, Rozen MJ: **Driving Toward Guiding Principles: A Goal for Privacy, Confidentiality, and Security of Health Information**. *J Am Med Informatics Assoc* 1999, **6**:122–133.

319. Poots AJ, Green SA, Ealing NHS, Barnes R, Bell D: **Using routine geo-coded data to identify geographical heterogeneity to reduce disparities : case studies in UK .** 2012(c):40–43.

320. Boyd JH, Ferrante AM, O'Keefe CM, Bass AJ, Randall SM, Semmens JB: **Data linkage infrastructure for cross-jurisdictional health-related research in Australia**. *BMC Health Serv Res* 2012, **12**:480.

321. Holman CDA, Bass AJ, Rosman DL, Smith MB, Semmens JB, Glasson EJ, Brook EL, Trutwein B, Rouse IL, Watson CR, de Klerk NH, Stanley FJ: **A decade of data linkage in Western Australia: Strategic design, applications and benefits of the WA data linkage system**. *Aust Heal Rev* 2008, **32**:766–777.

322. **CDRC - Obesity Network** [https://www.cdrc.ac.uk/research/obesity/]

323. Biehl A, Hovengen R, Grøholt E-K, Hjelmesæth J, Strand BH, Meyer HE: **Parental marital status and childhood overweight and obesity in Norway: a nationally representative cross-sectional study.** *BMJ Open* 2014, **4**:e004502.

324. Hampson SE, Tonstad S, Irgens LM, Meltzer HM, Vollrath ME: **Mothers' negative affectivity during pregnancy and food choices for their infants**. *Int J Obes* 2010, **34**:327–331.

325. Booth KM, Pinkston MM, Poston WSC: **Obesity and the built environment**. *J Am Diet Assoc* 2005, **105**(5 SUPPL.).

326. Villanueva K, Pereira G, Knuiman M, Bull F, Wood L, Christian H, Foster S, Boruff BJ, Beesley B, Hickey S, Joyce S, Nathan A, Saarloos D, Giles-Corti B: **The impact of the built environment on health across the life course: design of a cross-sectional data linkage study.** *BMJ Open* 2013, **3**:1–7.

327. Rodgers SE, Lyons RA, Dsilva R, Jones KH, Brooks CJ, Ford D V., John G, Verplancke JP: **Residential Anonymous Linking Fields (RALFs): A novel information infrastructure to study the interaction between the environment and individuals' health**. *J Public Health (Bangkok)* 2009, **31**:582–588.

328. Jones KH, Ford D V, Jones C, Dsilva R, Thompson S, Brooks CJ, Heaven ML, Thayer DS, McNerney CL, Lyons R a: **A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: a privacy-protecting remote access system for health-related research and evaluation.** *J Biomed Inform* 2014(January):1–9.

329. **SAIL Databank - The Secure Anonymised Information Linkage Databank** [https://saildatabank.com/saildata/sail-datasets/welsh-demographic-service-wds/]

330. Jones C, Brooks C, Arkley C, Jones C, Orton C, Mackey C, Mcnerney C, Thayer D, Brown G, Collins H, Hussain I, Peng J, Kennedy J, Biddle J, Au-yeung L, Ruschetti L, Cross L, Heaven M, Castagnet M, Bale M, Noyce R, Dsilva R, Lever R, Rees S, Rodgers S, Hindley S, Thompson S, Lee S, Lucas S: **SAIL 10 year anniverary brochure**. 2017.

331. **NHS Wales Informatics Service - Handling patient details** [http://www.wales.nhs.uk/nwis/page/52552]

332. **SAIL Databank - The Secure Anonymised Information Linkage Databank** [https://saildatabank.com/saildata/data-privacy-security/#protecting-identities]

333. Cummins S, Macintyre S: **Are secondary data sources on the neighbourhood food environment accurate? Case-study in Glasgow, UK**. *Prev Med (Baltim)* 2009, **49**:527–528.

334. **Local government in Wales - WLGA** [http://wlga.wales/local-government-in-wales]

335. **ANNEX B The Single Data List for 2017-18**. .

336. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, Moore L, O'Cathain A, Tinati T, Wight D, Baird J: **Process evaluation of complex interventions: Medical Research Council guidance**. *Bmj* 2015, **350**(mar19 6):h1258–h1258.

337. Jones I, Martin S, Whittington L: **Coping with the Cuts : Lessons from English Councils ' Responses to Budget Reductions Coping with the Cuts : Lessons from English Councils ' Responses to Budget Reductions**. 2015(December).

338. Defra: *INSPIRE Supports*. 2013.

339. Walter I: **Evidence Based Policy and Practice : Cross Sector Lessons From the UK Sandra Nutley**. 2002(July):2–3.

340. Pirog MA: **Data will derive innovation in public policy and management research in the next decade**. *J Policy Anal Manag* 2014, **33**:537–543.

341. **Food we can trust | Food Standards Agency** [https://www.food.gov.uk/about-us/about-the-fsa]

342. **Points of Interest** [https://www.ordnancesurvey.co.uk/business-and-government/products/points-of-interest.html]

343. **Frequently asked questions about the food hygiene rating scheme | Food Standards Agency** [https://www.food.gov.uk/multimedia/hygiene-rating-schemes/ratings-find-out-more-en/fhrs]

344. Ordnance Survey: **Points of Interest User guide**. 2012.

345. **Points of Interest | Business and government | Ordnance Survey** [http://www.ordnancesurvey.co.uk/business-and-government/products/points-of-interest.html]

346. **Food Standards Agency - Search for food hygiene ratings** [http://ratings.food.gov.uk/open-data/en-GB]

347. **Digimap** [http://digimap.edina.ac.uk/]

348. Larsen K, Buliung RN, Faulkner GE: **School travel route measurement and built environment effects in models of children's school travel behavior**. *J Transp Land Use* 2015, **2**:1–19.

349. Coulton CJ, Jennings MZ, Chan T: **How Big is My Neighborhood? Individual and Contextual Effects on Perceptions of Neighborhood Scale**. *Am J Community Psychol* 2013, **51**:140–150.

350. Fotheringham AS, Wong DWS: **The Modifiable Areal Unit Problem in Multivariate Statistical Analysis**. *Environ Plan A* 1991, **23**:1025–1044.

351. Buliung RN, Larsen K, Faulkner GEJ, Stone MR: **The "path" not taken: Exploring structural differences in mapped-versus shortest-network-path school travel routes**. *Am J Public Health* 2013, **103**:1589–1596.

352. Zwerts E, Allaert G, Janssens D, Wets G, Witlox F: **How children view their travel behaviour: A case study from Flanders (Belgium)**. *J Transp Geogr* 2010, **18**:702–710.

353. Kerr J, Rosenberg D, Sallis JF, Saelens BE, Frank LD, Conway TL: **Active commuting to school: Associations with environment and parental concerns**. *Med Sci Sports Exerc* 2006, **38**:787–794.

354. Stewart T, Schipperijn J, Snizek B, Duncan S: **Adolescent school travel: Is online mapping a practical alternative to GPS-assessed travel routes?** *J Transp Heal* 2017, **5**:113–122.

355. Schantz P, Stigell E: **A criterion method for measuring route distance in physically active commuting**. *Med Sci Sports Exerc* 2009, **41**:472–478.

356. Stigell E, Schantz P: **Methods for determining route distances in active commuting - Their validity and reproducibility**. *J Transp Geogr* 2011, **19**:563–574.

357. Broberg A, Sarjala S: **School travel mode choice and the characteristics of the urban built environment: The case of Helsinki, Finland**. *Transp Policy* 2015, **37**:1–10.

358. Chaix B, Kestens Y, Perchoux C, Karusisi N, Merlo J, Labadi K: **An interactive mapping tool to assess individual mobility patterns in neighborhood studies**. *Am J Prev Med* 2012, **43**:440–450.

359. **ALPHA - DECIPHer** [http://decipher.uk.net/public-involvement/young-people/]

360. **OS MasterMap ITN Layer | Business and government | Ordnance Survey** [http://www.ordnancesurvey.co.uk/business-and-government/products/itn-layer.html]

361. **Code-Point with polygons - locates every postcode unit in the UK with precision** [http://www.ordnancesurvey.co.uk/business-and-government/products/code-point-with-polygons.html]

362. **AddressBase Premium | Business and government | Ordnance Survey** [http://www.ordnancesurvey.co.uk/business-and-government/products/addressbase-premium.html]

363. Monk J: **Walking the Streets of London: Using Zadie Smith's "NW" to explore teenage, metropolitan ways of seeing and writing**. *English Educ* 2017, **51**:188–206.

364. **iTunes - Browse the top free apps on the App Store - Apple (UK)** [https://www.apple.com/uk/itunes/charts/free-apps/]

365. **Androidrank Android Market App Ranklist - android rating stats.** [http://www.androidrank.org/listcategory?category=&start=1&sort=0&price=all&hl=en]

366. Kiefer P, Giannopoulos I, Duchowski A, Raubal M: **Measuring Cognitive Load for Map Tasks Through Pupil Diameter**. In *Geogr Inf Sci 9th Int Conf GIScience 2016, Montr QC, Canada, Sept 27-30, 2016, Proc*. Edited by Miller JA, O'Sullivan D, Wiegand N. Cham: Springer International Publishing; 2016:323–337.

367. Hergan I, Umek M: **Comparison of children's wayfinding, using paper map and mobile navigation**. *Int Res Geogr Environ Educ* 2017, **26**:91–106.

368. Chillón P, Panter J, Corder K, Jones a. P, Van Sluijs EMF: **A longitudinal study of the distance that young people walk to school**. *Health Place* 2015, **31**(July 2007):133–137.

369. Cipeluch B, Jacob R, Winstanley A, Mooney P: **Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps**. In *Accuracy 2010 Symp*; 2010:337–340.

370. Downs RM, Liben L: **Fostering geographic literacy from early childhood: The control from early childhood: The contribution of interdisciplinary research**. *J Appl Dev Psychol* 1994, **15**:549–569.

371. Uttal DH: **Seeing the big picture: Map use and the development of spatial cognition**. *Dev Sci* 2000, **3**:247–264.

372. Korpilo S, Virtanen T, Lehvävirta S: **Smartphone GPS tracking—Inexpensive and**

efficient data collection on recreational movement. *Landsc Urban Plan* 2017, **157**:608–617.

373. Zhang M, Cheow E, Ho CSH, Ng BY, Ho R, Cheok CCS: **Application of Low-Cost Methodologies for Mobile Phone App Development**. *JMIR mHealth uHealth* 2014, **2**:e55.

374. Audrey S, Cooper AR, Hollingworth W, Metcalfe C, Procter S, Davis A, Campbell R, Gillison F, Rodgers SE: **Study protocol: the effectiveness and cost effectiveness of an employer-led intervention to increase walking during the daily commute: the Travel to Work randomised controlled trial**. *BMC Public Health* 2015, **15**:154.

375. **PostGIS — Spatial and Geographic Objects for PostgreSQL** [http://postgis.net/]

376. **pgRouting Project — Open Source Routing Library** [http://pgrouting.org/]

377. Mascheroni G, Ólafsson K: **The mobile Internet: Access, use, opportunities and divides among European children**. *New Media Soc* 2016, **18**:1657–1679.

378. Jones AP, Coombes EG, Griffin SJ, van Sluijs EM: **Environmental supportiveness for physical activity in English schoolchildren: a study using Global Positioning Systems**. *Int J Behav Nutr Phys Act* 2009, **6**:42.

379. **Public Health Wales - Welsh Network of Healthy School Schemes** [http://www.wales.nhs.uk/sitesplus/888/page/82249/]

380. **Healthier Checkout Lanes Coming to Aldi Supermarkets | Center for Science in the Public Interest** [https://cspinet.org/new/201601121.html]

381. Adjoian T, Dannefer R, Willingham C, Brathwaite C, Franklin S: **Healthy Checkout Lines: A Study in Urban Supermarkets**. *J Nutr Educ Behav* 2017, **49**:615–622.e1.

382. Winkler LL, Christensen U, Glümer C, Bloch P, Mikkelsen BE, Wansink B, Toft U: **Substituting sugar confectionery with fruit and healthy snacks at checkout – a win-win strategy for consumers and food stores? a study on consumer attitudes and sales effects of a healthy supermarket intervention**. *BMC Public Health* 2016, **16**:1184.

383. Graser A, Straub M, Dragaschnig M: **Towards an open source analysis toolbox for street network comparison: Indicators, tools and results of a comparison of osm and the official Austrian reference graph**. *Trans GIS* 2014, **18**:510–526.

384. Canavosio-Zuzelski R, Agouris P, Doucette P: **A Photogrammetric Approach for Assessing Positional Accuracy of OpenStreetMap© Roads**. *ISPRS Int J Geo-Information* 2013, **2**:276–301.

385. Haklay M, Weber P: **OpenStreet map: User-generated street maps**. *IEEE Pervasive Comput* 2008, **7**:12–18.

386. **Public Sector Mapping Agreement (PSMA)** [https://www.ordnancesurvey.co.uk/business-and-government/public-sector/mapping-agreements/public-sector-mapping-agreement.html]

387. **OpenStreetMap** [https://www.openstreetmap.org/]

388. Siebritz L, Sithole G, Zlatanova S: **Assessment of the homogeneity of volunteered geographic information in South Africa**. *22nd Congr Int Soc Photogramm Remote Sensing, ISPRS 2012* 2012, **39**(September):553–558.

389. Brovelli MA, Minghini M, Molinari M, Mooney P: **Towards an Automated Comparison**

of OpenStreetMap with Authoritative Road Datasets. *Trans GIS* 2017, **21**:191–206.

390. Embankment SB: **OS MasterMap ® Integrated Transport Network ™ (ITN) Layer Urban Paths Theme – coverage and technical information OS MasterMap Integrated Transport Network (ITN) Layer Urban Paths Theme – coverage and technical information**. 2010:1–16.

391. **OSM Wales** [http://download.geofabrik.de/europe/great-britain/wales.html]

392. **Rural and Urban Area Classification** [http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/rural-urban-definition-and-la/rural-urban-definition--england-and-wales-/index.html]

393. Dalton AM, Jones AP, Panter J, Ogilvie D: **Are GIS-modelled routes a useful proxy for the actual routes followed by commuters?** *J Transp Heal* 2014, **2**:1–11.

394. Estima J, Fonte CC, Painho M: **Comparative study of Land Use / Cover classification using Flickr photos , satellite imagery and Corine Land Cover database**. *Proc Agil Int Conf Geogr Inf Sci* 2014, **16**:205–220.

395. Neis P, Zielstra D: **Recent Developments and Future Trends in Volunteered Geographic Information Research: The Case of OpenStreetMap**. *Futur Internet* 2014, **6**:76–106.

396. Ma D, Sandberg M, Jiang B: **Characterizing the Heterogeneity of the OpenStreetMap Data and Community**. *ISPRS Int J Geo-Information* 2015, **4**:535–550.

397. Zhao P, Jia T, Qin K, Shan J, Jiao C: **Statistical analysis on the evolution of OpenStreetMap road networks in Beijing**. *Phys A Stat Mech its Appl* 2015, **420**:59–72.

398. **United Kingdom long distance paths - OpenStreetMap Wiki** [http://wiki.openstreetmap.org/wiki/United_Kingdom_long_distance_paths]

399. **Using OS OpenData to improve OpenStreetMap | Dan Karran** [https://www.dankarran.com/blog/2010/04/using-os-opendata-to-improve-openstreetmap]

400. Barrington-Leigh C, Millard-Ball A: **The world's open-source street map is more than 80% complete**. *PLoS One* 2017:1–21.

401. Davison KK, Werder JL, Lawson CT: **Children ' s Active Commuting to School : Current Knowledge and Future Directions**. *Rev Lit Arts Am* 2008, **5**:1–11.

402. Pont K, Ziviani J, Wadley D, Bennett S, Abbott R: **Environmental correlates of children's active transportation: A systematic literature review**. *Heal Place* 2009, **15**:827–840.

403. **PEAR project** [http://www.bristol.ac.uk/sps/research/researchprojectpages/pearproject/]

404. **OS MasterMap Topography Layer** [https://www.ordnancesurvey.co.uk/business-and-government/products/topography-layer.html]

405. Dr Foster Intelligence and Land Use Consultants: *Tackling the Takeaways: A New Policy to Address Fast-Food Outlets in Tower Hamlets*. London; 2011.

406. Mölter A, Lindley S: **Influence of walking route choice on primary school children's exposure to air pollution - A proof of concept study using simulation**. *Sci Total Environ*

2015, **530–531**:257–262.

407. **OS Open Rivers** [https://www.ordnancesurvey.co.uk/business-and-government/products/os-open-rivers.html]

408. **meridian 2** [http://digimap.edina.ac.uk/webhelp/os/data_information/os_products/meridian_2.htm]

409. **Road Safety Data - Datasets** [https://data.gov.uk/dataset/road-accidents-safety-data]

410. Hair JF, Anderson RE, Tatham RL, Black WC: *Multivariate Data Analysis*. Fifth. New Jersey: Prentice-Hall; 1984.

411. Bondy JA, Murty USR: **Graph theory with applications**. *Oper Res Q 19701977* 1976:270.

412. Dijkstra EW: **Comminication with an Automatic Computer**. 1959:1–174.

413. **Home to school travel and transport guidance Statutory guidance for local authorities**. 2014.

414. Schoeppe S, Duncan MJ, Badland H, Oliver M, Curtis C: **Associations of children's independent mobility and active travel with physical activity, sedentary behaviour and weight status: A systematic review**. *J Sci Med Sport* 2013, **16**:312–319.

415. Potoglou D, Arslangulova B: **Transportation Planning and Technology Factors influencing active travel to primary and secondary schools in Wales Factors influencing active travel to primary and secondary schools in Wales**. *Transp Plan Technol* 2016, **0**:1–20.

416. Jorgensen A, Hitchmough J, Calvert T: **Woodland spaces and edges: Their impact on perception of safety and preference**. *Landsc Urban Plan* 2002, **60**:135–150.

417. Özgüner H, Kendle AD: **Public attitudes towards naturalistic versus designed landscapes in the city of Sheffield (UK)**. *Landsc Urban Plan* 2006, **74**:139–1573.

418. Higgs S, Thomas J: **Social influences on eating**. *Curr Opin Behav Sci* 2016, **9**:1–6.

419. Coombes E, Jones A, Cooper A, Page A: **Does home neighbourhood supportiveness influence the location more than volume of adolescent's physical activity? An observational study using global positioning systems**. *Int J Behav Nutr Phys Act* 2017, **14**.

420. Lyons R a, Ford D V, Moore L, Rodgers SE: **Use of data linkage to measure the population health effect of non-health-care interventions.** *Lancet* 2013, **6736**:10–12.

421. Brophy S, Rees A, Knox G, Baker J, Thomas NE: **Child fitness and father's BMI are important factors in childhood obesity: A school based cross-sectional study**. *PLoS One* 2012, **7**:1–7.

422. **Childhood overweight and obesity statistics | Cancer Research UK** [http://www.cancerresearchuk.org/health-professional/cancer-statistics/risk/childhood-overweight-and-obesity#heading-One]

423. **WDS SAIL Databank - The Secure Anonymised Information Linkage Databank** [https://saildatabank.com/saildata/sail-datasets/welsh-demographic-service-wds/]

424. **LocatorHub | Esri UK** [http://www.esriuk.com/software/uk-products/LocatorHub]

425. **BMA - Join us** [https://www.bma.org.uk/join-

us?gclid=EAIaIQobChMIh8rzx7zh1wIVQ58bCh33TgcQEAAYASAAEgJMf_D_BwE]

426. **Health Research Authority** [https://www.hra.nhs.uk/]

427. Wales PH: **Public Health Wales - an Official NHS Wales website**. .

428. Hands BP, Chivers PT, Parker HE, Beilin L, Kendall G, Larkin D: **The associations between physical activity, screen time and weight from 6 to 14 yrs: The Raine Study**. *J Sci Med Sport* 2011, **14**:397–403.

429. Townsend N, Wickramasinghe K, Williams J, Bhatnagar P, Rayner M: **Physical activity statistics**. *Br Hear Found* 2015(January):1–128.

430. Reidpath DD, Burns C, Garrard J, Mahoney M, Townsend M: **An ecological study of the relationship between social and environmental determinants of obesity.** *Health Place* 2002, **8**:141–5.

431. Wales PH: **Public Health Wales - Welsh Network of Healthy School Schemes**. .

432. Seliske L, Pickett W, Rosu A, Janssen I: **The number and type of food retailers surrounding schools and their association with lunchtime eating behaviours in students**. *Int J Behav Nutr Phys Act* 2013, **10**:19.

433. **Facts on Pedestrian Casualties** [https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/448036/pedestrian-casualties-2013-data.pdf]

434. Boone-heinonen J, Gordon-larsen P, Kiefe CI: **Fast food restaurants and food stores: longitudinal associations with diet in young adults: The CARDIA Study**. 2012, **171**:1162–1170.

435. Duffey KJ, Gordon-Larsen P, Steffen LM, Jacobs DR, Popkin BM: **Regular Consumption from Fast Food Establishments Relative to Other Restaurants Is Differentially Associated with Metabolic Outcomes in Young Adults**. *J Nutr* 2009, **139**:2113–2118.

436. Pereira MA, Kartashov AI, Ebbeling CB, Van Horn L, Slattery ML, Jacobs PDR, Ludwig DS: **Fast-food habits, weight gain, and insulin resistance (the CARDIA study): 15-year prospective analysis**. *Lancet* 2005, **365**:36–42.

437. Faulkner GEJ, Buliung RN, Flora PK, Fusco C: **Active school transport, physical activity levels and body weight of children and youth: A systematic review**. *Prev Med (Baltim)* 2009, **48**:3–8.

438. Fonseca H, Silva AM, Matos MG, Esteves I, Costa P, Guerra A, Gomes-Pedro J: **Validity of BMI based on self-reported weight and height in adolescents**. *Acta Paediatr Int J Paediatr* 2010, **99**:83–88.

439. Magnusson K, Haugen IK, Østerås N, Nordsletten L, Natvig B, Hagen KB: **The validity of self-reported body mass index in a population-based osteoarthritis study**. *BMC Musculoskelet Disord* 2014, **15**:442.

440. Merrill RM, Richardson JS: **Validity of Self-Reported Height, Weight, and Body Mass Index: Findings from the National Health and Nutrition Examination Survey, 2001-2006**. *Prev Chronic Dis* 2009, **6**:A121.

441. **Welsh Government | Active travel: Walking and cycling** [http://gov.wales/statistics-and-research/active-travel/?lang=en]

442. **National Survey for Wales: Quality Report** [http://gov.wales/statistics-and-research/national-survey/?lang=en]

443. **School Health Research Network** [http://www.shrn.org.uk/]

444. Timperio A, Jeffery RW, Crawford D, Roberts R, Giles-Corti B, Ball K: **Neighbourhood physical activity environments and adiposity in children and mothers: a three-year longitudinal study**. *Int J Behav Nutr Phys Act* 2010, **7**:18.

445. Lohmann A, McMurran G: **Resident-Defined Neighborhood Mapping: Using GIS to Analyze Phenomenological Neighborhoods**. *J Prev Interv Community* 2009, **37**:66–81.

446. **Welsh Government | Well-being of Future Generations (Wales) Act 2015** [http://gov.wales/topics/people-and-communities/people/future-generations-act/?lang=en]

447. Paquet C, Daniel M: **Interactive effects of reward sensitivity and residential fast-food**. *Am J Clin Nutr* 2010:771–776.

448. **Welsh Government | Welsh Index of Multiple Deprivation** [http://gov.wales/statistics-and-research/welsh-index-multiple-deprivation/?lang=en]

449. Lloyd CD: **Spatial scale and small area population statistics for England and Wales**. *Int J Geogr Inf Sci* 2016, **30**:1187–1206.

450. Pires C, Agante L: **Encouraging children to eat more healthily: The influence of packaging**. *J Consum Behav* 2011, **10**:161–168.

451. **Welsh Government | Active Travel (Wales) Act 2013** [http://gov.wales/topics/transport/walking-cycling/activetravelact/?lang=en]

452. Chillón P, Evenson KR, Vaughn A, Ward DS: **A systematic review of interventions for promoting active transportation to school**. *Int J Behav Nutr Phys Act* 2011, **8**.

453. Audrey S, Batista-Ferrer H: **Healthy urban environments for children and young people: A systematic review of intervention studies**. *Heal Place* 2015, **36**:97–117.

454. Sarkar C, Webster C: **Urban environments and human health: current trends and future directions**. *Curr Opin Environ Sustain* 2017, **25**:33–44.

455. **PEAR: Physical Environment and Activity Relationships** [http://www.cedar.iph.cam.ac.uk/research/directory/pear/]

456. Obe RO, Hsu LS: *PgRouting: A Practical Guide*. Locate Press LLC; 2017.