Application of spatial behaviour in assessing food environment influence on health

- Integrating behaviour tracking with health, demographic and spatial data

Lyseen, Anders Knørr

DOI (link to publication from Publisher):
10.5278/vbn.phd.engsci.00046

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

APPLICATION OF SPATIAL BEHAVIOUR IN ASSESSING FOOD ENVIRONMENT INFLUENCES ON HEALTH

- INTEGRATING BEHAVIOUR TRACKING WITH HEALTH, DEMOGRAPHIC AND SPATIAL DATA

BY
ANDERS K. LYSEEN

DISSERTATION SUBMITTED 2016

AALBORG UNIVERSITY
DENMARK

APPLICATION OF SPATIAL BEHAVIOUR IN ASSESSING FOOD ENVIRONMENT INFLUENCES ON HEALTH

- INTEGRATING BEHAVIOUR TRACKING WITH HEALTH, DEMOGRAPHIC AND SPATIAL DATA

BY
ANDERS K. LYSEEN

AALBORG UNIVERSITY
DENMARK

DISSERTATION SUBMITTED FEBRUARY 2016

Dissertation submitted:	February 1, 2016
PhD supervisor:	Professor, dr.scient Henning Sten Hansen, *Aalborg University*
Assistant PhD supervisor:	Professor, PhD Bent Egberg Mikkelsen, *Aalborg University*
PhD committee:	Associate Professor, Head of Department Lars Bodum (Chairman), Aalborg University Professor Tuuli Toivonen University of Helsinki Professor, Dr. Hardy Pundt Hochschule Harz
PhD Series:	Faculty of Engineering and Science, Aalborg University

ISSN (online): 2246-1248
ISBN (online): 978-87-7112-492-7

Published by:
Aalborg University Press
Skjernvej 4A, 2nd floor
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2016

MANDATORY PAGE

Thesis Title
Application of spatial behaviour in assessing food environment influence on health
- *Integrating behaviour tracking with health, demographic and spatial data*

Name of PhD Student
Anders Knørr Lyseen

Name and Title of Supervisors
Henning Sten Hansen, Professor, dr.scient. (Main principal supervisor)
Bent Egberg Mikkelsen, Professor, PhD.

List of Papers

I. Lyseen, A. K. & Hansen, H. S. 2015. INSPIRE Compliance of Public Health Information – A Danish Case Study. *International Journal of Spatial Data Infrastructures Research* 10: 84-102.

II. Lyseen, A. K. & Hansen, H. S. 2014. Spatial and Semantic Validation of Secondary Food Source Data. *ISPRS International Journal of Geo-Information* 3, no. 1: 236-53.

III. Mikkelsen, B. E., Lyseen, A. K., Dobroczynski, M. & Hansen, H. 2014. Behavioural Nutrition & Big Data: How Geodata, Register data & GPS, Mobile Positioning, Wi-Fi, Bluetooth & Thermal Cameras can contribute to the Study of Human Food Behaviour. *Measuring Behavior 2014*, 9th International Conference on Methods and techniques in Behavioral Research, August 2014, Wageningen, The Netherlands.

IV. Lyseen, A. K., Hansen, H. S., Harder, H., Jensen, A. S. & Mikkelsen, B. E. Defining Neighbourhoods as a Measure of Exposure to the Food Environment. *International Journal of Environmental Research and Public Health* 12, no. 7: 8504-8525.

The papers are referred to in the text by their roman numerals.

RESUME IN ENGLISH

Globally, more people die from obesity than from malnutrition and underweight. A growing caseload of chronic illnesses, such as diabetes, cardiovascular diseases and cancers, are related to obesity and overweight. The growing number of overweight and obese people in the world is a major health concern for individuals and society. Proper eating habits combined with regular physical activity represent the most efficient method for preventing obesity and the development of lifestyle-related diseases. Knowledge of a population's health, societal and environmental influences, and behavioural patterns are highly valuable. However, countless factors influence humans' eating and physical activity behaviours. Personal factors, such as taste preference, convenience, knowledge, price, availability and accessibility, interact with the environment to influence food behaviour.

This study examines the application of spatial behaviour data in measuring exposure to the food environment. Second, the availability and integration of spatial data, demographic data and health data are examined. The socio-ecological model is used to provide a framework that demonstrates how different elements of society interact to shape individuals' eating and physical activity behaviours.

This study compares traditional methods of measuring exposure to the food environment with methods that use data from Global Positioning System (GPS) tracking. For each of the 187 participants, 11 different neighbourhoods were created in which exposure to supermarkets and fast food outlets were measured. A valid and complete food environment was established based on an analysis of two secondary food source datasets and a pre-classification method that minimised the time spent on field validation. The integration of behaviour data, spatial data, demographic data and health data are discussed, and current limitations are addressed.

In conclusion, this thesis shows that 'neighbourhood' clearly is a diffuse and blurred concept that varies in meaning depending on each person's perception and the study being conducted. Complexity and heterogeneity of human mobility no longer appear to correspond to the use of residential neighbourhoods; rather, these aspects of mobility emphasise the need for methods, concepts and measures of individual activity and exposure. The broad perspective of the socio-ecological model of obesity makes determination of relevant variables very complex. However, obtaining accurate and reliable data can be a lengthy and expensive process — one that is often confounded by privacy and confidentiality requirements, which need to be balanced against the need for high-quality analysis. Obtaining an overview of the available health data suitable for spatial applications is not easy. Most public health data do not have any spatial references, but these data should be linked to features with a spatial reference, for example, administrative units or addresses. According to Danish legislation, health information is private, which imposes great limitations on the use of health data.

Human health information should not be segregated, which is more or less the situation today; instead, it should be seamlessly integrated with other data.

Further efforts to anonymise data are required to ensure the privacy and confidentiality of health data. Spatial aggregation is key to making data anonymous. The lack of initiative to include spatial applications of health data in the agenda for the digitisation of health data hinders the integration of health data in a spatial data infrastructure. The issues of spatially aggregated health data and compliance must be addressed. The harmonisation and implementation of health data are unstructured and slowed by the lack of strategic promotion of spatial health data.

RESUME IN DANISH

Der er flere, der dør i verden af fedme end af fejlernæring og undervægt. Overvægt og fedme er skyld i et stigende antal tilfælde af kroniske sygdomme, som diabetes, hjerte-kar-sygdomme og cancer. Det stigende antal af overvægtige og svært overvægtige personer i verden er et kæmpe problem for såvel individer som for samfundet. Den mest effektive metode til forebyggelse af fedme og livsstilsrelaterede sygdomme er fornuftige spisevaner kombineret med regelmæssig motion. Viden om befolkningens sundhed, indflydelse fra miljø og samfund samt adfærdsmønstre er meget værdifulde, men der findes mange andre faktorer, som har indflydelse på personers spise- og motionsvaner. Personlige faktorer, som smagspræferencer, bekvemmelighed, viden, pris, tilgængelighed og udbud, interagerer med det fysiske miljø og har indflydelse på fødevarerelateret adfærd.

Dette studie undersøger anvendelsen af geografisk adfærdsdata til at måle personers eksponering til fødevareudbuddet. Sekundært undersøges tilgængeligheden af geografiske data, demografiske data og sundhedsdata, samt muligheden for at sammenstille disse. Som ramme er en socio-økologisk model anvendt. Den demonstrerer, hvordan forskellige elementer i samfundet interagerer og former individers spise- og motionsadfærd.

Studiet sammenligner traditionelle metoder for måling af eksponering til udbuddet af fødevarer med metoder, som anvender data fra overvågning med Global Positioning System (GPS). For hver af de 187 deltagere er der beregnet 11 forskellige 'neighbourhood', hvori eksponering til supermarkeder og fastfood restauranter er målt. To datakilder over fødevareudbydere er analyseret og danner baggrund for at etablere et validt og komplet datasæt over det geografiske fødevareudbud. En metode til før-klassifikation af fødevareudbydere minimerede tiden anvendt på feltvalidering. Integrering af adfærdsdata, geografisk data, demografisk data og sundhedsdata er diskuteret og nuværende begrænsninger er adresseret.

Studiet konkluderer, at 'neighbourhood' er et uklart og diffust koncept, hvis betydning varierer afhængig af individers opfattelse og de udførte studier. Menneskelig adfærd og bevægelighed er kompleks og heterogent og kan ikke længere begrebsliggøres ved brug af beboelsesområder. Bevægeligheden tyder snarere på et behov for metoder, koncepter og målinger af personlig aktivitet og eksponering. Det bredt favnende perspektiv i den socio-økologiske model for fedme gør det komplekst at fastsætte relevante variable. Sekundært er det en langsommelig og potentielt omkostningsfuld proces at indsamle nøjagtige og pålidelige data, hvilket yderligere kompliceres af krav om fortrolighed, og skal balanceres mod et behov for analyser af høj kvalitet. Det er vanskeligt at opnå et overblik over tilgængelige sundhedsdata, som kan anvendes til geografiske analyser. De fleste offentlige sundhedsdata har ikke nogen geografisk reference, men skal sammenkædes med elementer med en geografisk reference, som f.eks. administrative

enheder eller adresser. Ifølge dansk lovgivning er sundhedsdata fortrolige og fastsætter derved mange begrænsninger for brugen af sundhedsdata. Sundhedsdata bør ikke være adskilt fra andre data, hvilket mere eller mindre er situationen i dag, men i stedet nemt kunne sammenkædes med andre data.

For at sikre fortroligheden ved brug af sundhedsdata, er yderligere foranstaltninger krævet for at anonymisere data. Geografisk aggregering er nøglen til at anonymisere data. CPR-nummeret sikrer en sammenkædning af sundhedsdata med adressen, der sikrer den geografiske reference. Manglende initiativer på agendaen for digitalisering, som inkluderer geografisk anvendelse af sundhedsdata, begrænser integreringen af sundhedsdata i geografisk datainfrastruktur. Det skal adresseres, hvordan problemerne med aggregering af sundhedsdata og integrering til andre data løses. Generel harmonisering og implementering af sundhedsdata er ustruktureret og langsommeliggjort på grund af manglende initiativer og strategisk prioritering af geografiske sundhedsdata.

ACKNOWLEDGEMENTS

I wish to thank my supervisors, Henning Sten Hansen and Bent Egberg Mikkelsen, for their scientific expertise, constructive conversations and motivation. A big thanks goes to Henning Sten Hansen for his tireless and never-ending encouragement, especially during the last year, as well as for his ideas and comments relative to the scientific work.

A big thanks goes to my other colleagues at Aalborg University—Thomas, Morten, Rikke, Sune, Laurits and Henrik—for enjoyable conversations and professional discussions.

Last, but not least, I would like to thank my wife Mette because, without her continuous support, encouragement and inspiration, completion of this thesis would never have been possible. She means more than anything to me, and I owe her everything for the smiles and hugs she gave me in hard times.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

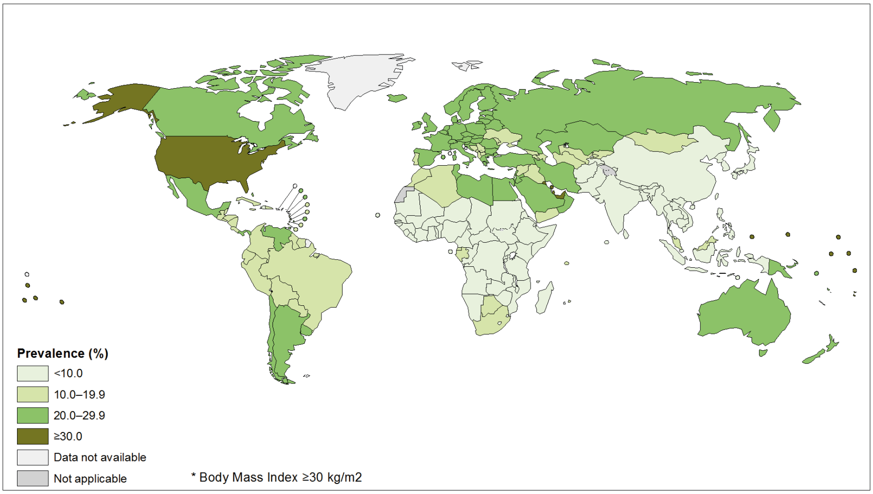
1.1 OBESITY AND THE PHYSICAL ENVIRONMENT

More than a decade ago, the number of overweight people in the world surpassed the number of underweight people [Gardner & Halweil, 2000]. Worldwide obesity has doubled since 1980. The World Health Organisation (WHO) estimates the overweight population (BMI > 25) today numbers more than 1.9 billion adults (18 years and older), with over 600 million of these being obese (BMI > 30) [WHO (1), 2015]. Many developing countries' healthcare systems are still struggling to control infectious diseases, but due to overweight and obesity advancing rapidly in these same developing countries, they also face a growing caseload of chronic illnesses, such as diabetes, cardiovascular diseases and cancers [Gardner & Halweil, 2000; Abbade et a., 2015]. However, the distribution of obese and underweight people shows clustering in different continents and countries, as presented by the obesity prevalence maps in Figure 1 [Gardner & Halweil, 2000]. Over- and underweight often leads to malnutrition, and in 2000, the World Health Organisation estimated that half of the earth's population suffers from malnutrition in some form [Gardner & Halweil, 2000].

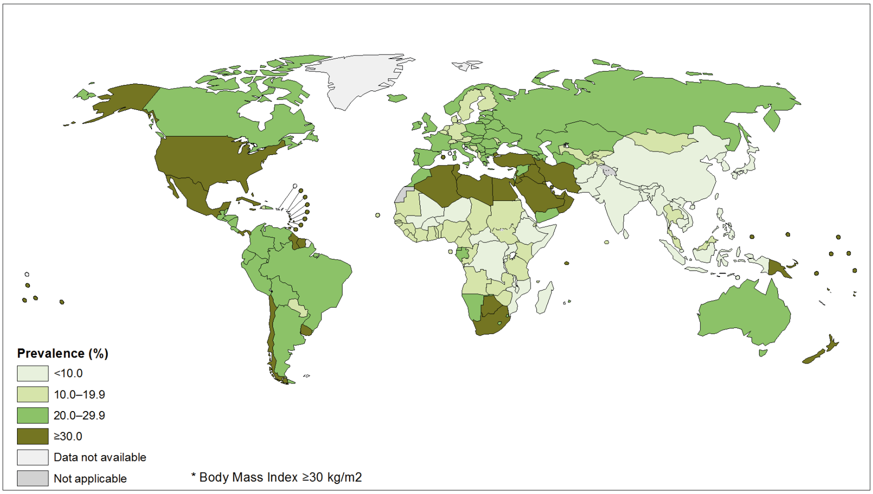
The growing number of overweight and obese people in the world is a major health concern [Wang et al., 2011]. Consequences of overweight and obesity include premature mortality, morbidity associated with chronic diseases and negative consequences for well-being and quality of life. The burden to individuals and society is mainly carried by an increased risk of cardiovascular diseases [Guh et al., 2009; Eeg-Olofsson et al., 2009; Rajkovic et al., 2014; Vázquez et al., 2014], type 2 diabetes [Guh et al., 2009; Arafat et al., 2014; Eeg-Olofsson et al., 2009; Tyrovalis et al., 2015] and several types of cancer [Renehan et al., 2008; Guh et al., 2009; Zhang et al., 2014]. However, overweight and obesity are also associated with an increased risk of asthma [Guh et al., 2009; Wang et al., 2008], infertility [Withrow & Alter, 2011] and osteoarthritis [Guh et al., 2009].

Obesity and overweight incur extensive costs to society, including economically and to the individuals affected, as well as impacting the individual's well-being. Research suggests that malnutrition has a huge cost for society because of how it hampers countries' development by lowering children's performance at school and decreasing worker productivity [Gardner & Halweil, 2000; Wang et al., 2011]. Expenditures related to obesity and overweight include not only the healthcare cost attributable to obesity-related diseases but also loss of productivity. Many countries are analysing the expenditures associated with obesity and overweight. In 2001, the economic costs of obesity in Canada totalled \$4.3 billion, where \$1.6 accounted for direct healthcare costs and \$2.7 for indirect costs (for example, lost productivity) [Katzmarzyk & Janssen, 2004]. In 2006-07, more than 21% of the total National Health Service costs in the UK (£5.8 billion) were used to treat poor-diet-related ill health, and approximately £5.1 billion

Prevalence of obesity*, ages 18+, 2014 (age standardized estimate)
Male



Prevalence of obesity*, ages 18+, 2014 (age standardized estimate) Female



The boundaries and names shown and the designations used on this map do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

World Health Organization

Data Source: World Health Organization Map Production: Health Statistics and Information Systems (HSI) World Health Organization

Figure 1. Distribution of obesity prevalence for (top) male and (bottom) female across countries.

was spent on overweight- and obesity-related treatment [Scarborough et al., 2011]. For 1995, the direct costs of inactivity and obesity were estimated to account for 9.4

percent of USA healthcare expenditures [Colditz, 1999]. Based on the 2008 National Health and Wellness Survey, a model by Wang and colleagues estimated a loss of 1.7-3 million productive person-years for working adults in the USA, which could represent an estimated cost as high as $390-580 billion [Wang et al., 2011]. Healthcare costs caused by overweight and obesity in the USA are projected to double every decade and account for 16-18 percent of health care costs in 2030 [Wang et al., 2008]. The combined costs are estimated to increase by $48-66 billion/year in the USA and by £1.9-2 billion in the UK by 2030 for the treatment of these preventable diseases [Wang et al., 2011]. Diabetes alone is estimated to account for $612 billion in health expenditures worldwide, representing 11% of total healthcare costs for adults in 2014 [International Diabetes Federation, 2014]. Withrow and colleagues estimate that obesity in general accounts for between 0.7 and 2.8 percent of a country's healthcare costs [Withrow & Alter, 2011]. As the individual country examples demonstrate, despite several models having been used to estimate the expenditures associated with overweight and obesity, all yield enormous cost estimates. Most countries' healthcare systems are facing increasing expenditures in the coming years because of the growth of the elderly population and more people becoming obese and developing chronic diseases. These trends put pressure on national healthcare systems and on local healthcare service providers. To change this negative trend, interventions that generate healthier habits have to be developed to reduce the prevalence of chronic diseases. Wang and colleagues considered a hypothetical example wherein the entire US population made a 1% reduction to their Body Mass Index (BMI) (roughly equal to 1 kg for an average-weight adult). Hypothetically, this would avert an estimated 2.1-2.4 million cases of diabetes, 1.4-1.7 million cases of cardiovascular disease and 73,000-127,000 cancer cases. Hence, stagnation or a decrease in overweight and obesity would contribute to a healthier and richer life for many people and a major reduction in healthcare costs.

Proper eating habits combined with regular physical activity represent the most efficient method for preventing obesity and the development of lifestyle-related diseases [Wing et al., 2001; Glanz et al., 2008; Abbade et al., 2015; Cordain et al., 2005]. Weight gain and obesity is obviously a problem reflecting excessive consumption of energy compared with the amount of energy expended through physical activity, but how to effectively change populations' unhealthy eating behaviours and sedentary behaviours remains unclear [Wing et al., 2001]. The promotion of healthy eating and physical activity often focuses on making healthier alternatives accessible, but individuals are often still in a position where they can choose between healthy and unhealthy alternatives. Unhealthy options are often presented as the easiest choices, whereas healthy alternatives seldom are presented in a way that makes the rational and healthier choices easy for consumers. Physical activity is proven to be healthy for metabolism, but because people cannot be forced to exercise, other solutions that encourage physical activity are needed. Many initiatives without a geographic component have been developed to promote healthy nutrition (e.g., by making consumers aware of caloric intake by using applications for mobile phones that encourage healthy eating).

Nudging is the newest term in shaping people's behaviour. Nudging can include a variety of approaches aimed at altering the social or physical environments to make certain behaviours more likely. Nudging is an effective tool for shaping environments that promote certain behaviour. Unfortunately, nudging also works to promote unhealthy eating, e.g., by placing chocolate by the checkout tills in supermarkets. Nudging can make us aware of the role that the social and physical environments play in shaping our health behaviour. However, evidence that examines the effectiveness of nudging on improving health needs further research [Marteau et al., 2011]. Therefore, knowledge of a population's health, societal and environmental influences, and the behavioural patterns preceding and following nudging interventions, are highly valuable. However, countless factors influence humans' eating and physical activity behaviours. Personal factors, such as taste preference, convenience, knowledge, price, availability and accessibility, interact with the environment to influence food behaviour. Individuals are born with certain genes, making them more or less susceptible to certain diseases. In the course of an individual's life, his or her socioeconomic status, social conditions and relations affect the individual's living conditions, which again influence the individual's lifestyle in terms of housing, mobility, nutrition, stimulation, etc. Whether the health risks are a result of one's genes, childhood, socioeconomic status, social conditions or relations, living conditions or lifestyle, reducing the number of negative impacts is important and challenging. This demonstrates the multiple factors that influence eating and physical activity behaviours. Capable of examining multiple levels of factors for public health promotion, the socio-ecological model is often used to provide a framework that demonstrates how different elements of society are combined to shape individuals' eating and physical activity behaviours [Fielding et al., 2010; Glanz et al., 2008]. The socio-ecological model provides a broader conceptualisation of important determinants of, in this case, obesity and overweight. However, the broad perspective of the socio-ecological model of obesity makes determination of relevant variables very complex [Abbade et al., 2015]. Second, information on the many factors can prove difficult to obtain at equivalent spatial and temporal levels, and many institutions and authorities might need to be involved in obtaining data.

1.2 DATA INTEGRATION AND COMPATIBILITY

Health, socioeconomic and spatial data are examples of the information that can be combined in the socio-ecological model to examine the influence of the physical environment on obesity. However, obtaining accurate and reliable data can be a lengthy and expensive process—one that is often confounded by privacy and confidentiality requirements, which need to be balanced against the need for high-quality analysis. The availability of data varies by country. In Scandinavia, data are readily available on individual or small-scale aggregate levels, and simple rules for anonymisation and ethics regulate research performed using these data. In many other countries, only highly aggregated data are available, and strict legislation on the use of the data can complicate accessing it. Furthermore, exclusion of key data sets due to privacy concerns or data being withheld due to commercial or political issues skews research results.

The INSPIRE Directive emphasises the availability of spatial data in Europe. The INSPIRE Directive establishes common rules, conditions and standards for spatial data, services and metadata. This provides the framework for better access and increased application of spatial data. Digitisation of public spatial data are on the agenda for many governments, and each country is developing their own national Spatial Data Infrastructures (SDI) [EDINA]. Commonly, national SDIs focus on traditional spatial data, such as addresses, property information, spatial planning, remote sensing data and environmental data. However, several countries worldwide are also digitalising their healthcare systems, and health data often applies INSPIRE principles [European Commission]. A few countries are taking this a step further by developing spatial data infrastructures that include health data; Denmark has only a few health datasets available [Danish Geodata Agency]. In the UK, an environment and health atlas [SAHSU] has been published. Additionally, the SDI Go-Geo [EDINA] delivers metadata, spatial data and interactive mapping to medical researchers, public health officials and the general public [Mathys and Boulos, 2011]. New Mexico has developed a web portal to integrate environmental information and health information [EPHT]. In Victoria, Australia, spatial data access and management is a priority, and an SDI has been developed to increase and strengthen effective collaborations with health projects. This further adds benefits through the increased use of under-used data [Thompson et al, 2009].

Joining health, socioeconomic and spatial datasets requires a common key attribute [Hansen, 2001]. Personal identifiers and addresses are vital key attributes for the integration of health, social and spatial data. Addresses are important and unambiguous database keys in many private and public registers in Denmark, and all Danish addresses have a spatial reference. Addresses are also registered in the Civil Person Register (CPR) [Pedersen, 2011], which contains information on individuals in Denmark. The CPR is key for linking health data [Bjerregaard and Larsen, 2011], social data, labour market data [Petersson et al, 2011] and education data [Jensen and Rasmussen, 2011]. Joining health, socioeconomic and spatial data to people's behaviour requires information about their whereabouts, choices and activities during the day. Traditionally, nutritional research has used food diaries, questionnaires and interviews to analysis of food behaviour [Kestens et al., 2012]. This is not unique to food studies; transport research also has previously frequently applied travel diaries, but implementation of tracking technologies has reduced some of the shortcomings, such as poor data quality, neglect of reporting short trips, total trip times and destination locations [Rainham et al., 2010; Murakami & Wagner, 1999]. Information on whereabouts combined with secondary food source data, including both governmental and commercial lists, can be used to measure the influence of food environments and foodscapes within health and nutritional research [Mikkelsen, 2011; Moore & Diez Roux, 2006; Neckerman et al., 2010; Sturm et al., 2008; Lytle, 2009; Glanz, 2009; Wang et al., 2007]. Measures of the food environments are needed to fully understand the relationship between these environments and obesity [Kelly et al., 2011].

Tracking devices and derived data are part of the new era of big data in research. Big data are generally referred to as huge volumes of data that are not handled by the usual methods and tools [Hampton et al., 2013]. Big data provide opportunities for advancing research across several sciences. Portable intelligent devices have created a series of new opportunities for convenient assessment of food choice and, as a result, food- and lifestyle-related behaviour is increasingly becoming the subject of measurement through the application of such mobile devices [Ngo et al., 2009; Thompson et al., 2010]. Intelligent devices, such as smartphones and touch pads, are increasingly used by consumers not only to get online using different wireless technologies but also for self-monitoring of lifestyle. For the research community, this development also offers new opportunities, insofar as the devices can also be used in a reverse mode to track the behaviour of individuals applying Global Positioning System (GPS), mobile positioning, Wi-Fi and Bluetooth. The collection of behaviour information leads to the creation of a very huge dataset, which induce problems for analysis. Hence, changing from place-based polygon features to people-based point features requires different analysis and an increased focus on data administration and cleaning.

Vast amounts of health, socioeconomic and spatial data are produced and stored in many countries. The availability of data varies and can be a lengthy and costly process. However, data reliability, methods for data handling and making sense of the hidden interconnected patterns of data and information are the primary obstacles. Application of big data requires new algorithms and methods for exploiting information in various applications, including health, smart cities and urban planning. On the other hand, the advanced development of space-time analysis and techniques provides a unique opportunity for discovering uncovered information at a near-real-time scale. Functionalities related to geographical information systems (GIS) offer simple representations of the physical environment, including its opportunities for physical activity and food.

1.3 RESEARCH AIMS

For more than two thousand years since Hippocrates' 'On Airs, Waters and Places', place has played an important role in understanding health and disease [Hippocrates, 1734]. Even two thousand years ago, Hippocrates already exhibited an understanding of the importance of cultural and environmental interactions on health, which were as important for studying human diseases and the promotion of health then as they are today. The population has changed significantly since Hippocrates, insofar as the population has become older, more obese and more sedentary. Both the agricultural revolution occurring 10,000 years ago and the industrial revolution beginning in the late 18th century have, through technological and economic innovations, influenced how and what we eat [Braidwood, 1960; Mantoux, 2013; Cowan, 1976]. The consequences on our health for some innovations have been positive, whereas other innovations have performed more harm to the populations' health.

Our current environment, characterised by cheap and readily available foods, promotes overeating and a higher consumption of energy-dense food, which is represented by large amounts of refined sugar and white flour [Ruhm, 2012]. Our living conditions are changing as more people move to urban areas and there is an abundance of food in many parts of the world [Abbade et al., 2015]. Migration to the city is estimated to possibly double the intake of sweeteners per capita. An example of the consequences of such migration is visible in China, where 23 percent of the urban Chinese population are overweight, compared with just over 10 percent of rural Chinese, as of 2000 [Gardner & Halweil, 2000]. Our living conditions have improved, and we enjoy innovations, such as cars, televisions and computers. However, ownership of one or more of these household devices is associated with obesity and type 2 diabetes [Lear et al., 2014]. Urbanisation, technological developments, a plethora of available food, radical changes in dietary habits through the 20th century and a sedentary lifestyle strongly influence people's behaviour and lead to growing numbers of overweight and obese people.

Environmental, policy and societal factors related to eating and physical activity behaviours are important contributors to the increased number of obese people [Sallis & Glanz, 2009]. Through changes to the physical environment, researchers have examined which factors influence healthy and unhealthy eating behaviours [French et al., 2001]. The food environment influences the types and amounts of food available and the opportunity for healthier dietary choices [McKinnon et al., 2009; Pearce et al., 2009]. Insights into food environments and nutritional behaviour can yield improvements in population health and nutrition [Thornton et al., 2012]. Local food environments have proven to be an indicator of the food behaviour of individuals [Mikkelsen, 2011; Moore & Diez Roux, 2006]. Studies in the USA, Canada, New Zealand and the UK indicate that the way cities are planned in terms of the food environment affects the behaviour of people in the local community and their health outcomes, such as BMI [Leung et al., 2011; Rossen et al., 2013; Oreskovic et al., 2009; Wang et al., 2007], body weight [Burgoine et al., 2014; Pearce et al. 2009], obesity [Jeffery et al., 2006; Liu et al., 2007], and diet [Thornton et al., 2012; Pearce et al. 2009; Laxer, 2013]. There are differences in the way cities are planned in different parts of the world, e.g., between North America and Europe. Even within Europe, differences between the way cities are planned and the structure of the food environment are evident, e.g., between Northern and Southern Europe.

Exposure to an environment is often conceptualised through and measured within neighbourhoods. However, the spatial area that defines a neighbourhood has proven difficult for researchers to define, with the result being large variation in the definitions of 'neighbourhood' used to study environmental exposure [Lytle, 2009]. Neighbourhoods are often used as the spatial element relative to people's behaviour. However, a conversion from place-based to people-based definitions of exposure measurement are occurring as tracking technologies gain ground in research [Burgoine et al., 2014; Burgoine et al., 2013; Odoms-Young et al., 2009; Kerr et al., 2011; Chaix et al., 2013]. Tracking technologies provide additional data that can be applied to examining the associations between the food environment and obesity.

The overall research objective of this thesis *is to examine the application of spatial behaviour in assessing food environment influences on health through integration with health, demographic and spatial data*. To achieve these objectives, the following four research issues will be addressed:

- Analysis of semantic and spatial validity of secondary food source data;
- Analysis of health data compliance with INSPIRE;
- Analysis of tracking technologies as applied to nutritional research; and
- Analysis of spatial exposure measures and the applicability of GPS.

1.4 THESIS ORGANISATION

The studies in Papers I, II and VI are based on data from Danish registers or data gathered on Danish population samples. In Paper III, the study takes a broader perspective; however, the point of origin remains the set of experiences occurring within the Danish context.
Figure 2 illustrates the organisation and the context of the papers in the thesis. The first section consists of an analysis of data availability, validity and interoperability by combining health, socioeconomic and spatial data in a geographical context. The second section consists of conceptualising the physical/food environment's association with people's behaviour and, essentially, its effect on peoples' health.

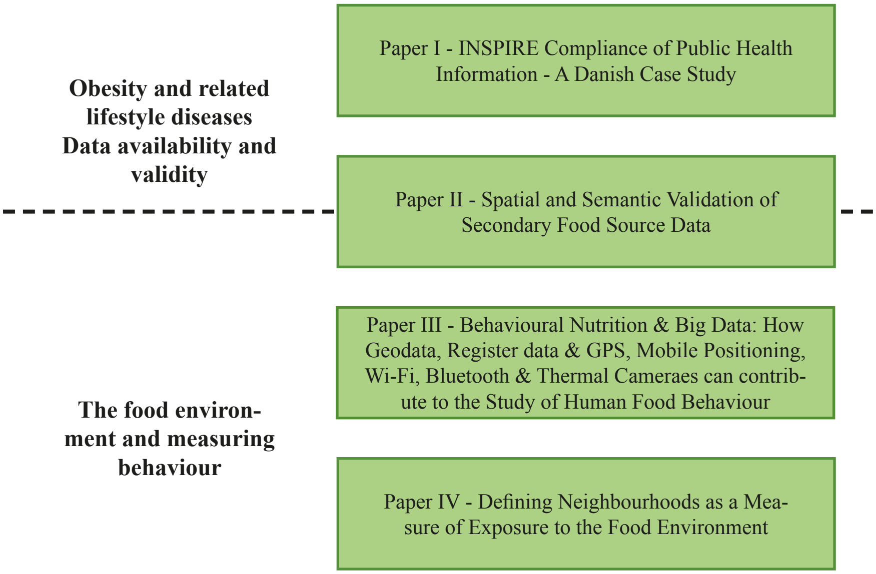


Figure 2. Organisation of the papers in the thesis.

Paper I analyses health data compliance with INSPIRE and the possibilities of combining health data, spatial data and socioeconomic data in Denmark. Some perspectives are extensible to other European countries because the INSPIRE directive is integrated into all European countries' spatial data infrastructures. Furthermore, the INSPIRE principles are valid guidelines for the efficient management of all types of data, not merely spatial data, thereby providing interoperability across datasets. Paper I comprehensively identifies the vast breadth of registers in Denmark that contain information on health and the possibilities for linking health data to socioeconomic data on the individual and aggregate levels. Health data are subject to restrictions because of privacy and confidentiality issues, which are discussed along with methods for making health data available in spatially aggregated form.

Methods for data handling and making sense of the hidden interconnected patterns in data and information requires insight into how the data are created and the validity of the information being registered. Concepts such as entity completeness, semantic accuracy and spatial accuracy are all part of defining the validity of data. All three concepts influence the applicability of the data. Data on the food environment are examined in Paper II. Secondary food source data are compared with field observations, and their completeness, semantic accuracy and spatial accuracy are analysed. Food environment data, together with data on the spatial distributions of obesity, diabetes and cardiovascular disease, are examples of data that are incorporated into the INSPIRE Directive's Human Health and Safety theme.

Place has traditionally provided the conceptual and analytic platform for studying the physical environment and its relation to food behaviour. Neighbourhoods are the predominant concept for studying the influence of the food environment. Human activities were previously closely attached to the locations of home and work. However, the perspective on human activities has shifted: complexity and heterogeneity of human behaviour today no longer appear to be confined to fixed spaces but, instead, require a more individualised conceptualisation. Paper III provides an overview of the available tracking technologies that can provide individual measures of activity patterns, and the paper offers an assessment of the pros and cons for different tracking technologies and application setups. In Paper VI, eleven different definitions of neighbourhoods are compared, including traditional place-based definitions and individual definitions based on GPS measurements. Food environment data from Paper II are used for an analysis of the differences between exposures to supermarkets and fast food outlets between neighbourhood definitions.

The topics of the papers presented in this thesis might seem heterogeneous, but the common dominator is that they all contribute to an examination of the puzzling relations between obesity-related diseases and people's behaviour in the food environment. With an ageing population and a growing number of obese people with lifestyle-related diseases, the relations between individual behaviour, place and health become more important for improving population health and reducing the costs of health care.

2. BACKGROUND

2.1 THE ECOLOGICAL PERSPECTIVE ON OBESITY

The ecological model is based on the principle that a single factor is not enough to explain why some people are more likely to, for example, become obese. The ecological model can enhance understanding of the dynamic relations between various individual and environmental factors. The ecological framework was initially introduced by Bronfenbrenner, who hypothesised that to understand human development, the entire ecological system in which humans grow and live needs to be taken into account [Bronfenbrenner, 2009]. The ecological model is composed of organised levels in a nested structure. The innermost level includes individual factors, such as age, gender, knowledge and preferences. The next level includes environmental features and the interconnections occurring within the environmental setting. The third level evokes the hypothesis that human development is affected by events occurring in settings where the person is not present. Last, social and cultural norms and values relate to all three aforementioned levels: for example, the cultural norms of home cooking and eating out vary among countries. An ecological model of obesity is presented in Figure 3.

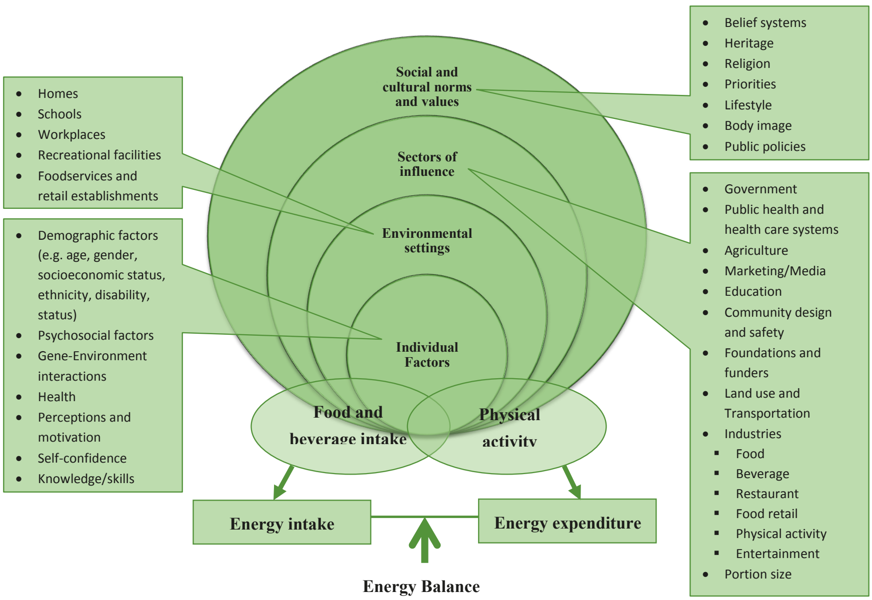


Figure 3. Ecological model of obesity. Inspired by the U.S. Department of Agriculture and U.S. Department of Health and Human Services. [U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2010]

The ecological model often combines factors such as individual, environmental settings, sectors of influence, and social and cultural norms and values [Glanz et al., 2008]. The physical environment is a concept that can belong to multiple levels, and changes here are expected to affect entire populations. Gathering data for all possible parameters in an ecological model is a daunting task that seems nearly impossible. However, in a Danish context, some data are available through registers, such as data on health, demographics, and locations of schools, industries, food retailers, etc. Information about individuals' preferences, perceptions and motivations are more difficult to obtain population-wide because these require interviews or questionnaires for each person.

People's behaviours are not traditionally included as parameters in the ecological model. However, parameters from all four levels in the ecological model influence people's behaviours in some way. Individual factors are the primary influence on an individual's behaviour, whereas environmental factors, such as the home neighbourhood, access to food retailers and the availability of food, indirectly influence food-related behaviour. Similarly, sectors influence individuals' food-related behaviours, such as the food industry's advertising and marketing, educational information and agricultural campaigns. The society in which an individual lives, which includes culture, religion and heritage, has a large indirect influence on eating behaviour.

The ecological model has been applied to examine the association between neighbourhood, physical activity and the food environment and weight status [Kegler et al., 2013; Christian et al., 2011; Kegler et al., 2012; Galvez et al., 2010]. The food environment influences the types and amounts of food available. Research has shown that food supply and intake, physical activity level and educational level influence the obesogenic situation [Abbade et al., 2015]. However, research on environmental influences on obesity has tended to focus more on physical activity than on food supply and intake. Sallis and colleagues proposed a socio-ecological model [Sallis et al., 2006] that they used to locate highly walkable neighbourhoods that were associated with physical activity, walking and biking [Van Dyck et al., 2012]. Access to parks, playgrounds and gyms in the local environment is associated with increased exercise for students [Caroll-Scoot et al., 2013]. An ecological approach to health promotion showed that change in the physical environment was associated with changes in health risk [Cohen et al., 2000].

2.2 THE FOOD ENVIRONMENT

The physical food environment includes places where food can be acquired, such as supermarkets, bakeries and restaurants. The food environment influences the types and amounts of food available and the opportunity for choosing a healthier diet [McKinnon et al., 2009; Pearce et al., 2009]. In theory, the food environment includes all food sources in a country. However, for an individual, the daily food environment only covers a geographical area whose boundaries tend to vary only slightly over time.

Reliable and valid measures of food environments are needed to fully understand the relationship between the food environment and health [Kelly et al., 2011]. Secondary food source data, including both governmental and commercial lists, are commonly used to measure food environments and foodscapes within health and nutritional research [Thornton et al., 2012; Mikkelsen, 2011; Neckerman et al., 2010; Sturm, 2008; Lytle, 2009; Glanz, 2009; Wang et al., 2009]. Knowledge of the validity of the secondary food source data are needed to fully understand the potential and limitations of its application. Hence, the analysis of, results of and conclusions based on secondary food data sources are influenced by completeness, thematic accuracy, geographical accuracy and contemporaneity. Unknown errors in food source data lead to misinterpretation of the results or under- and overestimation [Cummins & Macintyre, 2009] of, for example, the density of food retailers or the analysis of its associations with health and socioeconomic factors.
To conceptualise whether food retailers promote healthy or unhealthy eating behaviours, the dietary advice is taken into account. Therefore, fast-food outlets are often associated with unhealthy eating behaviours because of the high contents of saturated fat and salt and low content of wholemeal. Supermarkets are perceived to be the healthy counterpart because of the availability of varied fruit and vegetables supplies and lean meat products.

In 2013, the Danish Health and Medicines Authority updated their advice for a healthy diet [Danish Health and Medicines Authority, 2013]. Its diet advice is constructed to ensure that people eat the required nutrients in appropriate quantities. This ensures a healthy well-being that also prevents the development of lifestyle-related diseases. This diet advice provides guidelines for eating varied and greater amounts of fruit and vegetables, less saturated fats, and less salt and for choosing whole-grain and lean meat products. Similar dietary guidelines are found in many countries, e.g., in the United States of America [United States Department of Agriculture and U.S. Department of Health and Human Services, 2010].

2.3 NEIGHBOURHOODS AS A PROXY FOR MEASURING BEHAVIOUR

Applying neighbourhoods in the measurement of food exposure creates a manageable concept on which to analyse the effect of the exposure. However, variations in the definition of 'neighbourhood' show that not all definitions manage to conceive and measure the actual exposure equally well [Thornton et al., 2012; Burgoine, 2013]. Giles-Corti and colleagues evidenced this discrepancy upon finding little agreement in the literature on what an appropriate distance from home, work or school is when searching for relationships [Giles-Corti et al., 2005]. A study from Seattle found that 49% of participants had greater exposure to supermarkets outside their home neighbourhood

[Hurvitz & Moudon, 2012]. Similar results were found in Minnesota, where the participants had more than twice the exposure at work than at home [Jeffery et al., 2006].

The environment has been recognised to be important for understanding health-related behaviour [Ball et al., 2006; Lytle, 2009; Thornton et al., 2012]. Within nutritional research, there is an increased focus on measuring the impact of the food environment on health outcomes such as BMI [Leung et al., 2011; Rossen et al., 2013; Oreskovic et al., 2009; Wang et al., 2007], body weight [Burgoine et al., 2014; Pearce et al., 2009], obesity [Jeffery et al., 2006; Liu et al., 2007] and diet [Thornton et al., 2012; Pearce et al., 2009; Laxer, 2013]. Environmental exposure is often conceptualised through and measured within neighbourhoods. However, the spatial area that defines a neighbourhood has proven difficult for researchers to define, and the result is a large variety in the definitions of 'neighbourhood' used to study environmental exposure [Lytle, 2009]. The method used to define 'neighbourhood' is essential for researchers with respect to ensuring that it reaches optimal agreement with the actual exposure. However, for researchers to accomplish such agreement, they have to scrutinise behaviour carefully to fully understand the phenomenon. The way a neighbourhood is defined should reflect the context in which it is applied. Therefore, researchers have to make qualified assumptions about where people shop or eat out, the distance people are willing to travel for shopping or eating out, and other individual preferences [Lytle, 2009].

That there are challenges in defining neighbourhoods seems evident, and several studies seems to agree on some of the overall challenges [Burgoine et al., 2014; Burgoine et al., 2013; Odoms-Young et al., 2009; Kerr et al., 2011; Chaix et al., 2012]. Ball and colleagues [Ball et al., 2006] summarise this as follows: 1) people live and function in multiple context and settings; 2) people live and work in multiple geographic areas; and 3) there are different types of environmental influences, including built, natural, social, cultural and policy environments. Consequently, the methods used for defining 'neighbourhood' have to accommodate the individual characteristics of people's behaviour. An augmented focus on the individual is conceptualised by Rainham and colleagues as the change from a place-based to a people-based perspective with individual-based measures [Rainham et al., 2010].

Numerous examples are found in the literature that contradict the people-based approach through application of administrative divisions as the spatial area of a neighbourhood [Odoms-Young et al., 2009; Powell et al., 2011]. Census tracts [Apparicio et al., 2007; Block et al., 2004; Duran et al., 2013; Moore & Diez Roux, 2006], zip codes [Powell et al., 2011] or parishes are used as spatial representations of neighbourhoods for analyses of exposure to the food environment.
In the place-based perspective, all behaviour is located and centralised around the home. The importance of people's closeness and sense of belonging to a certain community and place is challenged by today's society. No matter to which beliefs one adheres, then the human mobility has increased substantially the last century and connectivity have made activities and places more dynamic.

Problematically, each individual is unique and, consequently, must be assumed to have his or her own perception of the neighbourhood concept. Complexity and heterogeneity of human mobility no longer appear to correspond to the use of residential neighbourhoods. Exposure to the food environment occurs in multiple environments, but to measure the impact of people's individual neighbourhoods in multiple environments is challenging.

The use of the term 'neighbourhood' in food environment research relies on spaces defined by fixed boundaries, such as administrative units, or a fixed distance, such as buffers, that define a school or residential neighbourhood [Lucan, 2015]. When referring to individual-measured areas, a more appropriate term instead of neighbourhood is 'activity spaces', as suggested by Zenk and colleagues [Zenk et al., 2011]. This division discrimination in terminology can potentially improve researchers' understanding of the differences between place-based and person-based exposure measures.
Technologies for tracking individuals' behaviours have been available for more than a decade. However, the development of lightweight, low-cost and accurate GPS devices and assisted GPS in smartphones has boosted the use of tracking within behavioural nutrition research. Tracking technologies can provide an individual measurement of space-time information about people's behaviour. The outcome of tracking can potentially consist of millions of data entries, which have to be handled and conceptualised to resemble a neighbourhood. The derived activity spaces are individual and not dependent on a fixed location. Therefore, commuting routes and leisure time activities are also included.
Even though many studies utilise neighbourhood as a concept, only few studies explore how neighbourhoods are defined or which definition is suitable for the study. A variety of neighbourhood definitions is applied in relation to measuring the impact of the food environment.

2.4 SPATIAL BEHAVIOUR TRACKING

There are several platforms that can be used to measure spatial behaviour. However, GPS is the oldest used technology and is widely applied. GPS tracking within behavioural nutrition research] has been limited [Zenk et al., 2011], but other research areas have embraced the technology and have used it for tasks such as measuring travel patterns [Murakami & Wagner, 1999; Stopher et al., 2008], tracking wildlife movement and habitats [Markham & Altmann, 2008], examining exposure to toxins, pesticides or air pollution [Phillips et al., 2001], following elderly individuals with Alzheimer's and other dementias [McShane & Skelt, 2009], and measuring physical activity within health research [Rodriguez et al., 2014].

Mobile positioning, Wi-Fi and Bluetooth signals are more novel technologies in terms of tracking and are less used in research than GPS. Application potentials of Bluetooth and Wi-Fi are similar and has been used for measuring travel time through airport security [Bullock et al., 2010], travel time on freeways [Haghani et al., 2010] and mapping large crowds at mass events [Versichele et al., 2012]. GSM has been sparsely used for tracking, with the exception of Ahas et al. [Ahas et al., 2008], who used it for GSM tracking of tourists. Smartphones are believed to have a large potential for use within the social sciences and for measuring behaviour through the use of apps and their large population penetration. The number of mobile phone subscriptions worldwide exceed the total world population in December 2014 [Kemp, 2014]. Unique numbers of mobile users worldwide surpassed 50% in September 1014 and rose by 5% in 2014 [Kemp, 2015], making profound the basis for the passive tracking of populations with Wi-Fi, Bluetooth or mobile positioning.

Knowing and understanding the strengths and weaknesses of tracking technologies are the keys to selecting one or more technologies fitting the scale and environment of the research. The best fit of a technology to a study is influence by the environment, the extent of the study area, the required accuracy in the positioning, the need for eliciting a response (active tracking) versus merely measuring movement (passive tracking) and the availability and pricing of the hardware and data.

2.5 SPATIAL DATA INFRASTRUCTURE (SDI)

The Danish national spatial data infrastructure is a common foundation for the management of geographical information and digital administration in Denmark [Hansen et al. 2011]. A well-functioning spatial data infrastructure is an important prerequisite for e-governance. The implementation of the INSPIRE Directive has emphasised SDI within key ministries and has resulted in several national services with free and easy access to spatial data on the environment, spatial planning, addresses, cadastral maps, and topography. However, until now, public health information has not been a part of the Danish national SDI. In 2008, the Danish government enacted The Act on Infrastructure for Geographic Information as a response to the European INSPIRE Directive. The act ensures the implementation of the INSPIRE Directive in Denmark and allows the common framework to be applied widely in the national geographical data infrastructure. The Danish spatial data infrastructure follows the basic principles of INSPIRE and establishes common rules, conditions and guidelines for data, services, technologies, metadata, and its organisation [Danish Ministry of the Environment, 2008]. The Danish infrastructure for spatial information consists of data and services through web portals or standardised services, such as Web Map Services (WMS) and Web Feature Services (WFS). A web-based portal publishes data and metadata such that users have one place to search and access the data [Danish Geodata Agency]. For a dataset to be a part of the Danish spatial data infrastructure, it must be digital, nationwide, nationally applicable, and statutory [Danish Ministry of the Environment, 2008].

In 2011, the Agency for Digitisation was established, and the national e-government strategy accelerated the process of modernising Danish society [Danish Agency for Digitisation, 2011]. The digitally based ambitions are described in the Danish e-government strategy of 2011-2015 [Danish Agency for Digitisation, 2011]. The government, municipalities and regions included in the strategy are tasked with increasing the momentum of digitisation in the public sector. Spatial information and the associated infrastructure play an important role in delivering data for public administration. One of the initiatives is common basic data [Danish Agency for Digitisation, 2012] for all public authorities. Since 2013, some spatial and register data have been freely available for the public [Hansen et al., 2013]. The strategies programme for basic data are directly based on the principles of the INSPIRE Directive to ensure consistency with INSPIRE in the development of the national infrastructure for geographical information and the digital public administration. INSPIRE also contributes to the coherence and exchange of data across public authorities through the use of international standards, including INSPIRE.

2.5.1 HEALTH DATA

Digitisation of public spatial data are on the agenda for many governments, and each country is developing its own national SDIs [EDINA]. Commonly, national SDIs focus on traditional spatial data, such as addresses, property information, spatial planning, remote sensing data and environmental data. However, at best, spatial health data have a very limited implementation in national SDIs: Denmark has only a few health datasets available [Danish Geodata Agency]. There are several international databases that distribute spatial health data, i.e., Worldmapper, those listed in by Mathys and Boulos [Mathys and Boulos, 2011], the World Health Organisation [WHO (2)], HIVMapper [HIVMapper] and HIV Spatial Data Repository [HIV Spatial Data Repository]. Only a few countries have or are developing spatial data infrastructures that include health data. In the United Kingdom, an environment and health atlas [SAHSU] has been published. Additionally, the SDI Go-Geo [EDINA] delivers metadata, spatial data and interactive mapping to medical researchers, public health officials and the general public [Mathys and Boulos, 2011]. New Mexico has developed a web portal to integrate environmental information and health information [EPHT]. In Victoria, Australia, spatial data access and management is a priority. An SDI to increase and strengthen effective collaboration within health projects and to add benefits through the increased use of under-utilised data are being developed [Thompson et al, 2009].

In 2012, the important Danish national health data at SSI was consolidated to ensure equal and transparent conditions for the use of data and to improve the quality of data and its sharing amongst health professionals and researchers. The role of SSI is to gather, analyse and disseminate data. Following the national e-government strategy of 2011-2015 [Danish Agency for Digitisation, 2011], a national strategy for the Digitisation of the Danish Healthcare Sector 2013-2017 was developed. One of the five main initiatives is the ‘better use of data’ [Danish Government et al, 2013]. This strategy should create a basis for the affordable maintenance of health data, the collaboration of information technology across the health sector, and the improved quality of health data through ensuring a reliable link between local health departments and

national registers [Danish Government et al, 2013]. The strategy facilitates the management and sharing of health data through a common infrastructure and standards for data, interfaces and services. However, the national strategy for the Digitisation of the Danish Healthcare Sector 2013-2017 does not mention INSPIRE nor the promotion of spatial applications of health data.
In Denmark, the Statens Serum Institute (SSI) maintains the health data registers. The National Institute of Public Health (SIF) maintains a few clinical registers and conducts a national health survey every three years. Table 1 lists registers that contain health data.

Health Register	**Responsible Authority**
The Danish Pathology Register	SSI
- 51 clinical registers	SSI & SIF
Cause of death register	SSI
National Patient Register	SSI
- Birth and fertility register	SSI
- Psychiatric Register	SSI
The Children's Database	SSI
The Conscription Register	SSI
The National Health Insurance Register	SSI
Rehabilitation Register	SSI
Health Service Provider Register	SSI
Central Business Register (CVR)	SSI
Register of Medicinal Product Statistics	SSI
Register on Drug Abusers in Treatment	SSI
The National Health Profile	SIF

Table 1. List of health registers in Denmark and the responsible institution.

All of the national registers maintained at SSI contain a personal identification number (CPR number) such that all registers can be combined [Pedersen, 2011]. The CPR number also provides an opportunity to geocode the data at the address level to support further spatial analysis [e.g., Storgaard et al., 2013]. Data from the National Health Profile are based on addresses and are aggregated at the municipality level; the data are freely available from a web portal.

2.5.2 SPATIAL DATA

The main sources of spatial data in Denmark are available from the Danish Geodata Agency's spatial data distributer (Kortforsyningen), the Danish Natural Environment

portal and the Danish Spatial Planning System web portal. The Danish Geodata Agency's spatial data distributor has spatial data that are freely available, such as ortho photos, topographic maps, place names and information registers (SNSOR in Danish), administrative boundaries, elevation models, cadastral data, historical maps and geo-keys [Danish Geodata Agency (2)]. At the Danish Natural Environment portal, spatial data related to the environment are available, and at the Danish Spatial Planning System web portal, spatially referenced plans at all governmental levels are published. All data are standardised, so they are available in the spatial reference system ETRS89 / UTM zone 32N (SRID = 25832) and in formats such as Shape (ESRI), Tab (MapInfo), MDB (GeoMedia) and GDB (ESRI-geodatabase format).
Geodata-info is the national counterpart to the INSPIRE portals that aims to search and discover spatial datasets and associated metadata in Denmark [Danish Geodata Agency].

2.5.3 BASIC DATA

The basic data programme in Denmark was established to ensure an effective use of governmental basic data through improved quality and a common platform for the distribution of data. The common platform for data distribution is currently under development, but the first type of data should be available in the fall of this year (2015) followed continuously by more data until Spring 2017, when the current plan will be fully integrated [Danish Geodata Agency (3)].
The basic data programme reflects data pertaining to the public sector, with information about people (CPR), companies (CVR), buildings (BBR), property (ESR), cadastres, addresses (DAWA— previously, AWS) and spatial data.

Addresses are easily recognisable and are used to locate residences in the Building and House Register (BBR) and the Central Business Register (CVR). Addresses are also registered in the Civil Person Register (CPR) [Pedersen, 2011], which contains information on individuals in Denmark. The CPR is key for linking health data [Bjerregaard and Larsen, 2011], social data, labour market data [Petersson et al, 2011] and education data [Jensen and Rasmussen, 2011]. The CPR contains addresses that can be linked to a spatially referenced address dataset and can thereby establish a relation between health data and spatial data, for example, on pollution in the Danish Natural Environment portal. All companies, institutions, and public service providers are registered in the CVR and are uniquely identified through the CVR number. The CVR also has information on addresses and industrial classifications. The relations between data directories in the basic data programme are illustrated in Figure 4.

Statistics Denmark is the national agency for statistical data, and the data are obtained from national public registers. The statistical data are outputted from individual data aggregated over administrative units, such as parishes, municipalities, regions or the entire country. Additionally, data may be aggregated over miscellaneous spatial units or the Danish National Grid. Aggregation is limited by enforced restrictions to ensure

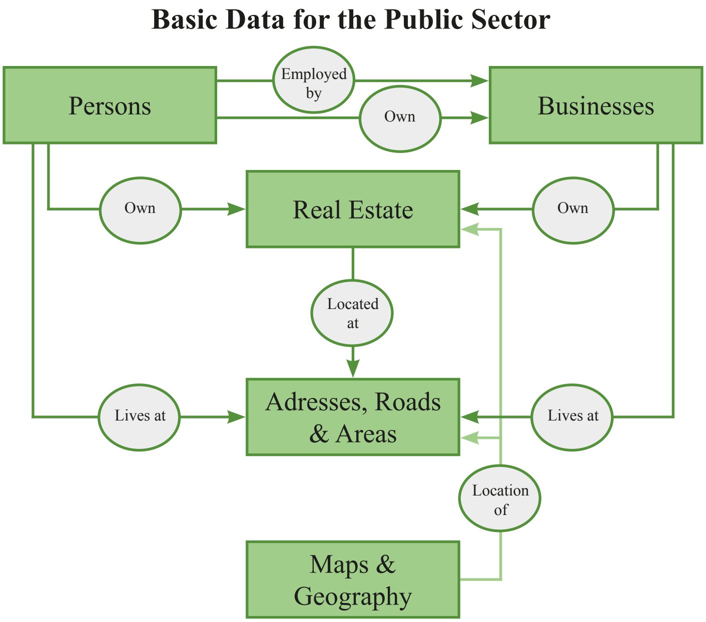


Figure 4. Relations in Basic Data for the public sector.

the confidentiality and privacy of individuals. The restrictions enforced by Statistics Denmark require that persons are not identifiable from the data. Therefore, the minimum number of persons or households in an aggregation is variable, depending on the type of data. However, threshold values of a minimum of 50 properties or 100 persons within each spatial unit are often required.

2.5.4 DATA INTEROPERABILITY

Spatially enabling public registers require a common key attribute [Hansen, 2001]. Address are an important and unambiguous database key in many private and public registers in Denmark, and all Danish addresses have a spatial reference. Figure 5 presents the associations between the spatial data repositories, the spatial reference data, data keys, national basic data registers, national health registers and Danish statistics data. Through data keys and spatial references, all data sources can be related.

Most public health data do not have any spatial references but need to be linked to features with a spatial reference, for example, administrative units or addresses. Health data can be geo-referenced and aggregated through a polygon feature dataset (e.g. administrative boundaries) from the Danish spatial geodata portal, as presented in Figure 6.

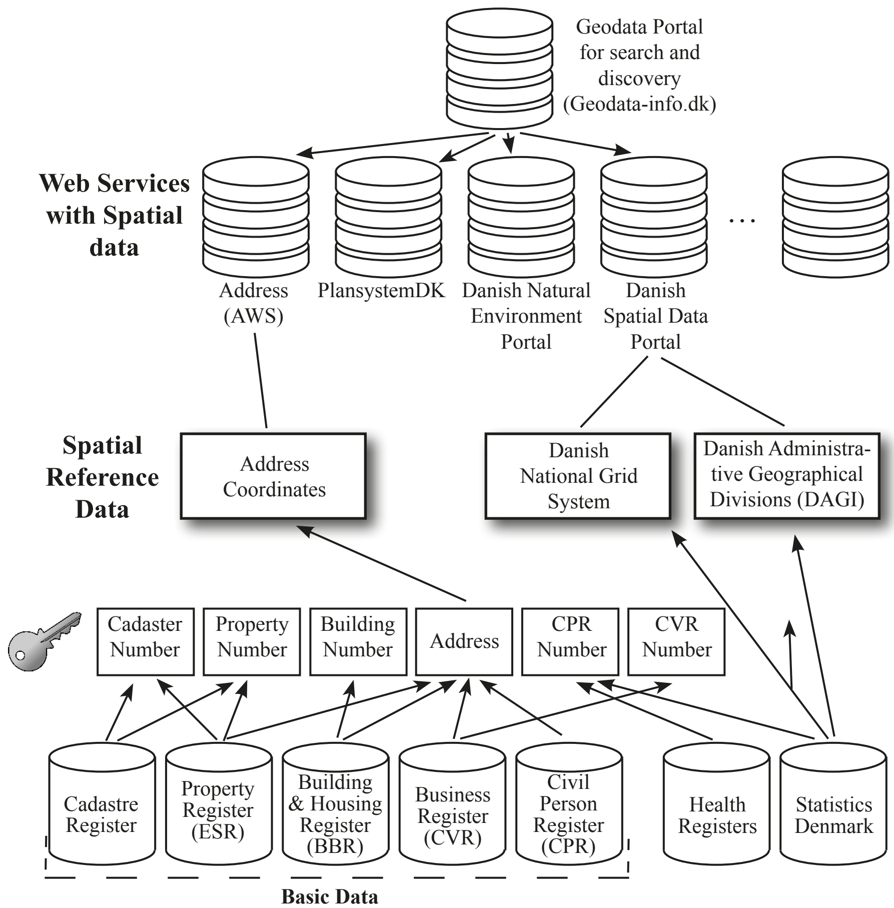


Figure 5. The associations among national registers through their key attributes and spatial references from spatial web services.

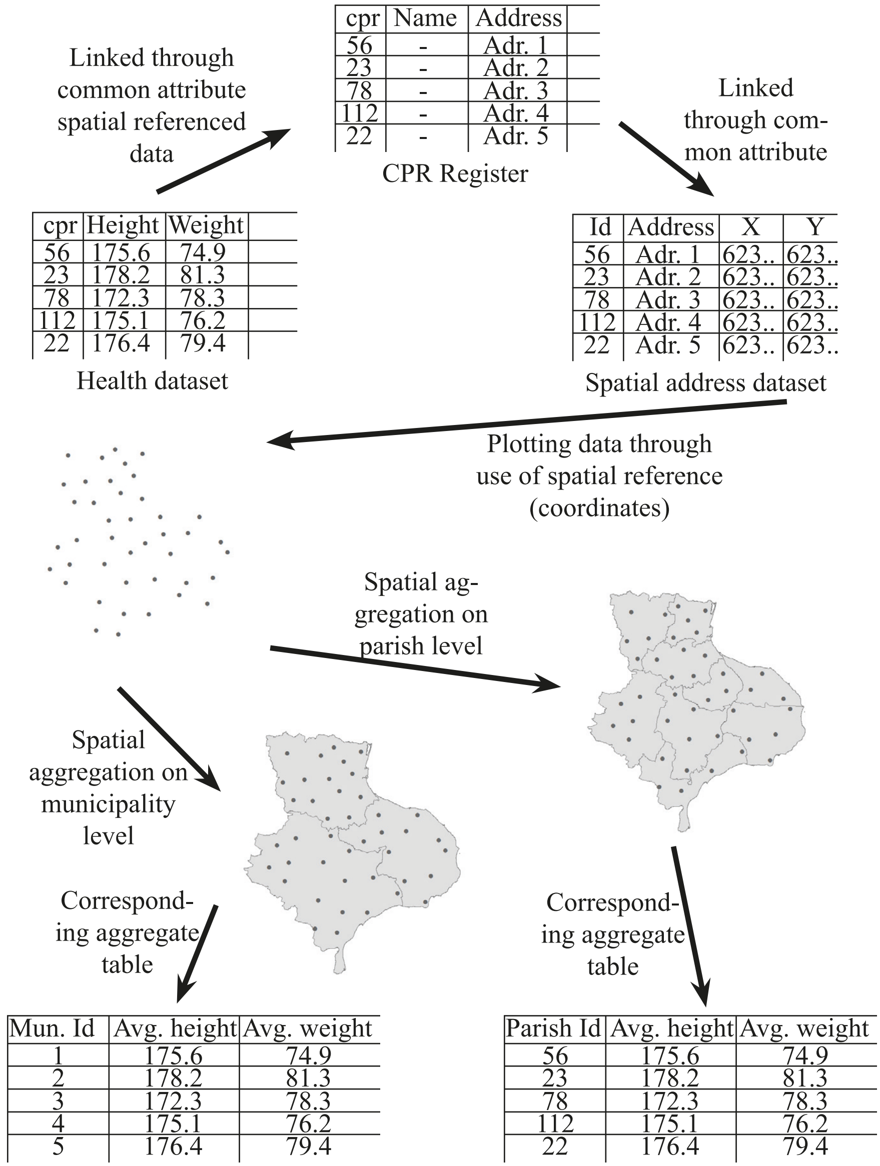


Figure 6. The process of geocoding health data via the CPR register and then spatially aggregating the health data into administrative units.

2.6 PRIVACY ISSUES

2.6.1 BEHAVIOUR TRACKING ISSUES

The use of wearable position technologies to follow the behaviour of individuals is an intrusion into people's privacy and violates many people's boundaries. Surveillance is frowned upon in many countries and raises a whole host of ethical issues for many people and governments. Privacy is a public concern, which causes debate about personal freedom and scientific ethics. Privacy, surveillance and data security are key aspects in both passive and active tracking. In active tracking, the person's identity is known to the researchers, who have the key responsibility, when storing, analysing and publishing the results of tracking, to ensure the confidentiality of the participating persons. European and national legal regulations for data collection must be followed to ensure the persons and their locations are not identifiable. Privacy and ethical concerns influence the type of people willing to participate in studies involving tracking. Some of the concerns revolve around the fear of being subject to surveillance and being followed and listened to everywhere one goes. The growing ICT generation may be more open to the technologies because being accustomed to positioning technologies may reduce these concerns. An explanation of the features of these technologies and a demonstration of the results of tracking could decrease the concerns of privacy violations and data security. Private tracking has to treat the people in a way that ensures their complete confidentiality, their data included. Thermal cameras ensure confidentiality compared with RGB-cameras, whereas the application of other methods are needed for BT, Wi-Fi and mobile positioning. There are several methods [Rainham et al., 2010] for assigning 'dummy' variables to ID a device instead of using the MAC address.

2.6.2 HEALTH DATA ISSUES

In different countries, there are different traditions in the registration of health data, which entail a variety of barriers in geocoding patient records. Northern European countries have a tradition for registering a large amount of information for every citizen, which potentially allows for an easier process when geocoding patient registers if location is included in this plethora of information. In other parts of the world, financing issues tends to restrict the availability of accurate databases, and the geocoding process will require much time and resources. Some countries have legally limited the data that each individual is permitted to register at a government level, which means that geographical references may not be recorded in patient records in some jurisdictions. Regardless of the problem each country has when geocoding its patient records, consideration for patient confidentiality is an issue when analysing the data and making results publicly available. This requires the data providers to ensure that the data are appropriately aggregated, either through legislation or pre-aggregation of the data.

According to Danish legislation, health information is private, which imposes great limitations on the use of health data. All registers that contain the CPR number are subject to severe privacy and confidentiality issues, legally and technically.
Danish legislation limits access to individual health information to employees within the sector with relevant needs and to researchers who have been granted permission through legal channels.

Health data include some of the most personal and private information on people. However, patient-identifiable data are critical to medical research insofar as updating, linking and validating data are impossible without identifiable data, and the implementation of potential confounders in the analysis is difficult [Haynes et al, 2007]. Spatial aggregation is a means of preserving confidentiality while maintaining an acceptable level of data usefulness [Boulos et al, 2009]. The privacy concerns additionally challenge health researchers due to the expensive and time-consuming methods required to secure data anonymity or confidentiality. Ultimately, there is always a trade-off between privacy concerns and the types and accuracy of possible spatial analyses of health [Boulos et al, 2009].

3. METHODS AND MATERIALS

This chapter includes a description of the study's geographical area and population sample, the data used in the study, the methods and the statistical approaches. All papers included in the thesis are based on Danish data, and the analyses are set in a Danish context. However, the results are discussed in a broader context.

3.1 STUDY AREA

The study area for both Papers II and VI are concentrated around Aalborg in Northern Jutland (Denmark). The study areas vary in size for Study II and VI. Study II consists of 49 parishes, whereas Study VI consists of the same 49 and an additional 16 parishes. In Study II, the study area is approximately 974 km^2, of which the city of Aalborg with the high-density housing constitutes approximately 8%. Approximately 15% of the population in the study area has an ethnicity other than Danish, and the levels of education and income are diverse across both the low- and high-density housing.

In Study VI, the population in the study area is approximately 230,000, and of that number, approximately 120,000 live in Aalborg. The study area is approximately 1552 km^2, of which Aalborg, with its high-density housing (mean ≈ 1700 people/km^2) only comprises 68.3 km^2 (≈4.4%). The remaining areas consist of small villages with populations of up to 7000 and low-density housing (mean ≈ 85 people/km^2). The study area's spatial expanse, relative location in Denmark and divide between urban and rural areas are presented in Figure 7.

Northern Jutland consists of 11 municipalities, five of which are defined as peripheral regions. Peripheral regions are characterised by, among other factors, a lower average income than the national average, a lower amount of commuting traffic and low or negative population growth. However, Aalborg attracts many young people and is the economic centre of the region. In Northern Jutland, approximately 50% of all people aged 16 to 25 lives in Aalborg, whereas these people are only approximately 17% of the entire population.

3.2 POPULATION SAMPLE

Selection of the population sample is an important part of research because the sample data are often used to draw conclusions about the entire population. It is important that samples provide a representative sample of the population. There are two main types of sample techniques: **probability sampling**, where the sample is randomly selected to ensure that each member of the population has equal chance of being selected, and **non-probability sampling,** which is used when the population is unknown (although this method tends to be more prone to bias). Probability sampling is optimal, but it is not always possible, and opportunity sampling can potentially be the solution.

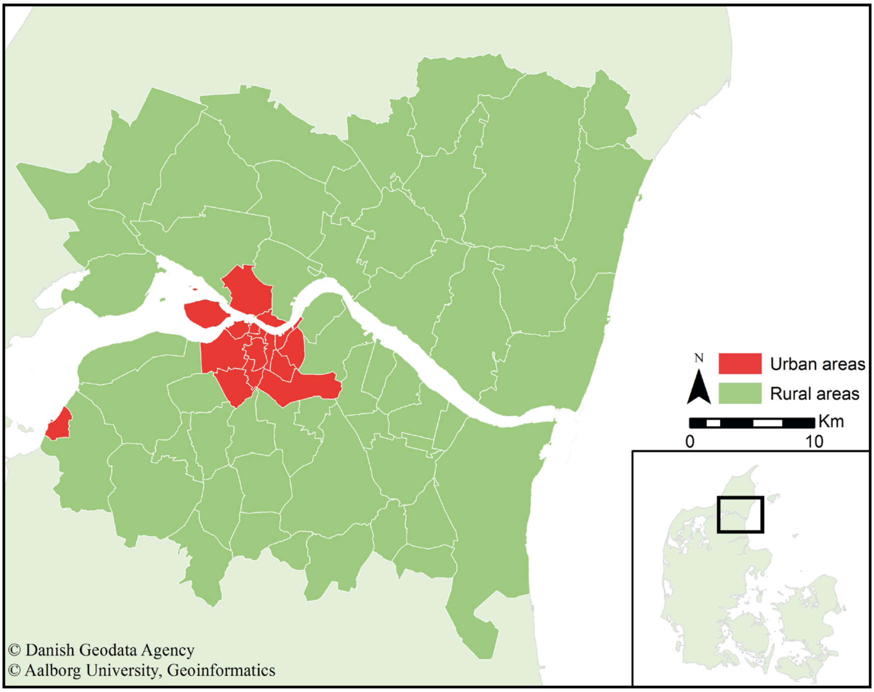


Figure 7. Presentation of the study area, the relative location in Denmark and the division between urban and rural areas.

In Paper VI, the population was limited to students enrolled at an educational institute in Aalborg. The population included approximately 7300 people, of which 223 people were willing to participate.

Gathering a sample population can prove difficult because some respondents might be reluctant to participate. There are many reasons why respondents are reluctant to participate in a survey. As Adler and Adler suggest, reluctant respondents can be found throughout society [Adler & Adler 2002]. However, they tend to fall into four groups:

- **Secretive respondents**—those with secrets who fear to be found out;
- **Sensitive respondents**—those who are sensitive about personal matters such as finances, health, disease or sexual conduct;
- **The advantaged**—those in positions of wealth, status or power; and
- **The disadvantaged**—respondents who might be at risk, engaged in criminal activities or distrustful of the survey's intentions.

In Paper VI, the respondents were asked to carry a GPS for a week. This fostered many questions from possible respondents, who were insecure about the GPS and what data would be visible from it. Information is important for many respondents to feel secure about participating in a survey. A thorough description of the survey that is easy to understand is a good approach to getting more respondents to participate in a survey. All subjects provided their informed consent for inclusion in the survey before they participated in the study and could opt out at any time by turning off the GPS device. The 223 respondents in Paper VI were distributed among six schools. The sample had a higher proportion of female (57%) than male (43%) participants. The participants' ages ranged from 16 to 23 years old, with an average age of 17.7 years.

The Global Positioning System tracked each person for one week of their typical school schedule. The GPS devices used in this study are the Flextrack Lommy Phoenix (Esbjerg, Denmark) and are approximately the same size as a mobile phone. The participants were asked to carry the device at all possible times during the week. The tracking resulted in 8.22 million records for the 223 participants. The number of loggings registered for each person varied from 579 to 128,679, with an average of 36,523. A threshold of 30 h (equal to waking hours for two days) of tracking was set as a minimum for the participants to be included in the study.

The final sample consists of 187 people (36 were excluded). The final sample population included 110 women (58.8%) and 77 men (41.2%) from 16 to 23 years of age (the mean age was 17.3 years old). The final sample included 93 people who live in a rural area and 94 people who live in an urban area.

3.3 DATA SOURCES

In Papers II and IV, data on food sources were used. Data were gathered from two nationwide directories (the CVR and Smiley). In both directories, branch codes were used to define food premises. The branch codes are based on the European NACE classification [Eurostat-European Commission, 2008]. The Smiley and CVR data were retrieved in June 2013.

3.3.1 CENTRAL BUSINESS REGISTER (CVR)

The CVR is a government register that contains information about businesses in Denmark. Information about the legal unit in the companies is uniquely identified through the CVR number, and under each legal unit, production units are identified through unique P-numbers. The P-number is used for a complete list of food retailers because each geographical individual retailer in a chain has its own P-number. The CVR is updated once each day, 5 days a week, all year. The database is administered and managed by the Danish Business Authority. The information is provided by the business owners, and it is their responsibility by law to keep the information up-to-date and correct. That the information about branch and address are kept up to date through third-party reporting implies that information consistency, accuracy and completeness could be doubtful. The CVR contains no information about the availability of foods such as fresh meat or vegetables in food selling premises or furnishing, opening hours or paying options in food-serving premises. Consequently, the NACE classification and business names are the only sources for identifying the different food premises. The 15 branches listed in Table 2 were identified in the CVR as food-selling or -serving premises by definition. [Eurostat-European Commission, 2008]

Classification	**NACE Code Used in CVR**	**NACE Code Used in Smiley**
Grocery shops and kiosks	47.11.10	47.11.00.A
Supermarket	47.11.20	
Discount supermarket	47.11.30	
Other non-specialized shops	47.19.00	
Greengrocer	47.21.00	47.21.00.A 47.21.00.B
Butcher shops and delis	47.22.00	47.22.00.A 47.22.00.B
Fish shops	47.23.00	47.23.00
Retail with bread, confectionery and sugar products	47.24.00	47.24.00.A 47.24.00.B
Retail with beverages	47.25.00	47.25.00
Other food in specialized shops	47.29.00	47.29.00.C 47.29.00.D 47.29.00.E
Gas stations	47.30.00	-
Full service restaurants	56.10.10	56.10.00.A
Pizzeria, ice cream, etc.	56.10.20	56.10.00.B
Bars, cafés, etc.	56.30.00	56.30.00

Table 2. List of NACE codes applied to limit the search to food retailers in Smiley and the Central Business Register (CVR).

3.3.2 SMILEY REGISTER

The Smiley register was introduced in 2001 and belongs under the Ministry Environment and Food, who administers food safety and hygiene regulations in Denmark. The register was created to register the food safety inspections of businesses and present the food safety level of each business to the public. Inspections are performed to ensure that shops and restaurants comply with the regulations. The inspection rates of the businesses are based on the health risk the branches constitute, ranging from two a year to one every second year. Businesses with non-perishable goods are inspected as needed. Consequently, updates of the register are similar to the inspection rate, which suggest retention of outdated data for up to two years. The register is updated every three months with the results of the latest inspections. The time between inspections and the three-month delayed update decreases the validity of the data, as it is less accurate and complete, while retaining outdated information. The relevant NACE classifications identified are listed Table 2 along with the indication of aggregated and disaggregated groups in Smiley compared with the use of the NACE codes in the CVR. The NACE classification and the business names are the only indicators for type of food premise as there is no information about merchandise, the menu, opening hours, table serving or paying options. [Danish Business Agency, 2013]

3.4 METHODS

The methods used in the thesis are data cleaning/preparation, geocoding, spatial analysis and statistical analysis. In Figure 8, a flow chart illustrates the progression of the methods.

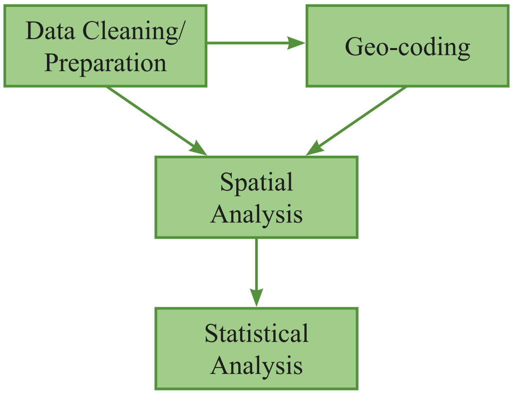


Figure 8. Progression of methods in thesis illustrated in a flow chart.

3.4.1 DATA PREPARATION

In all projects applying spatial data, it is important to clean and structure the data. Cleaning the data makes data more reliable and determines the validity of the data analysis and the results. Restructuring of data are often necessary such that they are usable for analysis in GIS software. Cleaning and structuring are essential in preparation of data for analysis and a time-consuming task.
In Paper I, data from two secondary food source directories were used. Data were in both directories delivered in excel format. A simple selection on municipality codes were created as an initially criterion to limit the data amount. Data quality varies and

data on food sources used in the study needs preparation/cleaning before used in spatial analysis. The quality of the food source data directories are doubtful for spatial use because of the invalidated address registration and questionable food source categorisation. Many of the addresses in the CVR and Smiley have misspelled road names and is therefore often misleading. For linking the data, the road numbers should always be used. However, the house numbers and especially the lettering accompanying the house numbers were often missing or characters did not correspond to the official address in Denmark. The data were cleaned by correcting the house numbers to the closest matching official addresses by removing or adding letters and adjusting house numbers to the closest even or odd number.

The addresses in the CVR were geocoded based on address reference data in the Universal Transverse Mercator (UTM) projection obtained from the Danish Geodata Agency.

The Smiley register contains WGS84 (World Geodetic System 84) coordinates for approximately 95 % of the population, which were transformed to UTM and used as their location data. The remaining records were geocoded through use of the address and reference data from the Danish Geodata Agency. The distribution of the Smiley and CVR directories is visualised in Figure 9.

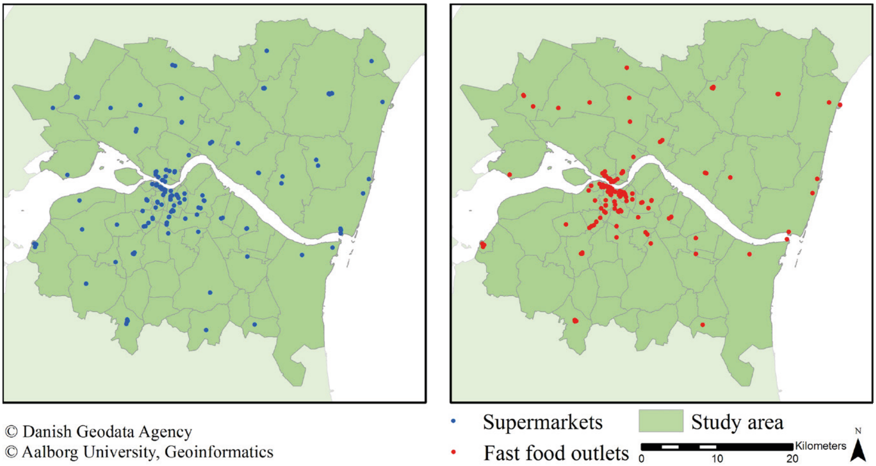


Figure 9. The spatial distribution of (top) supermarkets; and (bottom) fast food outlets within the study area.

For geocoding, an address 'key' was used consisting of the municipality code, street code and house number, including a letter. The municipality code is unique, the street codes are only unique within each municipality, and the house number and letter are only unique for each road. Combined correctly, the three codes make up a unique key for addresses in Denmark as visualised in Figure 10.

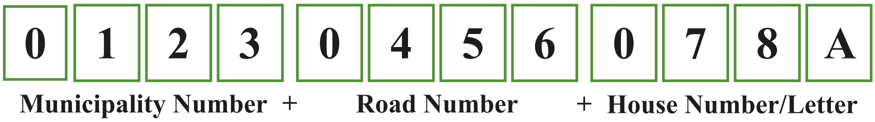


Figure 10. Combination of municipality number, road number and house number that constitutes the unique address key.

In Paper IV, data on spatial behaviour were measured by GPS, which tends to report some erroneous observations. GPS tracking is subject to several technical limitations when measuring space-time data [Kerr et al., 2011; Qi & Du, 2013]. Connection to an adequate number of satellites is critical because lack of such a connection can result in inaccurate position data or complete loss of data for a period.

The errors can be categorised as (1) outliers, either in attribute values—such as dummy values—extreme values—such as the number of satellites, horizontal delusion of precision (HDOP) and time to fix (TTF)—or extreme positions (e.g., on the equator); or (2) scatter, in the form of unnatural linear point patterns [Qi & Du, 2013; Barnett & Lewis, 1994; Hawkins, 1980]. However, outliers are not always erroneous observations. Outliers can be observations alleged to be erroneous and created through a faulty mechanism or can be observations, which clearly are valid, but lead to unexpected and interesting discoveries and hypothesis generation. The detection of outliers has a wide range of application in geographic information systems (GIS) and spatial databases and is essential for cleaning spatial data before its use in spatial analysis [Shekhar & Chawla, 2003; Shekhar et al., 2003]. Traditionally, the detection of outliers in spatial observations has been based on the non-spatial attributes—i.e., through the detection of observations whose non-spatial attribute values deviate significantly from the neighbouring observations' values [Lu et al., 2004; Chen et al., 2008].

Alternatively, the detection of outliers can be based on the spatial patterns of observations, which is especially relevant when using low-cost GPS units.
The data set was cleaned based on non-spatial attributes to remove faulty or inaccurate observations. The cleaning removed observations that did not fulfil the following requirements:

- Duplication of the GPS_timestamp attribute for each individual GPS unit;
- Less than four satellites used for calculating observation coordinates;
- TimeToFix attributes with a value greater than set timeout default value of 60 seconds;
- Horizontal Dilution of Precision (HDoP) values equal to zero or greater than eight; and
- Speeds greater than 150 km/h or greater than 120 km/h outside a 120-meter buffer around motorways.

The unnatural linear point patterns are detected by little or no change in the direction between three or more subsequent loggings, and the location of these loggings are outside a 50 metre buffer on the road network. Detection of outliers and scatter found 341,741 loggings that were perceived as erroneous data.

The GPS devices were set to register the location at 7-second intervals, which was the lowest interval possible for the devices used. However, due to external conditions (i.e., visibility to satellites and time to establish a fix), the logging interval varied up to 60 seconds. The calculation of several neighbourhood definitions in Paper IV assumes an even time interval between loggings (e.g., SD ellipses) because they are based on statistical assumptions. Spatial linear interpolation between subsequent loggings was applied to create an even time interval of 1 second between each pair of loggings. However, a 60-second threshold was set because the GPS creates a duplicate of the previous logging if it cannot obtain three consecutive measurements with an HDoP of less than 30 in 60 seconds. The consequence can be large time leaps for which it is difficult to estimate or guess the location. The interpolation results in a data set consisting of 60.18 million loggings, which corresponds to an average of three days and 17.4 hours of active tracking for each participant.

3.4.2 SPATIAL ANALYSIS

Tobler defined the first law of geography as ‘Everything is related to everything else, but near things are more related than distant things’ [Sui, 2004]. This definition of objects’ relations is the basis for defining neighbourhoods and using them as a concept for measuring their influence on people. The assumption is that all things that are located in close proximity to a person’s residence will influence that person’s behaviour.

In paper IV, neighbourhoods were created based on addresses and GPS loggingsas illustrated on Figure 11.
Based on the addresses of the 187 respondents, spatial buffers were calculated. Buffers are used to create a circular area at a specified distance, and they are quick to calculate, easy to understand and easy to compare because the area size is equal for all study subjects. Simple buffers are based on Euclidian distances, whereas buffers that are more complex are based on network analysis. The buffer distance should be appropriate for examining nutrition-related behaviours for the target group involved. This study applies two distances for defining the buffer size. A distance of 800 m was selected because it is approximately equal to a 10 min walk (5 km/h). Second, a distance of 1600 m (≈1 mile) was selected because it is frequently used in other studies [Leung et al., 2011; Burgoine et al., 2014; Laxer & Janssen, 2013; Giles-Corti et al., 2005; Block et al., 2004; Austin et al., 2005; Kestens & Daniel, 2010; Seliske et al., 2009]. A third neighbourhood definition is defined by combining the buffers for home and school and dissolving overlapping areas.

The administrative division ‘parishes’ were used as neighbourhood divisions because this definition corresponds to the spatially smallest administrative area. Each respondent was assigned to the parish in which his or her residential address was located to identify the neighbourhood. Administrative divisions are advantageous insofar as socioeconomic and demographic information often are aggregated based on these area definitions, which makes it simply to add demographic or socioeconomic information to a spatial analysis when using administrative divisions.

Based on the GPS loggings, three types of neighbourhoods were calculated: convex hull, standard deviational ellipses and path area. The convex hull area is created to represent the minimum bounding geometry enclosing all of the GPS loggings for an individual. The convex hull represents the maximum area in which an individual engaged in activities.

The standard deviational (SD) ellipses are created by calculating the standard deviation in the x-coordinates and y-coordinates from the mean centre of the coordinates. The ellipses do not represent the maximum area in which the individual could engage in activities but rather the area in which the individual is likely to be regularly involved in activities. This study applies one and two SD ellipses, which implies that approximately 68% and 95% or more of the GPS loggings are positioned within one or two standard deviations, respectively. The position of each GPS logging is a weight in calculating the ellipses extent. The GPS loggings therefore must represent an individual's whereabouts with even time intervals, which is performed through interpolation on the space-time data.

The path area represents the participants' travel patterns. For each GPS logging, the nearest road or path segment was determined through a near analysis. On the road and path segments, a 50-metre buffer was applied. The buffer is needed to capture the exposure to food outlets, for which spatial location often has an offset of 5–30 metre from roads. The path area is a narrow definition that encloses all of the GPS loggings, but compared with the convex hull, defines a smaller area of influence.

3.4.3 STATISTICAL ANALYSIS

In Paper II, a statistical analysis is used to examine the validity of two secondary food-source directories. This analysis is accomplished through the calculation of sensitivity, positive predictive value (PPV), negative predictive value (NPV) and the standard deviation of the spatial location of food sources. In Paper IV, the different neighbourhood measures of exposure to the food environment, spatial area and ability to capture behaviour are analysed using a one-way ANOVA (F-test), Welch's two-sample t-test and Tukey's HSD (honest significant difference) test. This is conducted to examine whether there are significant differences between the ability of neighbourhoods to measure exposure to food sources.

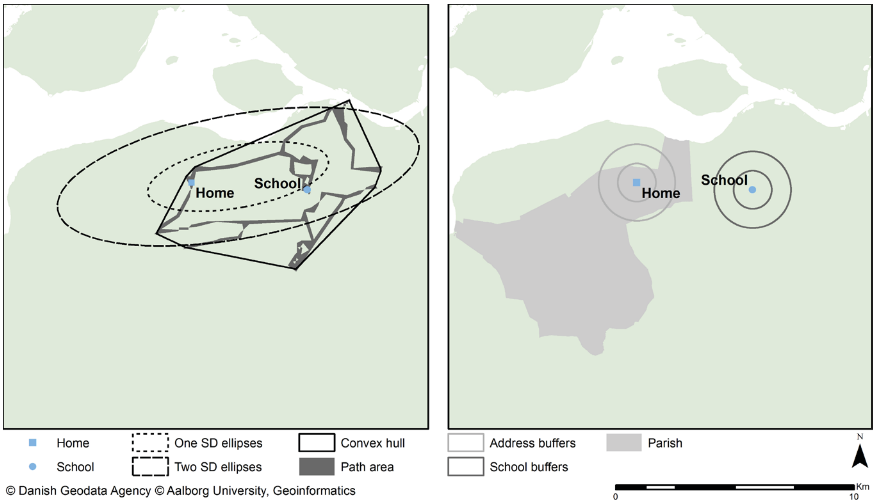


Figure 11. Visual representation of the neighbourhoods' spatial extents and definitions.

3.4.3.1 SENSITIVITY, PPV AND NPV

Sensitivity and PPV are calculated to establish the agreement level between the two food directories and the field observations. The results from the field observations are regarded as the 'gold standard'. The calculation was performed using a 2x2 table as described in Table 3.

		Field observation	
		Present	**Absent**
Food directories	**Present**	True positive (TP)	False positive (FP)
	Absent	False negative (FN)	True negative (TN)

Table 3. Illustration of the relations between true and false field observations and food directories.

Sensitivity corresponds to the proportion of food retailers observed through field observation that were listed in the food directories. Sensitivity is a measure of the completeness of the food directories using Eq. 1.

$$Sensitivity = \frac{TP}{TP + FN} \tag{1}$$

PPV corresponds to the proportion of food retailers listed in the directories observed through field observations using Eq. 2.

$$PPV = \frac{TP}{TP + FP} \quad (2)$$

Sensitivity and PPV were also calculated for the NACE classification, including both non-exact and exact classification matches between the NACE classification and the field observations, thereby providing a measure for the thematic accuracy of the government directories.

The pre-classification of the food retailers is evaluated through sensitivity, PPV and negative predictive value (NPV). NPV represents the proportion of observations pre-classified as not targeted as selling food and observed in the field observation as not selling food. NPV is calculated using Eq. 3.

$$NPV = \frac{TN}{TN + FN} \quad (3)$$

Categorisations of sensitivity, PPV and NPV follow the recommendations of Paquet and colleagues [Paquet et al., 2008]: < 0.30 (poor), 0.31 – 0.50 (fair), 0.51 – 0.70 (moderate), 0.71 – 0.90 (good) and > 0.91 (excellent).

3.4.3.2 SPATIAL STANDARD DEVIATION

The standard deviation (σ) between the food directories' locations and the field observations' RTK GNSS measurements is calculated as a measure of geographical accuracy. The standard deviation is calculated using Eq. 4. The distance (d) is calculated using the Pythagorean distance between the coordinates derived from the food directories and the coordinates from the field observation.

$$\sigma = \sqrt{\frac{\sum(d^2)}{n}} \quad (4)$$

The standard deviation is an indication of the dispersion from the expected or 'true' value. The observations measured by RTK GNSS have an accuracy of 1 – 2 centimetres in the plane [Geoteam, 2013]; hence, the coordinates measured by the RTK GNSS receiver are considered to represent the 'true value'.

3.4.3.3 ONE-WAY ANOVA, WELCH'S TWO SAMPLE T-TEST AND TUKEY'S HSD TEST

Mean values for the exposure to food outlets in each neighbourhood are used to analyse differences, with the null hypothesis being that any difference between the groups is a result of sampling error and the actual differences between the means effectively being zero. Welch's two-sample *t*-test is applied when comparing two groups, and the one-way ANOVA (*F*-test) is applied when comparing three or more groups. The means are compared to analyse whether there are significant differences in exposure to food sources across definitions of 'neighbourhood'.

One-way ANOVA assumes that the data are sampled from populations that follow a Gaussian distribution. Although this assumption is not very important with large samples, it is important with small sample sizes and particularly with unequal sample sizes. One-way ANOVA assumes that all of the groups have the same standard deviation. This assumption is not very important when all of the groups have the same or almost the same number of individuals. The sample sizes in this study are equal for all one-way ANOVA tests.

The one-way ANOVA compares several groups but does not provide information about one group's having a significantly different mean. The differences between groups might be due to errors in sampling, whereas other differences might have other sources. Therefore, a post hoc comparison test is conducted to examine the differences between pairs of each of the neighbourhood types. This identifies pairs of neighbourhoods that have significantly large differences that are not the result of sampling errors. These differences are calculated using Tukey's HSD (honest significant difference) test. Tukey's HSD test is weak, meaning it is limited in its ability to detect significant results. The test assumes normality for each group of data, independent observations within and among groups and homogeneity of variance. The test is quite robust to violations of normality and, to some extent, violations of homogeneity of variance for large samples. Tukey's HSD test requires previous calculation of one-way ANOVA and is calculated using Eq. 5. M_1 and M_2 are the means of the neighbourhood groups, MS_w is the mean square within groups from the one-way ANOVA, and n is the number per group.

$$HSD = \frac{M_1 - M_2}{\sqrt{MS_W(\frac{1}{n})}} \tag{5}$$

Welch's *t*-test is used to test the hypothesis that two independent or unpaired groups of data have equal means. The test is an adaption of Student's *t*-test but is used when the variances are possibly unequal. The test compares urban and rural samples, which are non-overlapping. The test assumes that the data are independent. Welch's *t*-test is calculated using Eq. 6, where is the group mean, S_i is the group variance and N_i is the group sample size.

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}} \tag{6}$$

The results are presented as the means and standard deviations for the spatial areas pertaining to the neighbourhood definitions, the ability to capture behaviour (GPS) and exposure to supermarkets and fast food outlets. For Welch's *t*-test and Tukey's HSD test, the results are presented with significance levels and lower/upper 95% confidence intervals for the differences. Significance levels are perceived as statistically

significant below the 0.05 level (*) and highly significant below the 0.01 level (**). The confidence intervals should either both be positive or negative to demonstrate that the difference is statistically significant to a degree that the results should be trusted. ANOVA is presented with F-test values and significance levels. High F-test values indicate that more pairs in the tested group have significant different mean values.
All statistical analyses are calculated using R [R Core Team, 2014].

4. RESULTS

4.1 ACCESS TO AND INTEGRATION OF HEALTH AND SPATIAL DATA (PAPER I)

This section describes the results from the analysis of the ability of Danish health data to be integrated into a national or European SDI and to be integrated with spatial data.

Denmark is dedicated to digitising its healthcare system. As a result, many advanced national registers with person-identifiable health data on the entire population are available. The registers are high-quality and accurate with broad coverage. This provides researchers the opportunity to perform high-quality analyses on the entire population or on selected cohorts and patient groups without time-consuming and expensive data collection.

Centralisation of health data at primarily one authority (SSI) is the first step towards improving the organisation of health data in Denmark. SSI plays an important role in the ongoing digitisation process of the Danish healthcare system. SSI is responsible for the organisation and control of health data. Thus, SSI has initiated the development of IT systems to support efficient and standardised reporting of health data throughout the healthcare system. The concept is that health data are registered and added to the registers through real-time updates. This ensures that the health data are available for everybody in the healthcare sector. The flow of health data are presented in Figure 12. Health professionals have dual roles in the system: professionals first register data on patients and then access information for treatment. SSI's role is to develop and maintain the data and IT systems and to promote the use of health data for research. Data from some health registers is applied at Danish Statistics, where the data are available in aggregated form. The aggregation is performed by region, municipality, parish or the national grid system; thus, some health data already have a spatial reference. Previously, the administration of health data were split between multiple units with different procedures. Organising the data at SSI has initiated the development of common standards for health data. This organisation will be an improvement that potentially enhances the use of health data.

Health data often suffers from some of the challenges that have been addressed with spatial data over the last two decades. Duplicate data registrations, e.g., clinical registers and the Danish Pathology Register, contradict the principle of collecting and storing data only once. Statutory data are a foundation for consistent registration and validity. The statutory requirement for providing data for the Children's Database is rather new, which results in discrepant numbers of records for different municipalities. With regard to spatial data, the requirements have also been that they should be stored electronically and be nationwide because this improves data use across different fields.

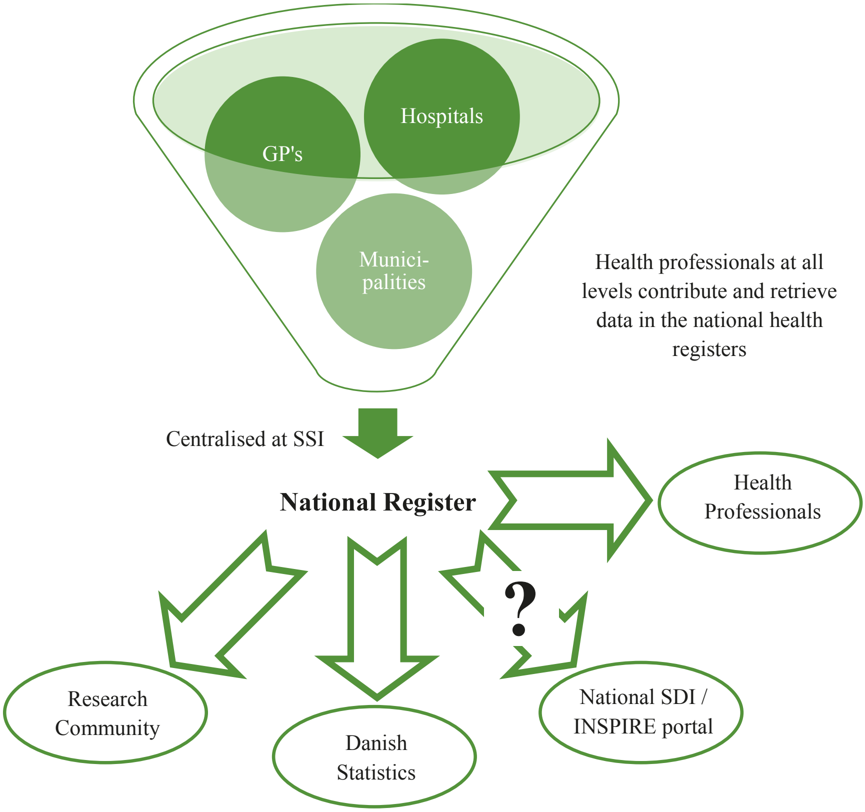


Figure 12. The process of geocoding health data via the CPR register and spatially aggregating the health data into administrative units.

Metadata represent a keystone for ensuring easy access to information. Currently, there is no easy access to metadata on health data in Denmark. Metadata should follow international standards (ISO, INSPIRE) so they can be included in the INSPIRE Geoportal and the Danish counterpart Geodata-info. Implementation of health data in the Danish SDI would be a major step forward in promoting spatial health research because it allows for effective access and use.

The national registers in Denmark all hold information, such as CPR numbers and addresses, for simple linkage of registers and spatial datasets through the use of geographical information systems. Therefore, environmental data and health data can be combined when analysing the causes and outcomes of diseases. However, health data and environmental data have very different requirements and restrictions with regard to spatial and privacy relevance. Health data are subject to heavy legal and ethical restrictions. Modifying data to an aggregated form is essential to bypass these restrictions. However, some degree of detail is inevitably lost in the process. Conversely, legal or sensitivity issues do not restrict environmental data, even though the data refer to a person's activities, e.g., pollution. Health data are based on individuals who are spatially dynamic. Aggregated health data are a momentary representation of the spatial distribution of incidence and prevalence rates. Environmental data are more stationary, as such data often only change over the long term. Administrative units or grid systems make the data easily transparent and combinable with other spatial data.

The rapid development of technology and spatial software over the last decade has important implications for health applications. However, without the proper structuring of health data, the preparation of health data for spatial analysis will continue to be tedious and time-consuming work. However, the harmonisation and implementation of health data in compliance with INSPIRE requires a huge amount of work. To overcome this hurdle, an overall policy is required for implementation, creation of metadata, linkage to spatial data and aggregation of sensitive and confidential data.

4.2 QUALITY OF SECONDARY FOOD SOURCE DATA (PAPER II)

This section describes the results from the analysis of the secondary food source data validation. The validation is made with regard to three parameters—completeness, thematic accuracy and spatial accuracy.

4.2.1 COMPLETENESS

Completeness is a comparison between the retailers listed in Smiley and the CVR and the field observations. From Smiley and the CVR, 285 and 199 retailers, respectively, were selected for field observations. During the field observations, 272 retailers from the Smiley directory and 164 retailers from the CVR directory were present. The PPV calculated for the retailers listed in Smiley that were present in the field observations was excellent (0.95). The PPV for the retailers listed in the CVR was good (0.82).

4.2.2 THEMATIC ACCURACY

The thematic accuracy is a measure of whether the listed food retailer type in the CVR and Smiley matches the type observed in the field. A total of 187 food retailers were observed in the field observations and also listed in Smiley, and 41 (21.93%) were observed that were unlisted in Smiley. One-third of the retailers listed in Smiley were not located in the field observations (n = 98), including those omitted because they were not targeted at selling food (n = 15) or were located in a restricted area (n = 76). The comparison for the Smiley directory is shown in Table 4.

The PPV calculated for the food retailers in Smiley that were present in the field observations was moderate (0.66), and the sensitivity for food retailers in the field observations listed in Smiley was good (0.82). The individually calculated sensitivities

		Supermarket		Specialty Stores		Restaurants		Bars, Cafés, etc.		Total	
		Present	**Absent**	**Present**	**Absent**	**Present**	**Absent**	**Present**	**Absent**	**Present**	**Absent**
Field observation	**Present**	40 (1*)	12	36 (2*)	6	99 (25*)	20	12 (0*)	3	187 (28*)	41
	Absent	3	-	8	-	75	-	12	-	98	-
Sensitivity		0.77		0.86		0.83		0.80		0.82	
PPV		0.93		0.82		0.57		0.50		0.66	

Table 4. Comparison of the food retailers listed in Smiley with those found in the field observations for each classification of food retailers and the total number (incorrectly classified retailers). PPV, positive prediction value.*

		Supermarket		Specialty Store		Restaurant		Fast Food		Bar, Cafés, etc.		Total	
		Present	**Absent**	**Present**	**Absent**	**Present**	**Absent**	**Present**	**Absent**	**Present**	**Absent**	**Present**	**Absent**
Field observation	**Present**	42 (2*)	10	13	25	22 (15*)	10	48 (2*)	33	18 (1*)	7	143 (20*)	85
	Absent	4	-	11	-	11	-	20	-	10	-	56	-
Sensitivity		0.81		0.34		0.69		0.59		0.72		0.63	
PPV		0.91		0.54		0.67		0.71		0.64		0.72	

Table 5. Comparison of the food retailers listed in the CVR with those found in the field observations for each classification of food retailers and the total number (incorrectly classified retailers).*

for each food retailer classification were good and ranged from 0.77–0.86. PPVs were also calculated for the individual classifications but with a larger dispersion from fair to excellent (0.50–0.93).
In Table 5, the comparison between the food retailers listed in the CVR and the food retailers found in the field observations is presented. One hundred and forty-three of the food retailers in the CVR were found in the field observations, and 55 were absent.

The PPV and sensitivity for the comparison of the CVR and field observations were, good (PPV = 0.72) and moderate (sensitivity = 0.63), respectively. The sensitivity for the individual food retailer classifications ranged from fair to good (0.34–0.81). PPV ranged from moderate to excellent (0.54–0.91).
Rural and urban areas were compared based on the number of food retailers listed in the CVR or Smiley and the field validation. However, only small differences were found in both PPV and sensitivity between the rural and urban areas for both the CVR and Smiley. However, there was a small tendency that retailers found during field observations in urban areas that were a bit more likely to be present in Smiley and the CVR.
A comparison of Smiley with the CVR is presented in Table 6. In the field observations, 228 food retailers were identified, but only 117 (51.32%) of these were listed in both the CVR and Smiley. Additionally, 15 observations from the field observations were not found in either the CVR or Smiley. The probability of a food retailer found in the field observations being listed in either the CVR or Smiley is excellent (sensitivity = 0.93).

		Field observation	
		Present	**Absent**
CVR	**Present**	117	26
	Absent	70	15

Table 6. Comparison of the food retailers found in the field observations being listed in Smiley and CVR.

4.2.3 SPATIAL ACCURACY

The field observation coordinates collected with the RTK GNSS receiver and those from Smiley (few geocoded) and the CVR (all geocoded) were compared based on joint Euclidian distance. The means and standard deviations for Smiley and the CVR are 23.74 ± 23.04 m and 18.74 ± 19.83 m, respectively. For Smiley, 97.33% of the records measured in the field were within 100 m of the listed coordinates and 87.70% were within 50 m. For CVR, all records measured in the field were within 100 m and 92.31% were within 50 m. For the 250 × 250 m cells, 12.30% of the records in Smiley and 12.59% of the records in the CVR were found outside the cell in which the listing was registered. None of the records in either Smiley or the CVR were found outside the parish in which the retailer was registered.

The errors between the locations in the registers and the measured locations were analysed for spatial patterns through the measurement of spatial autocorrelation (Moran's I) and high/low clustering (Getis-Ord General G). The results of the analysis were high positive z-scores for both spatial autocorrelation (Smiley 15.74; CVR 15.96) and high/low clustering (Smiley 8.66; CVR 11.18), indicating clustered results. The p-value was, on all occasions, below 0.001, indicating significant results.

The distribution of the clusters was analysed to determine whether the clusters are located in urban or rural areas. The analysis was conducted in the software ArcGIS Desktop 10.2 by ESRI using optimised hot-spot analysis (Getis-Ord Gi* Statistic) from the Spatial Statistics package. In Figure 13, the results are visualised. The clusters with low values (cold spots) are for both Smiley and the CVR located in the central part of Aalborg, whereas the clusters with high values are located in the sub-urban/rural areas for Smiley and in rural areas for the CVR.

4.3 MEASURING BEHAVIOUR—PERSONAL TRACKING (PAPERS III AND IV)

This section describes the results of analysing the potential application domains of different tracking technologies in nutritional research. Second, this section describes the results and experiences gained from using GPS to track personal behaviour in measuring exposure to food retailers.
There are several technologies that can be applied to track personal behaviour in a nutritional context. However, the context of nutritional research can range from the macro-level (nation or society) to the meso-level (community, village or city) to the micro-level (neighbourhood, household, family or person). The different context-levels variations in spatial scale, type of environment and required accuracy (even within levels—neighbourhood vs. person). Hence, a single technology is not preferable in every context. Each of the tracking technologies has pros and cons that influence the choice of technology.

The **environment** for a study is often indoor, outdoor, or a combination.

- GNSS and Assisted-GPS (A-GPS) technologies are only suitable for outdoor tracking.
- Thermal cameras require the environment to have open spaces with limited objects that block the view.
- BT, Wi-Fi and mobile positioning copes with both outdoor and indoor tracking.

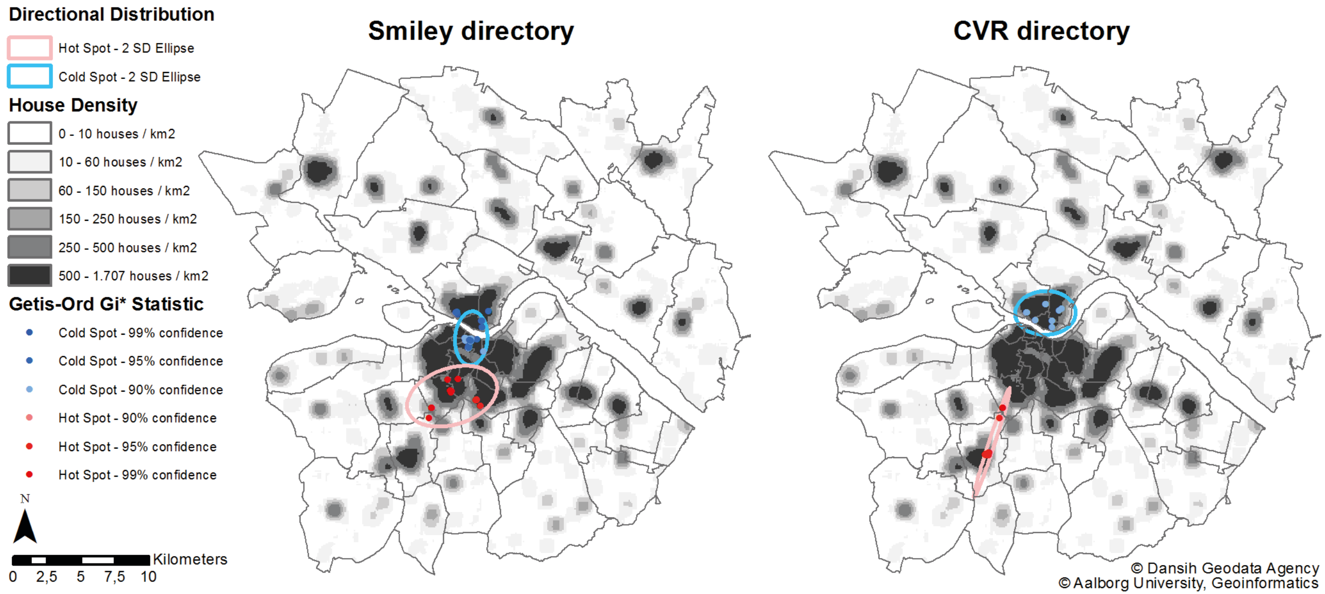


Figure 13. Map of hot/cold spots in the Getis-Ord Gi statistical analysis of the Euclidean distances between 'true' locations and the locations derived from the registers. Two standard deviational ellipses are visualised for the hot and cold spots.*

The **spatial extent of the study area** combined with the environment and the needed accuracy has a large impact on choosing the technology, the number of devices needed and thus the cost of the tracking.

- GNSS, A-GPS and mobile positioning are generally best suited for large areas because the individuals are carrying the technology necessary for tracking.
- BT, Wi-Fi and thermal cameras have stationary sensors. BT, Wi-Fi and thermal cameras are, in theory, plausible to use for large-scale tracking at a very high price for equipment proportional to the amount needed or with a low density of sensors only covering certain zones in the study area.

The **accuracy** in estimated positions of the devices is widely dependent on the environment, hardware and signals.

- GNSS and A-GPS depend heavily on the environment of the study. With high buildings and narrow streets, the accuracy can easily be as poor as 20-30 metres, depending on the equipment, time of day and whether the equipment can use satellites from more than one system. Bare field accuracy can be as good as 1-5 metres depending on the same elements.
- BT and Wi-Fi accuracy vary greatly based on two parameters, the range of the sensors and the possibility of triangulating between several sensors. Accuracy with no overlapping sensors will never be better than the scanning range of the sensors and because of the nature of the signals, the range of the sensors is a bit fuzzy, making a precise estimate of the positioning a bit uncertain.
- The accuracy of mobile positioning is the poorest among the technologies mentioned. There are several methods and network standards upon which to base tracking. Tracking solely on the cellular network yields accuracies ranging from under 100 metres to 20 kilometres—very much dependent on the environment and the density of the cellular network.
- Thermal cameras have proven to have a high position accuracy of 1-2 metres.

The **study population** can actively participate with their full consent or be passively monitored with or without their knowledge. With passive tracking, only the movement patterns of individuals are the output, whereas with active tracking, is it possible to integrate additional information about each person's health status, socioeconomic status, etc. Passive tracking technologies have the potential for following every person closely.

- GNSS and A-GPS both require acceptance by the respondents and can only be used for active tracking.

- Mobile positioning covers over 90% of the population worldwide without requiring their permission. Mobile positioning are best suited for passive tracking but could be used for active tracking.
- Thermal cameras are a passive-tracking technology that have the potential for following every person closely.
- BT and Wi-Fi can be used for passive tracking, although only a fraction of the population is registered. Both technologies can also be used for active tracking through the registration of MAC addresses.

Willingness to participate in tracking varies across the population, and the proportion that would complete the study satisfactorily is in the range of 50-60%.

The **technical knowledge** needed to apply the technologies for tracking differ.

- GNSS includes a huge number of off-the-shelf solutions and delivers output that needs little or no processing to be implemented into GIS.
- Mobile positioning, BT and Wi-Fi are all based on a cell structure, from which processing is needed to change the data to point features with coordinates.
- A-GPS requires an app to handle the tracking, which requires development or buying an existing platform that matches the purpose and hardware.
- Thermal cameras require the most processing as movement detection is still far from commercial use.

The technologies are, in theory, applicable worldwide, but in reality, GNSS, A-GPS and mobile positioning represent the only viable technologies because the rest require power supply or changing of batteries at regular basis. Mobile positioning is limited by the goodwill of the operating companies and the legal statutes in each country.

The **prize** of a complete tracking setup is influenced by the amount of devices, the type of technology and the accuracy. In many cases, A-GPS likely is the cheapest option because a large segment of populations already owns a smartphone and the technology is well suited for combination with BT and Wi-Fi tracking inside selected shops. Combining technologies is preferred to utilise the strengths of each technology in different environmental settings.

4.4 NEIGHBOURHOODS EXPOSURE MEASUREMENT (PAPER IV)

This section describes the results from analysis of a neighbourhood's ability to measure environmental exposure to food sources. Neighbourhoods were compared according to their spatial areas, the amount of behaviour they captured (GPS) and their exposures to supermarkets and fast food outlets. Comparisons are also made between urban and rural neighbourhoods and between place-based and person-based neighbourhoods.

4.4.1 COMPARISON OF NEIGHBOURHOOD AREA SIZES

Descriptive statistics for neighbourhood area sizes are presented in Table 7. The mean areas vary from 2.01 to 51.39 km^2. Significant dispersions occur for the neighbourhood type's parish, convex hull, one-SD ellipses and two-SD ellipses, which are reduced slightly upon dividing the sample into urban and rural areas. No variance exists between buffers around schools or addresses due to the equality of area sizes for all participants.

Neighbourhood		Area		Urban Area		Rural Area	
		Mean	σ	Mean	σ	Mean	σ
Place-based neighbourhoods							
Parish		17.80	20.05	5.70	3.94	30.02	22.28
Address 800 m buffer		2.01	-	2.01	-	2.01	-
Address 1600 m buffer		8.04	-	8.04	-	8.04	-
School 800 m buffer		2.01	-	2.01	-	2.01	-
School 1600 m buffer		8.04	-	8.04	-	8.04	-
Combined 800 m buffer		3.91	0.34	3.80	0.46	4.02	0
Combined 1600 m buffer		15.11	1.98	13.99	2.29	16.06	0.32
Person-based neighbourhoods							
Convex hull		51.13	82.30	21.14	34.93	81.45	103.02
One SD ellipses		17.78	40.55	4.53	6.26	31.17	54.08
Two SD ellipses		51.39	89.99	16.69	21.48	86.46	115.90
Path area		4.76	2.96	3.45	2.40	6.08	2.91
ANOVA	**F-test values**	39.83		24.34		35.48	
	Significance level	<0.001		<0.001		<0.001	
Tukey's HSD test		26 of 55 pairs have significant different means		27 of 55 pairs have significant different means		26 of 55 pairs have significant different means	

*Table 7. Mean area and standard deviation for neighbourhoods for total sample (**n** = 187), urban (**n** = 94) and rural (**n** = 93) areas. Lower portion of table presents results of ANOVA for neighbourhoods.*

Table 7 suggests that the areas of rural neighbourhoods are noticeably larger than those in urban neighbourhoods. Welch's *t*-test compares the area sizes for urban and rural neighbourhoods, and the results are presented in Table 8. All *t*-values from the test are positive, indicating that the rural areas are larger than the urban areas.

Neighbourhood	t	df	Sig. (Two-Sided)	95% Conf. Interval of the Differences	
				Lower	Upper
Place-based neighbourhoods					
Parish	10.369	97.691	** <0.001	19.667	28.977
Combined buffer 800 m	4.671	93.000	** <0.001	0.126	0.313
Combined buffer 1600 m	9.427	96.655	** <0.001	1.773	2.718
Person-based neighbourhoods					
Convex hull	5.349	112.674	** <0.001	37.973	82.646
1 std. deviational ellipses	4.721	94.439	** <0.001	15.439	37.853
2 std. deviational ellipses	5.709	98.246	** <0.001	45.521	94.022
Path area	6.731	177.809	** <0.001	1.856	3.395

Table 8. The results of Welch's ***t****-test comparing urban and rural neighbourhood area sizes. Buffers around the schools and addresses is omitted due to their being no difference.*

Neighbourhood		GPS Logging Count in Neighbourhoods	
		Mean	σ
Place-based neighbourhoods			
Parish		250,302.4 (73.98%)	142,687.9
Address 800 m buffer		248,100.1 (72.93%)	143,025.2
Address 1600 m buffer		256,252.2 (76.30%)	142,999.4
School 800 m buffer		46,299.8 (17.09%)	60,025.4
School 1600 m buffer		70,197.3 (25.50%)	91,348.3
Combined 800 m buffer		281,020.8 (84.71%)	149,402.6
Combined 1600 m buffer		290,858.0 (88.82%)	148,838.4
Person-based neighbourhoods			
Convex hull		321,796.1 (100%)	152,318.5
One SD ellipses		264,880.8 (81.64%)	132,870.1
Two SD ellipses		302,087.2 (94.35%)	145,004.9
Path area		321,796.1 (100%)	152,318.5
ANOVA	**F-test values**	509.8	
	Significance level	<0.001	
Tukey's HSD test		47 of 55 pairs have significant different means	

*Table 9. Mean counts of GPS loggings located within each neighbourhood (**n** = 187). The bottom of table presents the results of an ANOVA for logging counts in neighbourhoods.*

Neighbourhood		Supermarkets		Supermarkets Urban Areas		Supermarkets Rural Areas		Supermarkets pr. km²		Supermarkets pr. km² (Urban)		Supermarkets pr. km² (Rural)	
		Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ
Place-based neighbourhoods													
Parish		3.43	2.31	4.64	2.19	2.20	1.70	0.79	1.13	1.47	1.26	0.09	0.08
Address 800 m buffer		2.18	2.51	3.50	2.83	0.85	1.05	1.09	1.25	1.74	1.40	0.42	0.52
Address 1600 m buffer		6.01	5.87	10.14	5.55	1.83	1.87	0.74	0.72	1.24	0.68	0.22	0.23
School 800 m buffer		4.93	2.87	4.76	3.19	5.11	2.51	2.45	1.43	2.36	1.58	2.54	1.25
School 1600 m buffer		12.65	6.28	12.29	6.50	13.01	6.07	1.55	0.77	1.51	0.79	1.59	0.74
Combined 800 m buffer		6.79	3.75	7.61	4.38	5.96	2.75	1.75	1.00	2.02	1.17	1.48	0.68
Combined 1600 m buffer		16.70	7.61	18.61	8.36	14.75	6.23	1.12	0.55	1.33	0.59	0.90	0.38
Person-based neighbourhoods													
Convex hull		26.44	15.97	22.22	15.69	30.70	15.17	1.32	1.16	1.90	1.27	0.73	0.62
One SD ellipses		6.36	8.32	5.88	7.00	6.85	9.49	1.18	1.84	1.90	2.29	0.44	0.67
Two SD ellipses		20.04	18.75	16.15	14.07	23.97	21.89	0.97	1.05	1.47	1.22	0.44	0.41
Path area		11.44	6.25	10.41	6.30	12.47	6.06	2.82	1.45	3.39	1.53	2.23	1.09
ANOVA	F-test values	137.8		55.35		100.6		60.49		19.51		134.3	
	Sig. level	<0.001		<0.001		<0.001		<0.001		<0.001		<0.001	
Tukey's HSD test		44 of 55 pairs have significant different means		39 of 55 pairs have significant different means		42 of 55 pairs have significant different means		33 of 55 pairs have significant different means		19 of 55 pairs have significant different means		41 of 55 pairs have significant different means	

*Table 10. Mean exposure to supermarkets in each neighbourhood for total (**n** = 187), urban (**n** = 94), rural (**n** = 93) per km² (**n** = 187), per km² urban (**n** = 94) and per km² rural (**n** = 93) samples. The lower portion of the table presents the results of an ANOVA and Tukey's HSD for neighbourhoods.*

4.4.2 COMPARISON OF NEIGHBOURHOODS' ABILITY TO CAPTURE MEASURED GPS ACTIVITY

Each participant's activity was measured using GPS and the neighbourhood types' convex hull and path area by definition captured 100% of the activity. The mean amount of loggings within each neighbourhood type is presented in Table 9. The neighbourhood types, which most poorly captured the GPS-measured activities, were the 800 and 1600 m buffers around schools. The remaining mean values range from 72.93% to 94.35% GPS loggings within the neighbourhoods.

Tukey's HSD test was calculated to compare the individual pairs, and 47 out of 55 pairs were significantly different in mean amounts for loggings located within the neighbourhood boundaries. The results of Tukey's HSD test are available in Appendix V. Tests were conducted by dividing the data into urban and rural areas. Welch's t-test reported significant differences for the school 1600 m buffer (t = −3.220, sig = 0.001) and combined 800 m buffer (t = −4.894, sig < 0.001). In both cases, the urban neighbourhoods captured a significantly larger proportion than the rural sample.

4.4.3 COMPARISON OF EXPOSURE TO SUPERMARKETS IN NEIGHBOURHOODS

The number of supermarkets located within each neighbourhood served as a measure of the exposure to supermarkets, and the results are presented in Table 10. The mean number of supermarkets in each neighbourhood type has a strong positive linear relationship with the size of the neighbourhood areas (cor. coef. = 0.80 and *p* = 0.003).

Neighbourhood	t	df	Sig. (Two-Sided)	95% Conf. Interval of the Differences	
				Lower	Upper
Place-based neighbourhoods					
Parish	-8.489	175.316	**<0.001	-3.000	-1.868
Address buffer 800 m	-8.512	118.519	**<0.001	-3.267	-2.034
Address buffer 1600 m	-13.749	114.016	**<0.001	-9.508	-7.113
School buffer 800 m	0.839	176.145	0.403	-0.476	1.181
School buffer 1600 m	0.787	184.374	0.432	-1.091	2.538
Combined buffer 800 m	-3.086	156.853	**0.002	-2.705	-0.594
Combined buffer 1600 m	-3.567	171.904	**<0.001	-5.970	-1.716
Person-based neighbourhoods					
Convex hull	3.756	184.902	**<0.001	4.023	12.928
1 std. deviational ellipses	0.792	169.149	0.430	-1.443	3.376
2 std. deviational ellipses	2.902	156.711	**0.004	2.497	13.141
Path area	2.277	184.838	*0.024	0.275	3.842

*Table 11. The results of Welch's **t**-test comparing urban and rural neighbourhoods' exposure to supermarkets.*

Tukey's HSD tests were calculated to compare the individual pairs, and the proportions of significant pairs are presented in the last row of Table 10. A distinction is made between the number of significant pairs for the urban and rural samples for both the raw data count and the supermarkets per square kilometre. The complete results of Tukey's HSD test are available in Appendix V. The results of Welch's t-test, presented in Table 11, accentuate the significant differences for supermarket exposure in the urban and rural samples. Non-significant differences exist between one-SD ellipses and both school buffers that are most likely the result of the schools being identical for urban and rural participants. All of the place-based neighbourhoods have negative t-values, which indicate higher supermarket exposure in the urban sample. However, the t-values are positive for the individual-based neighbourhood types, which indicate a higher supermarket exposure in the rural sample.

4.4.4 COMPARISON OF EXPOSURE TO FAST-FOOD OUTLETS IN NEIGHBOURHOODS

Table 12 presents the results of fast-food exposure in neighbourhoods. More fast-food outlets per square kilometre are located near the schools than in other locations. Tukey's HSD tests were calculated to compare the individual pairs, and the proportions of significant pairs are presented in the last row of Table 12. Fewer significantly different pairs of neighbourhoods are found to experience fast-food exposure in urban areas than in the rural sample.

Neighbourhood	t	df	Sig. (Two-Sided)	95% Conf. Interval of the Differences	
				Lower	Upper
Place-based neighbourhoods					
Parish	-9.598	114.176	**<0.001	-5.754	-3.785
Address buffer 800 m	-7.288	96.203	**<0.001	-7.772	-4.445
Address buffer 1600 m	-9.988	98.067	**<0.001	-19.336	-12.926
School buffer 800 m	0.503	184.587	0.615	-2.581	4.350
School buffer 1600 m	0.593	184.756	0.554	-4.326	8.048
Combined buffer 800 m	-1.664	179.407	0.097	-6.862	0.583
Combined buffer 1600 m	-2.163	183.83	*0.032	-13.332	-0.613
Person-based neighbourhoods					
Convex hull	2.489	184.958	*0.014	1.871	16.160
1 std. deviational ellipses	−0.355	180.686	0.723	−5.798	4.029
2 std. deviational ellipses	1.697	167.599	0.092	−1.277	16.913
Path area	0.753	184.021	0.452	−2.258	5.045

*Table 13. The results of Welch's **t**-test for comparing fast-food outlet exposure in urban and rural neighbourhoods.*

Neighbourhood	Fast Food Outlets		Fast Food Outlets Urban Areas		Fast Food Outlets Rural Areas		Fast Food Outlets pr. km2		Fast Food Outlets pr. km2 (Urban)		Fast Food Outlets pr. km2 (Rural)	
	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ
Place-based neighbourhoods												
Parish	4.06	4.16	6.44	4.56	1.67	1.54	1.86	4.76	3.603	6.260	0.091	0.105
Address 800 m buffer	3.81	6.51	6.85	8.06	0.74	1.05	1.90	3.24	3.407	4.007	0.369	0.523
Address 1600 m buffer	9.79	13.71	17.81	15.45	1.68	2.54	1.20	1.68	2.188	1.898	0.206	0.311
School 800 m buffer	13.71	11.99	13.27	12.36	14.15	11.65	6.82	5.96	6.597	6.146	7.037	5.799
School 1600 m buffer	26.99	21.41	26.06	21.94	27.92	20.93	3.32	2.63	3.203	2.696	3.431	2.572
Combined 800 m buffer	16.47	12.97	18.03	14.06	14.89	11.64	4.28	3.47	4.842	3.901	3.703	2.893
Combined 1600 m buffer	33.07	22.27	36.54	23.02	29.57	21.02	2.24	1.59	2.649	1.753	1.820	1.291
Person-based neighbourhoods												
Convex hull	46.92	25.11	42.44	25.08	51.45	24.44	2.87	3.60	4.399	4.416	1.329	1.323
One SD ellipses	11.30	16.97	11.74	15.75	10.86	18.20	2.23	3.84	3.687	4.770	0.751	1.573
Two SD ellipses	34.98	31.61	31.10	26.11	38.91	36.05	2.12	4.10	3.499	5.375	0.729	0.883
Path area	24.29	12.65	23.60	13.18	24.99	12.12	6.45	4.53	8.213	5.198	4.658	2.797
ANOVA **F-test values**	110.7		42.69		78.45		46.06		14.78		80.82	
ANOVA **Sig. level**	<0.001		<0.001		<0.001		<0.001		<0.001		<0.001	
Tukey's HSD test	45 of 55 pairs have significant different means		35 of 55 pairs have significant different means		47 of 55 pairs have significant different means		29 of 55 pairs have significant different means		20 of 55 pairs have significant different means		36 of 55 pairs have significant different means	

*Table 12. Mean exposure to fast-food outlets in each neighbourhood for total (**n** = 187), urban (**n** = 94), rural (**n** = 93) per km² (**n** = 187), per km² urban (**n** = 94) and per km² rural (**n** = 93) samples. The lower portion of the table presents results of ANOVA and Tukey's HSD for neighbourhoods.*

The results of Tukey's HSD test were significantly different between the fast-food outlet exposure in rural neighbourhoods (47/36 of 55) and some of the urban neighbourhoods (35/20 of 55). Welch's t-test compares fast-food exposure in the urban and rural neighbourhoods. The results of the t-tests are presented in Table 13. Significant differences exist between the mean exposures to fast-food outlets for the home-based neighbourhood's parish, addressing both 800-metre and 1600-metre buffers. For all three, the t-values are negative, denoting higher exposure in the urban sample. The results of Welch's t-test for comparing fast food exposure per square kilometre in urban and rural neighbourhoods resulted in significant differences for all neighbourhoods except both school buffers. The t-values are all negative, which indicate a higher fast-food exposure per square kilometre in the urban sample.

5. DISCUSSION

This chapter provides a discussion of the study's methodological strengths and limitations and of the study's results in the Danish context. The relevance and utility of the results are discussed in an international context.

The ecological model of nutrition and physical activity indicates that the interrelation between these two factors is influenced by wide range of multi-level parameters. There exists a need for combining data from several sources to understand the impact of people's behaviour on their dietary habits and disease outcomes. Increased availability of data and improved development and integration of technologies were the basis for the primary research objectives. The European INSPIRE Directive and an open/free data strategy in Denmark have increased the availability of data, but the examination of data validity and integration are still in their early stages. New data from alternative data sources, such as GPS, Wi-Fi and Bluetooth, are emerging as facilitators for gathering data on people's behaviour. The hypothesis was that these technologies could be used to challenge the use of more traditional methods of data collection, such as self-reported data.

5.1 VALID FOOD SOURCE DATA—EXPENSIVE AND TIME-CONSUMING?

Food source data provide fundamental information for much spatially oriented nutrition research. However, many studies confirm that valid and accurate food source data are often not readily available. Often, time-consuming and expensive field validation efforts are carried out to ensure the validity of secondary food source data. In Paper II, data from the Smiley register and the CVR register are used as secondary sources on food retailer data. The data in the Smiley register and the CVR register are, at best, doubtful. In Paper II, the identification of food retailers in public spaces using individual lists from secondary sources show limited utility as a measure of the food environment. This is because the thematic accuracy for the directories are represented by a PPV of 66% for Smiley and 72% for the CVR, representing the proportions of food retailers listed in the directories that are actually food retailers. Likewise, the sensitivity values of 82% for Smiley and 63% for the CVR represent the proportion of food retailers found through the field observations that were listed in the directories. The results have similarities to previous studies by Smiley [Toft et al., 2011], wherein an identical sensitivity of 82% was achieved, though the PPV was a great deal higher at 92%. The higher PPV obtained was most likely the result of that study's being limited to fast-food retailers.

Geographic accuracy clearly influences the applicability of the data. Analyses aggregating retailers over large areas or analysing distances to the nearest food retailer are less affected by geographical inaccuracy, particularly if the food environment is dense with retailers. On the other hand, areas with few food retailers and analyses at

small scales are vulnerable to geographic inaccuracy. In areas with a high density of food retailers, the distance in the analysis will theoretically have no impact because the direction of the errors should be random. Whether this holds true is doubtful, but it calls for further research to fully understand the nature of the errors. The aggregation of retailers over small areas will create errors, as exemplified in the CVR directory. In the CVR, 92.31% of the records were within 50 m, and according to the standard deviations, 95% should be within 58 m, but when aggregated into 250 × 250 m cells, more than 12% were aggregated incorrectly.
The completeness and thematic accuracy of the data demonstrates that if the raw data were used in research, there would exist a huge overrepresentation of food retailers, similar to findings from other studies [Liese et al., 2010]. The misclassification of retailers poses a major problem if analysing small retailer groups, such as specialty stores, whereas the errors have less of an impact on large groups, such as restaurants or supermarkets. The completeness of both the CVR and Smiley are poor in their raw state, as they are both missing retailers and have retailers that are in restricted areas, misclassified and non-existent

Previous studies have stated that individual lists of food retailers have limited utility for identifying food stores, but combining the lists improves the likelihood of a retailer being a food store [Hosler & Dharssi, 2010]. Combining the CVR and Smiley produced the same results, with sensitivity increasing to 93% but PPV still falling short of targets. Hence, a combination of the two directories is not a method for reaching a valid list of food retailers without field observations or another method. The expenses and time attached to field validation makes this method less attractive; therefore, alternatives were sought in Paper II. The approach applied a pre-classification method based on a combination of the NACE Classification and the business name. This has previously been successfully applied in Denmark on fast-food outlets [Toft et al., 2011], though as a general method for classification of food retailer types, it is less accurate. However, pre-classification could be a good supplement to field validation by limiting the number of retailers that need field validation to approximately 25 percent. The method has been tested only twice in Denmark, with various success, and more studies in different countries are needed to confirm or reject the potential of this method. The method demands knowledge about the tradition and culture of the food retailers, as well as the language to determine which words the classification should be based on. Whether this is possible in other countries is unknown. Alternatives to field validation and pre-classification could be methods from the big-data concept. Generally, big data have the potential to analyse unstructured data from several sources, such as restaurant homepages, commercial listings (google, yellow pages) or social media (Facebook, Twitter). The vast amounts of data could potentially help validate the secondary food source data but also could increase knowledge that is often difficult to gather without field validation (e.g., opening hours, the nutritional value of the menu).

5.2 INTEGRATION OF DATA—A ROAD TO USEFUL AND VALUABLE DATA …

The concept of the ecological model is to combine knowledge about several parameters to create a more detailed framework with which to improve the understanding of the mechanics that influence eating behaviour and physical activity level. In Denmark, a huge amount of information about individuals (CPR, health registers, etc.) is registered, as well as for the environment in which they live (BBR, ESR, CVR, Smiley, PlansystemDK, etc.). The locations of food sources, educational facilities, agriculture and land use in general are not random but are the result of a decade's regulation and well-established planning laws. Additionally, some parameters are more difficult targets of information gathering, such as priorities, life style and heritage. Whether it is possible or even a good idea to combine all these data from multiple structured and unstructured data into one model is not the focus for this project. Instead, the project's focus is on the availability of data and the possibility of integrating these data. Previously, data were kept separate with no common data model by which to facilitate integration. However, this has changed with the INSPIRE Directive and the open/free data policy. Common data models and data keys have improved the quality of several registers in Denmark. The Building and Housing Register (BBR) and Common Management Information System data (FLIS) are good examples of registers in which the Danish government has facilitated improvements of data quality and integration between data. In contrast stands the CVR register, with its poor accuracy. In Denmark, the government has chosen to use many resources to improve the quality of data in anticipation of the data's being used more frequently and thereby creating value. Of key importance is the possibility of integrating several data sources, as data often creates more value and insight through integration with other data. In many other countries, governments do not act as facilitating parties, and other initiatives and projects are needed, such as the Open Street Map, where everybody can participate in improving data quality. However, creating open-source projects to improve all data are not a valid approach, e.g., health data and other sensitive information will be much more difficult to gather from a larger population sample.

In Denmark, health data are still separated from other data, and implementation of health data in the Danish SDI would be a major step forward in promoting spatial health research because it allows for effective access and use. Second, data might improve the quality and efficiency of the healthcare system. The national registers in Denmark all hold information, such as CPR numbers and addresses, for simple linkages of registers and spatial datasets through the use of geographical information systems. Therefore, environmental data and health data can be combined when analysing the causes and outcomes of diseases. However, health data and environmental data have very different requirements and restrictions with regard to spatial and privacy relevance. Health data are subject to heavy legal and ethical restrictions. Modifying data to an aggregated form is essential to bypass these restrictions. However, some degree of detail is inevitably lost in the process. Conversely, legal or sensitivity issues do not restrict environmental data, even though the data refer to a person's activities, e.g., pollution. Health data are based on individuals who are spatially dynamic. Aggregated health data are a momentary representation of the spatial distribution of incidence and prevalence rates. Environmental data are more stationary insofar as they often only change over the long term. Environmental data are not subject to generalisation, in contrast to the aggregated health data. However, environmental data, such as UV and pollen, experience continuous extension and fluctuation over a yearly cycle. These types of data are often based on the interpolation between discrete features and average values over several years. People frequently move; thus, a higher frequency for updating health data in the SDI is needed. Environmental data, however, changes slowly over long periods and often does not significantly affect the spatial area across a year. Hence, an approach completely different from the one today should be used to populated health data into the national SDI.

Health registers have sensitive information, and most people would not like their personal medical history to be freely available. All parties involved with health data must ensure the confidentiality of individuals and protect the data both legally and technically. Patient-identifiable data are critical to medical research: updating, linking and validating data are impossible without person identifiable data, and the implementation of potential confounders in the analysis is difficult (Haynes et al, 2007). There is always a trade-off between the requirements of researchers and privacy concerns, despite the indisputable value of patient-identifiable data. Danish legislation limits granting access to individual health information to employees within the sector with relevant needs and to researchers who were granted permission through legal channels. Person-identifiable data cannot be part of INSPIRE, but by making the data anonymous, the data can legally be a part of the national geoportal. Aggregating health data into administrative units or grid systems is a possibility for making the data anonymous. However, the spatial units must contain enough observations such that individual persons or families are not recognisable. Deciding between administrative units or a particular grid size for aggregation is difficult due to the various needs at multiple spatial and temporal scales. However, grid systems have the advantage of remaining the same over time and avoiding the modifiable area unit problem (MAUP) (Openshaw, 1983).

Metadata are critical for an efficient SDI to support good data management and exchange information. The creation and publication of metadata are perceived as tedious and time-consuming; however, without metadata, there is a risk that datasets become redundant. Documentation of Danish health care data are unstructured and is probably mainly a result of the previous administration. Restructuring the metadata for health is required for consistency and organisation. There are several studies on the validation of registers with health data in Denmark (Abildstrøm and Madsen, 2011; Bjerregaard and Larsen, 2011; Gjerstorff, 2011; Green, 2011; Lyseen and Hansen, 2014; Sortsø, 2011; Thygesen et al, 2011), which could be applied for the creation of metadata information for health datasets. This would bring the metadata for health data in line with the metadata included in other national registers and spatial data, making integration of the data easier.

5.3 NEIGHBOURHOODS OR ACTIVITY SPACES? DIFFERENCE AND APPLICABILITY IN NUTRITION RESEARCH

The use of the term 'neighbourhood' in food environment research pertains to spaces defined by fixed boundaries, such as administrative units, or by a fixed distance, such as buffers, that define a school or residential neighbourhood. When referring to individually measured areas, a more appropriate term instead of neighbourhood is 'activity spaces', as suggested by Zenk and colleagues [Zenk et al., 2011]. This differentiation between terms can potentially improve researchers' understanding of the differences between the place-based and person-based exposure measures.

The understanding of place as a concept ranges from the individual adhering to his or her own unique place determined by everyday life and behaviour to the claim that the individual unconsciously relates his or her behaviour and choices to more structured patterns based on social and physical environment characteristics [Lytle, 2009]. However, a discussion about place is often ignored due to pragmatic considerations, such as data only being accessible at the level of administrative units. Administrative divisions as a concept for place are therefore often the natural choice for many researchers, without considering the administrative divisions' ability to encapsulate the relevant behaviour. The consequence is an erroneous assumption or generalisation that all individuals have equivalent behavioural patterns, thereby limiting exposure to a confined area and limiting diversity in food supply choices.

Such place-based neighbourhood definitions do not take into account diversity in individual behaviour. This problem is most likely the result of assuming people carry out most of their activities in their residential location, which is contradicted by the high mobility observed in the participant sample in Paper IV. The participants in this study are young adults, and most have a high mobility level, even absent the ability to drive a car. The participant's mobility must be taken into account because it weakens the influence of residential neighbourhoods. However, other studies with low-mobility group samples, such as the elderly and disadvantaged people, are probably more sensitive to residential neighbourhood exposure [Chaix et al., 2012].

Defining individual activity spaces is advantageous for providing increased specificity in a multiple-space exposure measurement. However, as Ball and colleagues note, collecting activity-space attribute data can be time- and labour-intensive because the individual activity spaces do not align spatially with existing administrative divisions [Ball et al., 2006]. The activity spaces defined by the individual's behaviour most likely vary in the size of the area, which increases the complexity of analysis when comparing different individuals' exposure. Moreover, comparisons across different studies are very difficult if the areas of their activity spaces vary in size. The equal size of neighbourhoods based on buffers makes them easier to compare between studies in different countries. However, buffers are limited to a few locations, and as revealed in Paper IV, buffers and administrative divisions have similar problems in capturing exposure during commuting or leisure-time activities.

The results of Paper IV are consistent with several other studies [Ball et al., 2006; Lytle, 2009; Burgoine & Monsivais, 2013; Chaix et al., 2012; Rainham et al., 2010] advocating for more individually based neighbourhood definitions that take into account multiple environments for exposure beyond the home, school or work communities. Exposure during commuting time and leisure activities are particularly difficult to incorporate when neighbourhoods are place-based. Kwan further questions the use of arbitrary definitions of neighbourhoods instead of considering the actual spaces in which individuals' exposures occur [Kwan, 2009]. The main objections to the static and administratively bounded spatial definitions in ecological exposure measures found in this study and accentuated by Kwan are the assumptions that (1) residential neighbourhoods are the most relevant in affecting food exposure and (2) individuals who live in the same spatial areas experience the same level of exposure, regardless of time spent in the area and residential locations within the area [Kwan, 2009]. The results from this study contradict these assumptions insofar as individuals also spend a substantial time outside their residential neighbourhoods, and variability in the sizes of individual activity spaces mirrors the variety observed in individuals' exposures.

5.4 MEASURING PEOPLE'S MOVEMENT—STRENGTHS AND LIMITATIONS IN A RESEARCH CONTEXT

Many studies that use GPS to measure people's movement or whereabouts often highlight how tracking is an objective measure compared with self-reported information. I will not disagree that GPS is a 'more' objective measure. However, I will object to this being an entirely objective measure because knowledge that their behaviour is being monitored also influences individuals' behaviour. People who know they are being monitored might not behave as they regularly do, and the opportunity to measure people without their awareness could yield different insights to their behaviour. There are clear ethical issues with tracking people without their consent but also great potential value. Video surveillance has existed for years, and many countries have strict laws about where, when and for which purposes video surveillance is allowed. However, tracking people without their consent by using the novel technologies of Bluetooth and Wi-Fi is less established, and most countries lack clear laws that address ethical concerns related to these technologies. A decade ago, tracking was limited to GPS, but the next generation of tracking by means of Bluetooth and Wi-Fi is already gaining ground in principal, larger cities. Smart city projects are establishing frequent Wi-Fi points in cities that provide free access to Wi-Fi. However, at same time, the movements of people throughout the city could potentially be tracked through these points. Where are the boundaries to such data collection, and for what can these type of data be used? In a maybe-not-that-far-off future, Wi-Fi hot spots could be citywide or perhaps even nationwide tracking mechanisms for the entire population. The potential uses for this type of data are huge, but such data also introduce both ethical and modelling challenges that need to be addressed.

5.5 DANISH-BASED STUDY—IS THERE ANY GENERIC VALUE TO THE STUDY?

The culture, laws and traditions of Denmark influence the context and settings in which this study has been performed. In Denmark, there is a long tradition of registering data on people, buildings, properties, businesses, etc. These data are not available for research in many countries; therefore, analysis often has to be made on survey data from a population sample. These data can often be expensive and time-consuming to gather. However, the idea of opening up national registers for public use is catching on in other countries and has the potential for increasing the value of data through its more frequent use. Establishing strategies at the national level are one option for improving the use and integration of such data. Conversely, this is not an option for all countries, due to different traditions in the registration of data and the countries' willingness to maintain records on people. The bases for conducting ecological analyses are obviously very different from country to country, but ensuring integration between data sources and the validation of data are important in all studies.

The food environment in Denmark is highly regulated because of planning laws that have restricted the free market regarding the location of food sources. How this regulation has shaped the supply and demand of food sources in different environments and how this influences individuals' eating behaviours are difficult to estimate from a national study. Through cross-country studies, these influences could potentially expand knowledge of food exposure and behaviour.

Tracking technologies have afforded new opportunities for researchers to gather information about an individual's whereabouts. In Denmark, tracking by GPS has been used in several studies, whereas studies using other technologies such as mobile positioning, Bluetooth and Wi-Fi are limited. Bluetooth and Wi-Fi are novel technologies for tracking behaviour, whereas mobile positioning has been available for some time. In Denmark, mobile positioning is restricted and is not currently a viable option for tracking an individual's whereabouts. In other countries, mobile positioning is not restricted in the same way and therefore makes mobile positioning a preferred option for some studies due to its high population penetration.

Defining neighbourhoods or activity spaces are difficult in all studies. Without proper attention to this part of a study, the basis for inferring valuable and correct conclusions is difficult. Many studies apply standard spatial divisions, such as administrative divisions or buffers, on a location. However, there are general problems with adopting static definitions of individuals' exposure. The availability of food retailers is not relevant because retailers are located within a pre-defined boundary or are within a certain distance. People are individuals who by default cannot be compared worldwide. So when researchers apply a buffer distance in Europe because another researcher in the USA or China used a certain distance, this is a major problem. These methods were initially appropriate when research was developing. However, different people even within the same country and city are more diverse today than ever. Consequently, standardised definitions of neighbourhoods or even activities might not be diverse enough to capture this individuality. Different cultures, heritages, traditions, preferences, etc., influence the choices people make as well as the environment. To capture the effect of the environment, all of these interpersonal influences have to be added to the model.

5.6 METHODOLOGICAL STRENGTHS AND LIMITATIONS OF THIS STUDY

In Paper II, all food sources have been implemented in the study. Several studies only look at one or two food sources and therefore miss many sources of potential exposure. Studies often use supermarkets as a source for healthy eating and fast-food outlets as an unhealthy eating source. However, there are many other options for unhealthy/healthy foods, such as gas stations, moveable food stalls, etc. Furthermore, supermarkets do not guarantee healthy eating behaviour; similarly, many fast-food outlets have healthy choices. Classifying food sources as healthy or unhealthy based on their name and classification are not necessarily accurate, and the nutritional value of each food source menu is not taken into account. In Paper IV, the diversity of food sources are not incorporated into the study because the objective focused on examining different spatial extents of neighbourhoods and not on the exact exposure. However, because supermarkets and fast-food outlets are used as proxies for healthy and unhealthy food, this could lead others to do the same. This was not the intention, and future research should focus on all sources in the food environment.

In Papers II and IV, the area of interest is limited by a set of spatial boundaries. These boundaries are based on administrative units and are not equivalent for both papers. The validity (sensitivity and PPV) of food sources has therefore only been established for a part of the area used in Paper IV. Second, the boundaries limit the area and location of food sources in an unfortunate way that leads to a problem known as the edge effect. Limiting the study to areas in which the population lives could be a problem. The location, type and number of food sources just outside the boundary is unknown and could potentially influence the results. In Paper IV, the food environment outside the boundary is unknown. However, the behaviours of the participants are known, and little time is spent outside the study's spatial boundaries. Nevertheless, all studies should examine whether the edge effect exerts an influence when deciding to spatially limit the area.

Individual characteristics (e.g., personal preferences) as confounders are crucial to take into account when analysing relationships between the food environment and health outcomes [Ball et al., 2006; Lytle, 2009]. However, not all preferences can be adjusted through common confounders such as income, ethnicity and education level. Consequently, methods used for defining neighbourhoods must accommodate the individual's behavioural characteristics [Chaix et al., 2012]. However, to achieve this effect, researchers must carefully scrutinise the behaviour to be measured to fully understand the phenomenon. The manner in which a space is defined should reflect the context in which it is applied [Kwan, 2012]. Therefore, to measure exposure to the food environment, researchers must make qualified assumptions about where people shop, the distances they are willing to travel to shop and other individual preferences. Thus, paying attention to the individual is important when developing studies of the interaction between the population and the environment. As Larson and Story concluded, most food environment studies have methodological problems that reduce the credibility of their findings [Larson & Story, 2009]. Problems occur with assessing physical access to food sources in the environment and linking access to a food source with food purchases and intake. Further analysis of individual behaviour could potentially be used to link food source exposure to individual food purchasing by analysing movement and stop flows in space-time data.

The survey period of one week in Paper IV is a short time frame for the analysis of the participants' behaviour. Short tracking periods could include infrequently visited locations and vice versa [Chaix et al., 2012]. This phenomenon illustrates a shortcoming of GPS technologies because recording continuous involvement at such a level for longer periods is difficult. The development of tracking technologies is a fast-growing field, and technologies such as Bluetooth, Wi-Fi and cellular phone networks could potentially be used to track participants in a way that requires less involvement from the individuals, mostly because all these technologies are included in most mobile phones today and therefore do not require participants to carry and maintain additional devices. The development of these technologies provides a promising improvement for empirical place research [Paper III].

In Paper IV, the GPS devices used were set to measure at seven-second intervals, which was the minimum interval available between loggings. A seven-second interval between registrations is a short time and discharges the battery faster than at a longer interval. A short interval between registrations is preferable for some uses, but the logging interval could probably be 15 s or more to measure the extent of the activity spaces. However, some problems occur with high registration frequency. Activity measured by GPS can experience periods with loss of data that interferes with the registration interval. Activity space measures as standard deviational ellipses are calculated from the centre of gravity of the measured point locations, and uneven intervals between registrations therefore affect the extent of the calculated spaces. Several methods have been proposed for resolving this issue by estimating missing data [Zenk et al., 2011] or interpolating between registrations. Furthermore, studies' ability to measure individuals' use of food retailers is dependent on a short interval between registrations. To detect stops at food retailers, several consecutive registrations at the same location are needed. Determining a maximum interval between registrations is difficult without further research, but a long interval between registrations results in a smaller dataset that is easier to analyse. Studies that apply GPS to measure activity must consider the accuracy required (interval between registrations) and the expected travel types and speed of participants.

Any study of this type must use the appropriate spatial area to measure the exposure. However, many studies have applied place-based neighbourhoods with little care to identifying these areas [Kwan, 2009]. Among the most discussed methodological issues in research applying spatial data are the Modifiable Area Unit Problem (MAUP). MAUP refers to the issue that the areal units to which data are assigned might influence results. Neighbourhoods based on administrative divisions or buffers, as in Paper IV, are highly susceptible to the MAUP. Place-based neighbourhoods allow little variation between individuals compared with person-based neighbourhoods. Large differences exist between individual activity spaces, such as the convex hull and standard deviational ellipses, where the standard deviation for each type of activity space is larger than the mean area size. This finding clearly indicates a large spread between individual activity spaces. Considering the actual spatial and temporal exposure would allow for a more accurate measure of exposure and address the MAUP [Kwan, 2009]. This result would allow individuals to have individualised exposure measures despite living in the same neighbourhoods.

6. CONCLUSION

This thesis shows that the use of traditionally defined neighbourhoods as a proxy for food environment exposure significantly differs from the use of individually defined activity spaces. People's behaviours are based on their perceptions of what reality is, not on reality itself. The world that is perceived is the world important to behaviour analysis. The food environment is perceived differently for individuals, and each person acts on his or her own perception. One environment is therefore not equal for all. All people have individual perceptions of the world and, therefore, individual food environments in which they interact. Human behaviour is influenced by individual attitudes, motives, interests, experiences and expectations, and by food sources' size, proximity and similarity, as well as the social and cultural situations. However, the same parameters influence people in different ways. Therefore, equivalent influences do not equally influence the same behaviours or exposures.

Current conceptualisations of food exposure in neighbourhoods are inadequate for addressing the complexity of human behaviour. 'Neighbourhood' is a fuzzy concept that varies in meaning depending on the study being conducted and on each person's individual perception of his or her neighbourhood. The complexity and heterogeneity of human mobility no longer appear to correspond to the use of residential neighbourhoods, which emphasises the need for methods and measures of individual activity and exposure. Paper IV shows that exposure to the food environment occurs in multiple environments, but measuring individuals' activity spaces in multiple environments is challenging. The lack of focus on neighbourhood or activity space definitions in studies of the food environment is unfortunate, mainly due to the large amount of research analysing relationships between the food environment and health outcomes in which no evidence demonstrates that the neighbourhood exposures used coincide with the actual exposure.

This thesis demonstrates that tracking technologies can provide space-time data on the behaviour of individuals and that these data can be used to define activity spaces for measuring exposure to the food environment. GPS has proven itself useful for tracking behaviour, whereas Bluetooth, Wi-Fi and mobile positioning remain novel technologies with little application in tracking behaviour. GPS tracking is expensive and time-consuming, whereas Bluetooth, Wi-Fi and mobile positioning potentially can reach a larger population sample at a lower cost. However, the technologies are still novel for tracking and need further research.

Measuring exposure to food sources require valid and complete data on the food sources, often over large areas. This thesis shows that the quality of food source data can be improved by using multiple sources. Combining the CVR and Smiley resulted in an excellent sensitivity (0.93), with only 15 retailers missing from both directories; however, without field observations, the retailers not targeted as selling food in the public space could not be removed from the directories, again leading to a misrepresentation of food retailers. The use of a pre-classification method can limit the required amount of field observations. Field observations are expensive and time-consuming to perform, and limiting the needed field observations by a factor of four is valuable. Adaption of the pre-classification method to other Danish and possibly Scandinavian directories is plausible with the current characteristics of the pre-classification, due to similarities in language, tradition and culture. Application of pre-classification to other countries' directories is believed to be possible if the criteria for classifying food retailers are modified to the cultures and traditions of the country's language and food environment.

Denmark has a vast amount of data in digital databases, but there are several issues that must be addressed before data are easy to integrate across spatial, demographic and health data. Issues with health data relate to duplicate registrations of the same data and a lack of access to structured metadata. Further efforts to anonymise data are required to ensure the privacy and confidentiality of health data. Spatial aggregation is the key to making the data anonymous, whereas the CPR ensures the linkage between the health data and the addresses that serve as the spatial reference. The lack of initiative to include spatial applications of health data on the agenda for the digitisation of health data hinders the implementation of health data in INSPIRE and the further use of health data. The issues of spatially aggregated health data and compliance must be addressed. The harmonisation and implementation of health data in accordance with the INSPIRE Directive are unstructured and slowed by a lack of strategic promotion of spatial health data. A large step towards integrating health data with spatial data are the removal of silo-based approaches to data storage and administration. Common keys across registers in health, demographic and spatial data can ensure easy connection of several data sources in large-scale ecological models. In an ecological model of obesity and physical activity, many parameters influence individuals' behaviour and health outcomes. Understanding the influence of food behaviour and the risk of getting chronic illnesses as diabetes, cardiovascular diseases and cancers is important for the general well-being of many countries' populations. There is a strong need to be able to integrate behaviour measures with food exposure and food intake as well as demographic data and health outcomes.

REFERENCES

Abbade, E. B., Dewes, H. and Ganji, V. (2015)
Behavioral and Societal Drivers of an Obesogenic Environment Worldwide. Nutrition & Food Science 45, no. 2.

Adler, P. A. and Adler, P. (2002)
The reluctant respondent. In J. F. Gubrium and J. A. Holstein (Editors), Handbook of interview research: Context and method Sage. Thousand Oaks, California.

Ahas, R., Aasa, A., Roose, A., Mark, Ü. and Silm, S. (2008)
Evaluating Passive Mobile Positioning Data for Tourism Surveys: An Estonian Case Study. Tourism Management 29: 469–486.

Apparicio, P., Cloutier, M. and Shearmur, R. (2007)
The Case of Montréal's Missing Food Deserts: Evaluation of Accessibility to Food Supermarkets. International Journal of Health Geographics 6, no. 1: 1–13.

Arafat, M., Salam, A. and Arafat, O. (2014)
The Association of Type 2 Diabetes with Obesity and Other Factors: in Multinational Community. International Journal of Pharmacy and Pharmaceutical Sciences 6, no. 9.

Austin, S. B., Melly, S. J., Sanchez, B. N., Patel, A., Buka, S. and Gortmaker, S. L. (2005)
Clustering of Fast-Food Restaurants Around Schools: A Novel Application of Spatial Statistics to the Study of Food Environments. American Journal of Public Health 95: 1575–1581.

Ball, K., Timperio, A. F. and Crawford, D. A. (2006)
Understanding Environmental Influences on Nutrition and Physical Activity Behaviors: Where Should We Look and What Should We Count? International Journal of Behavioral Nutrition and Physical Activity 3: 1–8.

Barnett, V. and Lewis, T. (1994)
Outliers in spatial statistical data. John Wiley, 3rd edition.

Bjerregaard, B. and Larsen, O. B. (2011)
The Danish Pathology Register. Scandinavian Journal of Public Health 39, no. 7: 72–74.

Block, J. P., Scribner, R. A. and DeSalvo, K. B. (2004)
Fast Food, Race/ethnicity, and Income: A Geographic Analysis. American Journal of Preventive Medicine 27, no. 3: 211–17.

Boulos, M. N. K., Curtis, A. J. and Abdelmalik, P. (2009)
Musings on Privacy Issues in Health Research Involving Disaggregate Geographic Data about Individuals. International Journal of Health Geographics 8: 46.

Braidwood, R. J. (1960)
The Agricultural Revolution. WH Freeman.

Bronfenbrenner, U. (2009)
The Ecology of Human Development: Experiments by Nature and Design. Harvard university press.

Bullock, D., Haseman, R., Wasson, J. and Spitler, R. (2010)
Automated Measurement of Wait Times at Airport Security. Transportation Research Record: Journal of the Transportation Research Board 2177: 60–68.

Burgoine, T. and Monsivais, P. (2013)
Characterising Food Environment Exposure at Home, at Work, and along Commuting Journeys Using Data on Adults in the UK. International Journal of Behavioral Nutrition and Physical Activity 10, no. 1: 1–11.

Burgoine, T., Forouhi, N. G., Griffin, S. J., Wareham, N. J. and Monsivais, P. (2014)
Associations between Exposure to Takeaway Food Outlets, Takeaway Food Consumption, and Body Weight in Cambridgeshire, UK: Population Based, Cross Sectional Study. British Medical Journal (BMJ) 348, no. 5: 1464–1464.

Carroll-Scott, A., Gilstad-Hayden, K., Rosenthal, L., Peters, S. M., McCaslin, C., Joyce, R. and Ickovics, J. R. (2013)
Disentangling Neighborhood Contextual Associations with Child Body Mass Index, Diet, and Physical Activity: The Role of Built, Socioeconomic, and Social Environments. Social Science & Medicine, Social Determinants of Child Health 95: 106–114.

Chaix, B., Kestens, Y., Perchoux, C., Karusisi, N., Merlo, J. and Labadi, K. (2012)
An Interactive Mapping Tool to Assess Individual Mobility Patterns in Neighborhood Studies. American Journal of Preventive Medicine 43: 440–450.

Chaix, B., Méline, J., Duncan, S., Merrien, C., Karusisi, N., Perchoux, C., Lewin, A., Labadi, K. and Kestens, Y. (2013)
GPS Tracking in Neighborhood and Health Studies: A Step Forward for Environmental Exposure Assessment, a Step Backward for Causal Inference? Health & Place 21: 46–51.

Chen, D., Lu, C., Kou, Y. and Chen, F., (2008)
On detecting spatial outliers. Geoinformatica 12, no. 4: 455–475

Christian, H., Giles-Corti, B., Knuiman, M., Timperio, A. and Foster, S. (2011)
The Influence of the Built Environment, Social Environment and Health Behaviors on Body Mass Index. Results from RESIDE. Preventive Medicine 53, no. 1–2: 57–60.

Cohen, D. A., Scribner, R. A. and Farley, T. A. (2000)
A Structural Model of Health Behavior: A Pragmatic Approach to Explain and Influence Health Behaviors at the Population Level. Preventive Medicine 30, no. 2: 146–154.

Colditz, G. A. (1999)
Economic Costs of Obesity and Inactivity. Medicine & Science in Sports & Exercise 31, no. 1: 663.

Cordain, L., Eaton, S. B., Sebastian, A., Mann, N., Lindeberg, S., Watkins, B. A., O'Keefe, J. H. and Brand-Miller, J. (2005)
Origins and Evolution of the Western Diet: Health Implications for the 21st Century. The American Journal of Clinical Nutrition 81, no. 2: 341–354.

Cowan, R. S. (1976)
The 'Industrial Revolution' in the Home: Household Technology and Social Change in the 20th Century. Technology and Culture: 1–23.

Cummins, S. and Macintyre, S. (2009)
Are Secondary Data Sources on the Neighbourhood Food Environment Accurate? Case-Study in Glasgow, UK. Preventive Medicine 49, no. 6: 527–528.

Danish Agency for Digitisation.
The Digital Path to Future Welfare - eGovernment Strategy 2011-2015. Rosendahls-Schultz Distribution. http://www.digst.dk/~/media/Files/Digitaliseringsstrategi/Tilgaengelig_engelsk_strategi.pdf [Accessed 8 November 2015].

Danish Agency for Digitisation. (2012)
Fact Sheet - Basic Data. http://www.digst.dk/Servicemenu/English/Digitisation/~/media/Files/English/Fact_sheet_BasicData_pdf.pdf [Accessed 8 June 2014].

Danish Business Agency.
Virk – Virk Data. http://datacvr.virk.dk/data/ [Accessed 17 September 2013].

Danish Geodata Agency.
Geodata-Info. http://www.geodata-info.dk/Portal/List.aspx. [Accessed 8 June 2014].

Danish Geodata Agency (2).
Open Public Geodata Overview. http://eng.gst.dk/media/gst/2364689/Openpublicgeodataoverview1.pdf. [Accessed 9 June 2015]

Danish Geodata Agency (3).
Tidsplan | Datafordeler.. http://datafordeler.dk/tidsplan/. [Accessed 6 September 2015]

Danish Government, Local Government Denmark, and Danish Regions. (2013)
Making eHealth Work, National Strategy for Digitalisation of the Danish Healthcare Sector 2013-2017. http://www.ssi.dk/English/HealthdataandICT/~/media/Indhold/DK%20-%20dansk/Sundhedsdata%20og%20it/NationalSundhedsIt/Om%20NSI/Strategy2013-17.ashx. [Accessed 8 June 2014].

Danish Health and Medicines Authority. (2013)
De Ti Kostråd. http://sundhedsstyrelsen.dk/da/nyheder/2013/de-ti-kostraad. [Accessed 14 Marts 2015].

Danish Ministry of the Environment. (2008)
Infrastructure for Spatial Information Act. https://www.retsinformation.dk/Forms/r0710.aspx?id=122571. [Accessed 8 June 2014].

Duran, A. C., Diez Roux, A. V., Latorre, M. D. R, and Jaime, P. C. (2013)
Neighborhood Socioeconomic Characteristics and Differences in the Availability of Healthy Food Stores and Restaurants in Sao Paulo, Brazil. Health & Place 23: 39–47.

EDINA.
GoGeo, The Place to Discover Geospatial Information and Services for those in Education and Research. http://www.gogeo.ac.uk/gogeo/index.htm. [Accessed 8 June 2014].

Eeg-Olofsson, K., Cederholm, J., Nilsson, P. M., Zethelius, B., Nunez, L., Gudbjörnsdóttir, S. and Eliasson, B. (2009)
Risk of Cardiovascular Disease and Mortality in Overweight and Obese Patients with Type 2 Diabetes: An Observational Study in 13,087 Patients. Diabetologia 52, no. 1: 65–73.

EPHT.
New Mexico Environmental Public Health Tracking Program. https://nmtracking.org/nm_epht/ [Accessed 1 February 2014]

European Commission.
About INSPIRE. INSPIRE - Infrastructure for Spatial Information in the European Community. http://inspire.ec.europa.eu/index.cfm/pageid/48. [Accessed 8 June 2014.]

Eurostat-European Commision. (2008)
NACE Rev. 2—Statistical Classification of Economic Activities in the European Community. Eurostat Methodologies and Working Papers, Office for Official Publications of the European Communities. Luxembourg, Luxembourg.

Fielding, J. E., Teutsch, S. and Breslow, L. (2010)
A Framework for Public Health in the United States. Public Health Reviews 32, no. 1: 174–189.

French, S. A., Story, M. and Jeffery, R. W. (2001)
Environmental influences on eating and physical activity. Annual Review of Public Health 22, no. 1: 309-335.

Galvez, M. P., Pearl, M. and Yen, I. H. (2010)
Childhood Obesity and the Built Environment: A Review of the Literature from 2008-2009. Current Opinion in Pediatrics 22, no. 2: 202–207.

Gardner, G. and Halweil, B. (2000)
Overfed and Underfed: The Global Epidemic of Malnutrition. WorldWatch Paper. Worldwatch Institute.

Geoteam. (2013)
Gpsnet.dk – Danmarks Præcise Netværk. http://www.geoteam.dk/produkter/gpsnet-dk/entreprenoerer.html. [Accessed 9 September 2015].

Giles-Corti, B., Timperio, A., Bull, F. and Pikora, T. (2005)
Understanding Physical Activity Environmental Correlates: Increased Specificity for Ecological Models. Exercise & Sport Sciences Reviews October 2005 33, no. 4: 175–181.

Glanz, K. (2009)
Measuring Food Environments: A Historical Perspective. American Journal of Preventive Medicine 36, no. 4: 93–98.

Glanz, K., Rimer, B. K. and Viswanath, K. (2008)
Health Behavior and Health Education: Theory, Research, and Practice. John Wiley & Sons.

Guh, D. P., Zhang, W., Bansback, N., Amarsi, Z., Birmingham, C. L. and Anis, A. H. (2009)
The Incidence of Co-Morbidities Related to Obesity and Overweight: A Systematic Review and Meta-Analysis. BioMed Central (BMC) Public Health 9, no. 1: 1–20.

Haghani, A., Hamedi, M., Sadabadi, K., Young, S. and Tarnoff, P. (2010)
Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors. Transportation Research Record: Journal of the Transportation Research Board 2160: 60–68.

Hampton, S. E, Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L., Duke, C. S. and Porter, J. H. (2013)
Big Data and the Future of Ecology. Frontiers in Ecology and the Environment 11, no. 3: 156–162.

Hansen, H. S. (2001)
A quasi-four dimensional database for the built environment. In: Westort, C.Y. (ed.) DEM 2001. Lecture Notes in Computer Science 2181: 48–59. Springer, Heidelberg.

Hansen, H. S., Schrøder, L., Hvingel, L. and Christiansen, J. S. (2011)
Towards Spatially Enabled e-Governance – A Case Study on SDI Implementation. International Journal of Spatial Data Infrastructure Research 6: 73-96.

Hansen, H. S., Hvingel, L. and Schrøder, L. (2013)
Open Government Data – a key element in the digital society. Lecture Notes in Computer Science 8061: 167 – 180.

Hawkins, D. M. (1980)
Identification of Outliers. Taylor & Francis.

Haynes C. L., Cook, G. A. and Jones, M. A. (2007)
Legal and Ethical Considerations in Processing Patient-Identifiable Data without Patient Consent: Lessons Learnt from Developing a Disease Register. Journal of Medical Ethics 33, no. 5: 302–307.

Hippocrates, C. F and Watts J. (1734)
Hippocrates upon air, water, and situation: upon epidemical diseases; and upon prognosticks, in acute cases especially. To this is added (by way of comparison) Thucydides's account of the plague of Athens, the whole translated, methodis'd, and illustrated with useful and explanatory notes. London; Printed for J. Watts.

HIVMapper. (2014)
Stat Compiler. http://www.hivmapper.com/ [Accessed 8 June 2014].

HIV Spatial Data Repository.
Spatial Data Repository. http://www.hivspatialdata.net/ [Accessed 8 June 2014].

Hosler, A.S. and Dharssi, A. (2010)
Identifying retail food stores to evaluate the food environment. American Journal of Preventive Medicine 39, 41–44.

Hurvitz, P. M. and Moudon, A. V. (2012)
Home Versus Nonhome Neighborhood: Quantifying Differences in Exposure to the Built Environment. American Journal of Preventive Medicine 42, no. 4: 411–417.

International Diabetes Federation.
Diabetes Atlas. International Diabetes Federation. http://www.idf.org/diabetesatlas. [Accessed 16 March 2015]

Jeffery, R.W., Baxter, J., McGuire, M. and Linde, J. (2006)
Are Fast Food Restaurants an Environmental Risk Factor for Obesity? International Journal of Behavioral Nutrition and Physical Activity 3: 2.

Jensen, V. M. and Rasmussen, A. W. (2011)
Danish Education Registers. Scandinavian Journal of Public Health 39, no. 7: 91–94.

Katzmarzyk, P. T. and Janssen, I. (2004)
The Economic Costs Associated With Physical Inactivity and Obesity in Canada: An Update. Canadian Journal of Applied Physiology 29, no. 1: 90–115.

Kegler, M. C., Swan, D. W., Alcantara, I., Feldman, L. and Glanz, K. (2013)
The Influence of Rural Home and Neighborhood Environments on Healthy Eating, Physical Activity, and Weight. Prevention Science 15, no. 1: 1–11.

Kegler, M. C., Alcantara, I., Veluswamy, J. K., Haardörfer, R., Hotz, J. A. and Glanz, K. (2012)
Results From an Intervention to Improve Rural Home Food and Physical Activity Environments. Progress in Community Health Partnerships 6, no. 3: 265–277.

Kemp, S. (2014)
More Mobile Connections Than People. We Are Social. http://wearesocial.net/blog/2014/12/mobile-connections-people/. [Accessed 8 November 2015].

Kemp, S. (2015)
Digital, Social & Mobile Worldwide in 2015. We Are Social. http://wearesocial.net/blog/2015/01/digital-social-mobile-worldwide-2015/.

Kelly, B., Flood, V. M. and Yeatman, H. (2011)
Measuring Local Food Environments: An Overview of Available Methods and Measures. Health & Place 17, no. 6: 1284–1293.

Kerr, J., Duncan, S. and Schipperjin, J. (2011)
Using Global Positioning Systems in Health Research: A Practical Approach to Data Collection and Processing. American Journal of Preventive Medicine 41, no. 5: 532–540.

Kestens, Y. and Daniel, M. (2010)
Social Inequalities in Food Exposure around Schools in an Urban Area. American Journal of Preventive Medicine 39: 33–40

Kestens, Y., Lebel, A., Chaix, B., Clary, C., Daniel, M., Pampalon, R., Theriault, M., Subramanian, S.V.P. and Miranda, J. J. (2012)
Association between Activity Space Exposure to Food Establishments and Individual Risk of Overweight. PLoS ONE 7, no. 8: 1-13.

Kwan, M. (2009)
From Place-Based to People-Based Exposure Measures. Social Science and Medicine 69, no. 9: 1311–1313.

Kwan, M. (2012)
The Uncertain Geographic Context Problem. Annals of the Association of American Geographers 102, no. 5, 958–968.

Larson, N. and Story, M. (2009)
A Review of Environmental Influences on Food Choices. Annals of Behavioral Medicine 8: 56–73.

Laxer, R.E.; Janssen, I.
The Proportion of Excessive Fast-Food Consumption Attributable to the Neighbourhood Food Environment among Youth Living within 1 Km of Their School. Applied Physiology, Nutrition, and Metabolism 2013, 39, 480–86.

Lear, S. A., Teo, K., Gasevic, D., Zhang, X., Poirier, P., Rangarajan, S., Seron, P., Kelishadi, R., Tamil, A. M., Kruger, A, Iqbal, R., Swidan, H., Gómez-Arbeláez, D., Yusif, R., Chifamba, J., Kutty, V. R., Karsidag, K., Kumar, R., Li, W., Szuba, A., Avezum, A., Diaz, R., Anand, S. S., Rosengren, A. and Yusuf, S. (2014)
The Association between Ownership of Common Household Devices and Obesity and Diabetes in High, Middle and Low Income Countries. Canadian Medical Association Journal 186, no. 4: 258–266.

Leung, C. W., Laraia, B. A., Kelly, M., Nickleach, D., Adler, N. E., Kushi, L. H. and Yen, I. H. (2011)
The Influence of Neighborhood Food Stores on Change in Young Girls' Body Mass Index. American Journal of Preventive Medicine 41: 43–51.

Liese, A. D., Colabianchi, N., Lamichhane, A. P., Barnes, T. L., Hibbert, J. D., Porter, D. E., Nichols, M. D. and Lawson, A. B. (2010)
Validation of 3 food outlet databases: Completeness and geospatial accuracy in rural and urban food environments. American Journal of Epidemiology 172, no. 11: 1324–1333.

Liu, G. C., Wilson, J. S. B., Qi, R. C. and Ying, J. D. (2007)
Green Neighborhoods, Food Retail and Childhood Overweight: Differences by Population Density. American Journal of Health Promotion 21, no. 4: 317–325.

Lu, C., Chen, D. and Kou, Y. (2004)
Multivariate Spatial Outlier Detection. International Journal on Artificial Intelligence Tools 13, no. 4: 801–811.

Lucan, S. C. (2015)
Concerning Limitations of Food-Environment Research: A Narrative Review and Commentary Framed around Obesity and Diet-Related Diseases in Youth. Journal of the Academy of Nutrition and Dietetics 115, no. 2: 205–212.

Lytle, L. A. (2009)
Measuring the Food Environment: State of the Science. American Journal of Preventive Medicine 36, no. 4: 134–144.

Mantoux, P. (2013)
The Industrial Revolution in the Eighteenth Century: An Outline of the Beginnings of the Modern Factory System in England. Routledge.

Markham, A. C. and Altmann J. (2008)
Remote Monitoring of Primates Using Automated GPS Technology in Open Habitats. American Journal of Primatology 70, no. 5: 495–499.

Marteau, T. M., Ogilvie, D., Roland, M., Suhrcke, M. and Kelly, M. P. (2011)
Judging Nudging: Can Nudging Improve Population Health? British Medicine Journal (BMJ) 342: 228.

Mathys, T. and Boulos, M. N. K. (2011)
Geospatial Resources for Supporting Data Standards, Guidance and Best Practice in Health Informatics. BioMed Central (BMC) Research Notes 4, no. 1: 19.

McKinnon, R. A., Reedy, J., Morrissette, M. A., Lytle, L. A. and Yaroch, A. L. (2009)
Measures of the Food Environment: A Compilation of the Literature, 1990–2007. American Journal of Preventive Medicine 36, no. 4: 124–133.

McShane, R. and Skelt L. (2009)
GPS Tracking for People with Dementia. Working With Older People 13, no. 3: 34–37.

Mikkelsen, B. E. (2011)
Images of Foodscapes: Introduction to Foodscape Studies and Their Application in the Study of Healthy Eating out-of-Home Environments. Perspectives in Public Health 131, no. 5: 209–216.

Mikkelsen, B. E., Lyseen, A. K., Dobroczynski, M and Hansen, H. S. (2014)
Behavioural Nutrition & Big Data: How Geodata, Register Data & GPS, Mobile Positioning, Wi-Fi, Bluetooth & Thermal Cameras Can Contribute to the Study of Human Food Behaviour. Wageningen, The Netherlands, 2014.

Moore, L. V. and Diez Roux, A. V. (2006)
Associations of Neighborhood Characteristics With the Location and Type of Food Stores. American Journal of Public Health 96, no. 2: 325–331.

Murakami, E. and Wagner, D. P. (1999)
Can Using Global Positioning System (GPS) Improve Trip Reporting? Transportation Research Part C: Emerging Technologies 7, no. 2–3: 149–165.

Neckerman, K. M., Bader, M. D. M., Richards, C. A., Purciel, M., Quinn, J. W., Thomas, J. S., Warbelow, C., Weiss, C. C., Lovasi, G. S. and Rundle, A. (2010)
Disparities in the Food Environments of New York City Public Schools. American Journal of Preventive Medicine 39, no. 3: 195–202.

Ngo, J., Engelen, A., Molag, M., Roesle, J., García-Segovia, P. and Serra-Majem, L. (2009)
A review of the use of information and communication technologies for dietary assessment. British Journal of Nutrition 101: 102-112.

Odoms-Young, A. M., Zenk, S. and Mason, M. (2009)
Measuring Food Availability and Access in African-American Communities: Implications for Intervention and Policy. American Journal of Preventive Medicine 36, no. 4: 145–150.

Oreskovic, N. M. A., Winickoff, J. P. B., Kuhlthau, K. A. C., Romm, D. C. and Perrin, J. M. B. (2009)
Obesity and the Built Environment among Massachusetts Children. Clinical Pediatrics 48 no. 9: 904–912.

Paquet, C., Daniel, M., Kestens, Y., Léger, K. and Gauvin, L. (2008)
Field validation of listings of food stores and commercial physical activity establishments from secondary data. International Journal of Behavioral Nutrition and Physical Activity 5: 58.

Pearce, J. A., Hiscock, R. A., Blakely, T. B. and Witten, K. C. (2009)
A National Study of the Association between Neighbourhood Access to Fast-Food Outlets and the Diet and Weight of Local Residents. Health and Place 15: 193–197.

Pedersen C. B. (2011)
The Danish Civil Registration System. Scandinavian Journal of Public Health 39, no. 7: 22–25.

Petersson, F., Baadsgaard, M. and Thygesen, L. C. (2011)
Danish Registers on Personal Labour Market Affiliation. Scandinavian Journal of Public Health 39, no. 7: 95–98.

Phillips, M. L., Hall, T. A., Esmen, N. A., Lynch, R. and Johnson, D. L. (2001)
Use of Global Positioning System Technology to Track Subject's Location during Environmental Exposure Sampling. Journal of Exposure Analysis & Environmental Epidemiology 11, no. 3: 207.

Powell, L. M., Han, E., Zenk, S. N., Khan, T., Quinn, C. M., Gibbs, K. P., Pugach, O., Barker, D. C., Resnick, E. A., Myllyluoma, J. and Chaloupka, F. J. (2011)
Field Validation of Secondary Commercial Data Sources on the Retail Food Outlet Environment in the U.S. Health & Place 17, no. 5: 1122–1131.

Qi, F. and Du, F. (2013)
Trajectory Data Analyses for Pedestrian Space-Time Activity Study. Journal of Visualized Experiments 72.

R Core Team. (2014)
R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org/ [Accessed 19 February 2015].

Rajkovic, N., Zamaklar, M., Lalic. K, Jotic, A., Lukic, L., Milicic, T., Singh, S., Stosic, L. and Lalic., N. M. (2014)
Relationship between Obesity, Adipocytokines and Inflammatory Markers in Type 2 Diabetes: Relevance for Cardiovascular Risk Prevention. International Journal of Environmental Research and Public Health 11, no. 4: 4049–4065.

Renehan, A. G., Tyson, M., Egger, M., Heller, R. F. and Zwahlen, M. (2008)
Body-Mass Index and Incidence of Cancer: A Systematic Review and Meta-Analysis of Prospective Observational Studies. The Lancet 371, no. 9612: 569–578.

Rainham, D., McDowell, I., Krewski, D. and Sawada, M. (2010)
Conceptualizing the Healthscape: Contributions of Time Geography, Location Technologies and Spatial Ecology to Place and Health Research. Social Science & Medicine 70, no. 5: 668–676.

Rodriguez, D. A., Brown, A. L. and Troped, P. J. (2014)
Portable Global Positioning Units to Complement Accelerometry-Based Physical Activity Monitors. Medicine and Science in Sports and Exercise 37, no. 11: 572–581.

Rossen, L. M., Curriero, F. C., Cooley-Strickland, M. and Pollack, K. M. (2013)
Food Availability En Route to School and Anthropometric Change in Urban Children. Journal of Urban Health 90, n. 4: 653–666.

Ruhm, C. J. (2012)
Understanding Overeating and Obesity. Journal of Health Economics 31, no. 6: 781–796.

SAHSU.
The environment and health atlas for England and Wales. Small Area Health Statistics Unit. http://www.envhealthatlas.co.uk/ [Accessed 1 February 2014].

Sallis, J. F., Cervero, R. B., Ascher, W., Henderson, K. A., Kraft, M. K. and Kerr, J. (2006)
An Ecological Approach to Creating Active Living Communities. Annual Review of Public Health 27: 297–322.

Sallis, J. F. and Glanz, K. (2009)
Physical Activity and Food Environments: Solutions to the Obesity Epidemic. Milbank Quarterly 87, no. 1: 123–154.

Scarborough, P., Bhatnagar, P., Wickramasinghe, K. K, Allender, S, Foster, C. and Rayner, M. (2011)
The Economic Burden of Ill Health due to Diet, Physical Inactivity, Smoking, Alcohol and Obesity in the UK: An Update to 2006–07 NHS Costs. Journal of Public Health 33, no. 4: 527–535.

Seliske, L. M., Pickett, W., Boyce, W.F. and Janssen, I. (2009)
Density and Type of Food Retailers Surrounding Canadian Schools: Variations across Socioeconomic Status. Health Place 15, no. 3: 903–907

Shekhar, S. and Chawla, S. (2003)
Spatial databases: a tour. Prentice Hall.

Shekhar, S., Lu, C. and Zhang, P. (2003)
A Unified Approach to Detecting Spatial Outliers. GeoInformatica 7, no.2: 139

Stopher, P., FitzGerald, C. and Zhang, J. (2008)
Search for a Global Positioning System Device to Measure Person Travel. Transportation Research Part C: Emerging Technologies 16, no. 3: 350–369.

Storgaard, R. L., Hansen, H. S., Aadahl, M. and Glumer, C. (2013)
Association between neighbourhood green space and sedentary leisure time in a Danish population. Scandinavian Journal of Public Health 41: 846-852.

Sturm, R. (2008)
Disparities in the Food Environment Surrounding US Middle and High Schools. Public Health 122, no. 7: 681–690.

Sui, D. Z. (2004)
Tobler's First Law of Geography: A Big Idea for a Small World? Annals of the Association of American Geographers 94, no. 2: 269–277.

Toft, U., Erbs-Maibing, P. and Glümer, C. (2011)
Identifying fast-food restaurants using a central register as a measure of the food environment. Scandinavian Journal of Public Health 39, no. 8: 864–869.

Thompson, F. E., Subar, A. F., Loria, C. M., Reedy, J. L. and Baranowski, T. (2010)
Need for technological innovation in dietary assessment. Journal of American Diet Associasion 110: 48.

Thompson, J., Eagleson, S, Ghadirian, P. and Rajabifard, A. (2009)
SDI for Collaborative Health Services Planning. Global Spatial Data Infrastructures World Conference, Rotterdam, The Netherlands.

Thornton, L. E., Pearce, J. R., Macdonald, L., Lamb, K. E. and Ellaway, A. (2012)
Does the Choice of Neighbourhood Supermarket Access Measure Influence Associations with Individual-Level Fruit and Vegetable Consumption? A Case Study from Glasgow. International Journal of Health Geographics 11: 29.

Tyrovolas, S., Koyanagi, A., Garin, N., Olaya, B., Ayuso-Mateos, J. L., Miret, M, Chatterji, S., Tobiasz-Adamczyk, B., Koskinen, S., Leonardi, M. and Haro, J. M. (2015)
Diabetes Mellitus and Its Association with Central Obesity and Disability among Older Adults: A Global Perspective. Experimental Gerontology 64: 70–77.

U.S. Department of Agriculture, U.S. Department of Health and Human Services. (2010)
Dietary Guidelines for Americans. Washington, DC: U.S. Government Printing Office. 7th Edition.

Van Dyck, D., Cerin, E., Conway, T. L., Bourdeaudhuij, I. D., Owen; N., Kerr; J., Cardon, G., Frank, L. D., Saelens, B. E. and Sallis, J. F. (2012)
Perceived Neighborhood Environmental Attributes Associated with Adults' Transport-Related Walking and Cycling: Findings from the USA, Australia and Belgium. International Journal of Behavioral Nutrition and Physical Activity 9, no. 70: 1–14.

Vázquez, L. A., Rodríguez, Á., Salvador, J., Ascaso, J. F., Petto, H. and Reviriego, J. (2014)
Relationships between Obesity, Glycemic Control, and Cardiovascular Risk Factors: A Pooled Analysis of Cross-Sectional Data from Spanish Patients with Type 2 Diabetes in the Preinsulin Stage. BioMed Central (BMC) Cardiovascular Disorders 14, no. 1: 1–8.

Versichele, M., Neutens, T., Delafontaine, M. and Van de Weghe, N. (2012)
The Use of Bluetooth for Analysing Spatiotemporal Dynamics of Human Movement at Mass Events: A Case Study of the Ghent Festivities. Applied Geography 32: 208–220.

Wang, M. C., Kim, S., Gonzalez, A. A., MacLeod, K. E. and Winkleby, M. A. (2007)
Socioeconomic and Food-Related Physical Characteristics of the Neighbourhood Environment Are Associated with Body Mass Index. Journal of Epidemiology and Community Health 61, no. 6: 491–498.

Wang, Y., Beydoun, M. A., Liang, L., Caballero, B. and Kumanyika, S. K. (2008)
Will All Americans Become Overweight or Obese? Estimating the Progression and Cost of the US Obesity Epidemic. Obesity 16, no. 10: 2323–2330.

Wang, Z. and Lee, C. (2010)
Site and neighborhood environments for walking among older adults. Health Place 16, no. 6:1268-1279.

Wang, Y. C., McPherson, K., Marsh, T., Gortmaker, S. L and Brown, M. (2011)
Health and Economic Burden of the Projected Obesity Trends in the USA and the UK. The Lancet 378, no. 9793: 815–825.

WHO (1). (2015)
Obesity and Overweight. WHO. http://www.who.int/mediacentre/factsheets/fs311/en/. [Accessed 14 March 2015].

WHO (2).
GeoNetwork - The Portal to Spatial Data and Information. http://apps.who.int/geonetwork/srv/en/main.home [Accessed 8 June 2014].

Wing, R. R., Goldstein, M. G., Acton, K. J., Birch, L. L., Jakicic, J. M., Sallis, J. F., Smith-West, D., Jeffery, R. W. and Surwit, R. S. (2001)
Behavioral Science Research in Diabetes Lifestyle Changes Related to Obesity, Eating Behavior, and Physical Activity. Diabetes Care 24, no. 1: 117–123.

Withrow, D. and Alter, D. A. (2011)
The Economic Burden of Obesity Worldwide: A Systematic Review of the Direct Costs of Obesity. Obesity Reviews 12, no. 2: 131–141.

Zhang, W., Bai, X., Ge, H., Cui, H., Wei, Z. and Han, G. (2014)
Meta-Analysis in the Association between Obesity and Risk of Thyroid Cancer. International Journal of Clinical and Experimental Medicine 7, no. 12: 5268–5274.

Zenk, S. N., Schulz, A. J., Matthews, S. A., Odoms-Young, A., Wilbur, J., Wegrzyn, L., Gibbs, K., Braunschweig, C. and Stokes, C. (2011)
Activity Space Environment and Dietary and Physical Activity Behaviors: A Pilot Study. Health & Place 17, no. 5: 1150–1161.

APPENDICES

APPENDIX I – PAPER I: INSPIRE COMPLIANCE OF PUBLIC HEALTH INFORMATION – A DANISH CASE STUDY

International Journal of Spatial Data Infrastructures Research, 2015, Vol.10, 84-102

INSPIRE Compliance of Public Health Information – A Danish Case Study*

Anders K. Lyseen[1], Henning Sten Hansen[2]

[1,2]Department of Development and Planning, Aalborg University, Denmark, alyseen@plan.aau.dk ; hsh@plan.aau.dk

Abstract

Geographical information systems have become important to research, planning, commercial businesses, and health organisations in the public and private sectors. Data management and sharing are advantageous considering that repeating tasks is costly. The existence of several versions of the 'same' dataset raises concerns over data reliability and authority. Digitisation, which largely involves spatial information, is one approach for sharing data. Thus, digitisation is a vital part of the Danish e-government strategy. A well-functioning spatial data infrastructure (SDI) is an important prerequisite for e-governance. Implementation of the INSPIRE Directive has placed emphasis on SDI within key ministries and has resulted in several national services with free access to spatial data. However, until now, public health information has not been a part of the Danish SDI. In Denmark, several organisations have created independent public health datasets, and the infrastructure of the data is undocumented. Obtaining an overview of the available health data suitable for spatial applications is not easy. Most public health data do not have any spatial references, but it should be linked to features with a spatial reference, for example, administrative units or addresses. According to Danish legislation, health information is private, which imposes great limitations on the use of health data. Human health information should not be isolated, which is more or less the situation today, but rather seamlessly combined with other data. The aim of the current research is to identify available public health data in Denmark, including links to spatially referenced features, and to analyse its compliance with

DOI: 10.2902/1725-0463.2015.10.art4

International Journal of Spatial Data Infrastructures Research, 2015, Vol.10, 84-102

the principles of the INSPIRE Directive. The INSPIRE Directive includes the theme of human health and safety, and it is the basis for identifying available health data.

Keywords: Health, Spatial Data Infrastructure, INSPIRE, Open Data

1. INTRODUCTION

Geographical information systems have become important in research, planning, commercial business, and health organisations in the public and private sectors. Geography is important in understanding the dynamics of health, and the location or spatial reference is often the key to joining many health data and spatial datasets (Boulos, 2005). Epidemiologists and medical geographers have applied the spatial dimension to health data to understand the causes and spread of diseases. The earliest examples date back to Finke and Seaman in the 1790s (Barrett, 2000). Presumably, the most famous example is the mapping of cholera in 1854 (Snow, 1855). Human health is influenced by the environment in which the person lives and by socioeconomic factors. The spread of disease and other threats to health (e.g., air pollution) do not follow administrative borders. Human health information should not be isolated from spatial datasets, as is occurring today. Hence, there is a need to join health and environment datasets (Thompson et al, 2009).

Most countries face increases in the expenditures for healthcare, which are largely a result of an ageing population and chronic diseases associated with particular lifestyles. An effective primary healthcare system plays an important role in managing the increasing demands on the health care system. For the healthcare system to be effective, collaboration through sharing of health data and data on accessibility of health services for targeting limited resources are required (Thompson et al, 2009). An efficient spatial data infrastructure (SDI) would enable effective access and use of health datasets and would possibly improve the quality and efficiency of the healthcare system (Murdoch and Detsky, 2013). Data management and sharing have many benefits considering that we cannot afford to repeat tasks due to limited resources. Multiple versions of the 'same' dataset raise concerns regarding the reliability and authority of datasets.

Digitisation of public spatial data is on the agenda for many governments, and each country is developing their own national SDIs (EDINA). Commonly, national SDIs focus on traditional spatial data, such as addresses, property information, spatial planning, remote sensing data and environmental data. However, at best, spatial health data have a very limited implementation in national SDIs; Denmark has only a few health datasets available (Danish Geodata Agency). There are several international databases that distribute spatial health data, i.e., Worldmapper, those listed in (Mathys and Boulos, 2011), the World Health Organisation (World Health Organisation (1)), HIVMapper (HIVMapper) and HIV Spatial Data Repository (HIV Spatial Data Repository). Only a few countries have or are developing spatial data infrastructures that include health data. In the United Kingdom, an environment and health atlas (SAHSU) has been published. Additionally, the SDI Go-Geo (EDINA) delivers metadata, spatial data and interactive mapping to medical researchers, public health officials and the general public (Mathys and Boulos, 2011). New Mexico has developed a web portal to integrate environmental information and health information (EPHT). In Victoria, Australia, spatial data access and management is a priority. An SDI to increase and strengthen effective collaboration within health projects and add benefits through increased use of under-used data is being developed (Thompson et al, 2009).

Health data include some of the most personal and private information on people. However, patient-identifiable data are critical to medical research considering that updating, linking and validating data are impossible without identifiable data, and the implementation of potential confounders in the analysis is difficult (Haynes et al, 2007). Addresses are needed for analysis and spatial aggregation of data. Spatial aggregation is a means of preserving confidentiality while maintaining an acceptable level of data usefulness (Boulos et al, 2009). The privacy concerns additionally challenge health researchers due to the expensive and time-consuming methods to secure data anonymity or confidentiality. Ultimately, there is always a trade-off between privacy concerns and the types and accuracy of possible spatial analyses of health (Boulos et al, 2009).

Within the healthcare sector, there are vast amounts of data, but the application of spatial information in the healthcare sector has been ad hoc and uncoordinated (Thompson et al, 2009). The large number of agencies involved in healthcare systems requires common data systems to coordinate the sharing of datasets efficiently (Thompson et al, 2009). The data assembly at the Statens Serum Institute (SSI) in Denmark has combined previously independent data administrations and centralised the administration of the central health care registers in Denmark. However, obtaining an overview of available health data suitable for spatial applications is not easy. Most public health data do not have any spatial references, but it needs to be linked to features with a spatial reference, for example, administrative units or addresses.

Therefore, the aim of the current research is to identify available health data in Denmark, including links to spatially referenced features, and to analyse its compliance with the principles of the INSPIRE Directive.

The paper begins with a description of the Danish SDI and the implementation of the INSPIRE Directive in Denmark. Then, the Danish health data infrastructure is

International Journal of Spatial Data Infrastructures Research, 2015, Vol.10, 84-102

described, along with initiatives for the development of the data infrastructure in the Danish healthcare system. The national registers, spatial web services, and their relationships are introduced. The compliance of health data with the INSPIRE Directive is analysed. Finally, the issues related to the compliance and the inevitable privacy concerns of health data are discussed.

2. THE DANISH SPATIAL DATA INFRASTRUCTURE

The Danish national spatial data infrastructure is a common foundation for the management of geographical information and digital administration in Denmark (Hansen et al. 2011). A well-functioning spatial data infrastructure is an important prerequisite for e-governance. The implementation of the INSPIRE Directive has emphasised SDI within key ministries and has resulted in several national services with free and easy access to spatial data on the environment, spatial planning, addresses, cadastral maps, and topography. However, until now, public health information has not been a part of the Danish national SDI. In 2008, the Danish government enacted The Act on Infrastructure for Geographic Information as a response to the European INSPIRE Directive. The act ensures the implementation of the INSPIRE Directive in Denmark and that the common framework can be applied widely in the national geographical data infrastructure. The Danish spatial data infrastructure follows the basic principles of INSPIRE and establishes common rules, conditions and guidelines for the data, services, technologies, metadata, and the organisation (Danish Ministry of the Environment, 2008). The Danish infrastructure for spatial information consists of data and services through web portals or standardised services, such as Web Map Services (WMS) and Web Feature Services (WFS). A web-based portal publishes data and metadata such that users have one place to search and access the data (Danish Geodata Agency, 2014). For a dataset to be a part of the Danish spatial data infrastructure, it must be digital, nationwide and nationally applicable, and statutory (Danish Ministry of the Environment, 2008).

In 2011, The Agency for Digitisation was established, and the national e-government strategy accelerated the process of modernising Danish society (Danish Agency for Digitisation, 2011). The digital-based ambitions are described in the Danish e-government strategy of 2011-2015 (Danish Agency for Digitisation, 2011). The government, municipalities and regions included in the strategy should increase the momentum of digitisation in the public sector. Spatial information and the associated infrastructure play an important role in delivering data for public administration. One of the initiatives is common basic data (Danish Agency for Digitisation, 2012) for all public authorities. Since 2013, spatial and register data have been freely available for the public (Hansen et al., 2013). The strategies program for basic data are directly based on the principles of the INSPIRE Directive to ensure consistency with INSPIRE in the development of the national infrastructure for geographical information and the digital public administration.

INSPIRE also contributes to the coherence and exchange of data across public authorities using international standards, including INSPIRE. Data must comply with the INSPIRE principles to be included in the Danish SDI and INSPIRE.

The principles of the INSPIRE Directive are also utilised as a basis for the development of non-spatial public data collections and the general digital infrastructure in Denmark.

3. HEALTH DATA INFRASTRUCTURE

In 2012, the important national health data at SSI was consolidated to ensure equal and transparent conditions for the use of data and to improve the data quality and sharing for health professionals and researchers. The role of SSI is to gather, analyse and disseminate data. Following the national e-government strategy of 2011-2015 (Danish Agency for Digitisation, 2011), a national strategy for the Digitisation of the Danish Healthcare Sector 2013-2017 was developed. One of the five main initiatives is ‘better use of data’ (Danish Government et al, 2013). This strategy should create a basis for the affordable maintenance of health data, collaboration of information technology across the health sector, and improved quality of health data through ensuring a reliable link between the local health departments and the national registers (Danish Government et al, 2013). The strategy facilitates the management and sharing of health data through a common infrastructure and standards for data, interfaces and services. However, the national strategy for Digitisation of the Danish Healthcare Sector 2013-2017 does not mention INSPIRE or the promotion of spatial applications of health data.

The INSPIRE Directive includes the theme of human health and safety (INSPIRE Thematic Working Group Human Health and Safety, 2011), defined as ‘The geographical distribution of dominance of pathologies, information indicating the effect on health or well-being of humans linked directly or indirectly to the quality of the environment’. The INSPIRE Directive on human health and safety covers a range of data on diseases and health-related problems, as well as other indicators of health effects that are linked directly or indirectly to the environment. The theme of human health and safety involves health conditions of individuals and the population related to the INSPIRE theme population distribution – demography (INSPIRE Thematic Working Group Population Distribution, 2012). The characteristics of a population at a relevant spatial scale are important for analysing human health. The international classification of disease and related health problems (ICD) (World Health Organisation) are used to categorise disease, health-related conditions and external causes of disease and injury within the INSPIRE Directive. Diseases, injuries and accident data are expressed as raw incidence, prevalence or mortality rates under the INSPIRE Directive. Health data can be stratified by gender, age, socioeconomic indicators, or living conditions (urban or rural) (INSPIRE Thematic Working Group Human Health and Safety, 2011).

4. DANISH REGISTERS AND SPATIAL WEB SERVICES

Spatial enabling public registers require a common key attribute (Hansen, 2001). The address are an important and unambiguous database key in many private and public registers in Denmark, and all Danish addresses have a spatial reference. The address is easily recognisable and is used to locate residences in the Building and House Register (BBR) and the Central Business Register (CVR). Addresses are also registered in the Civil Person Register (CPR) (Pedersen, 2011), which contains information on individuals in Denmark. CPR is key for linking health data (Bjerregaard and Larsen, 2011), social data, labour market data (Petersson et al, 2011) and education data (Jensen and Rasmussen, 2011). CPR contains addresses that can join a spatially referenced address dataset and can thereby establish a relation between health data and spatial data on, for example, pollution in the Danish Natural Environment portal. All companies, institutions, and public service providers are registered in the CVR and are uniquely identified through the CVR number. The CVR also has information on addresses and industrial classifications. CPR, CVR, BBR, the Property Register (ESR) and the Cadastre Register are part of the Basic Data programme in Denmark.

Statistics Denmark is the national agency for statistical data, and the data are obtained from national public registers. The statistical data are output from individual data aggregated over administrative units, such as parishes, municipalities, regions or the entire country. Additionally, data may be aggregated over miscellaneous spatial units or the Danish National Grid. Aggregation is limited by enforced restrictions to ensure the confidentiality and privacy of individuals. The restrictions enforced by Statistics Denmark require that persons are not identifiable in the data. Therefore, the minimum number of persons or households in an aggregation is variable, depending on the type of data. However, threshold values of a minimum of 50 properties or 100 persons within each spatial unit are often required.

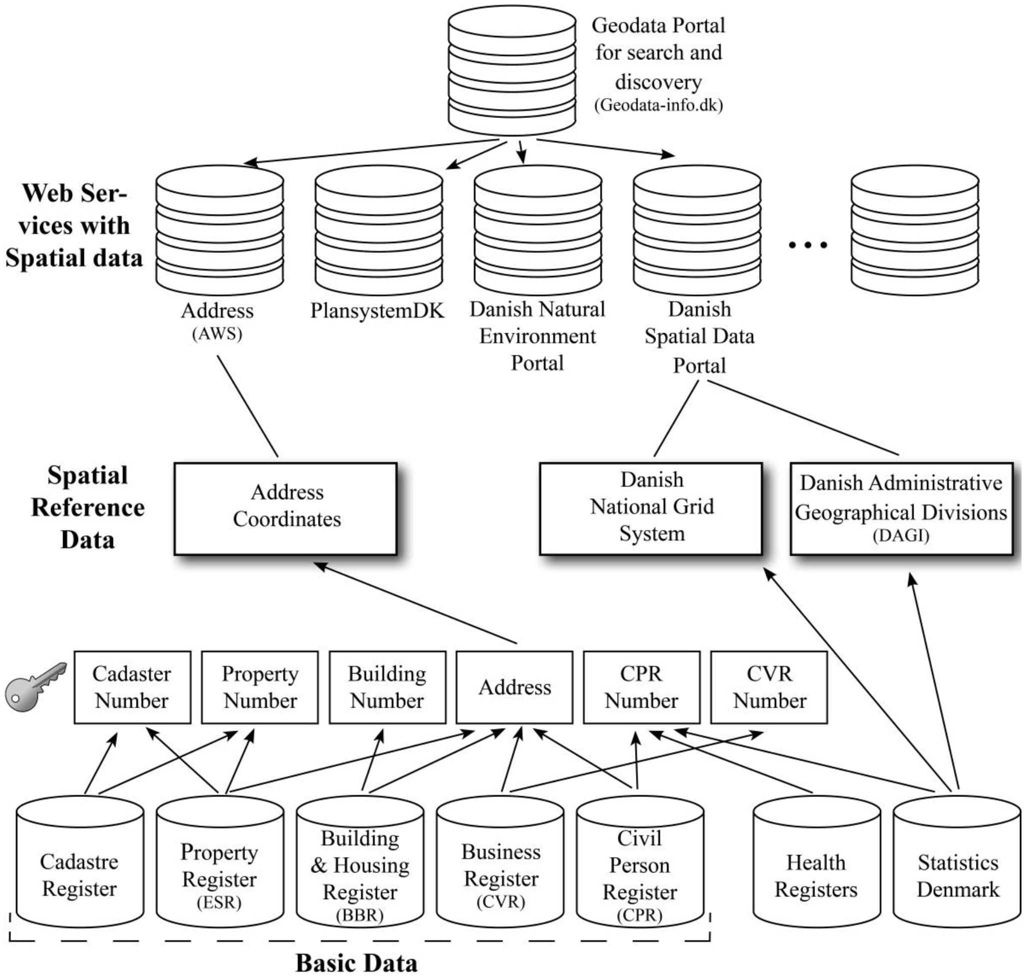
Geodata-info is the national counterpart to the INSPIRE portals that aims to search and discover spatial datasets and associated metadata in Denmark (Danish Geodata Agency). Figure 1 presents the associations between the spatial data repositories, the spatial reference data, data keys, national basic data registers, national health registers and Danish statistics data. Through data keys and spatial references, all the data sources can be related.

In Denmark, the Statens Serum Institute (SSI) maintains the health data registers. The National Institute of Public health (SIF) maintains a few clinical registers and

International Journal of Spatial Data Infrastructures Research, 2015, Vol.10, 84-102

conducts a national health survey every three years. Table 1 lists registers that contain health data.

Figure 1: The Association between National Registers through their Key Attributes and the Spatial Reference from Spatial Web Services.



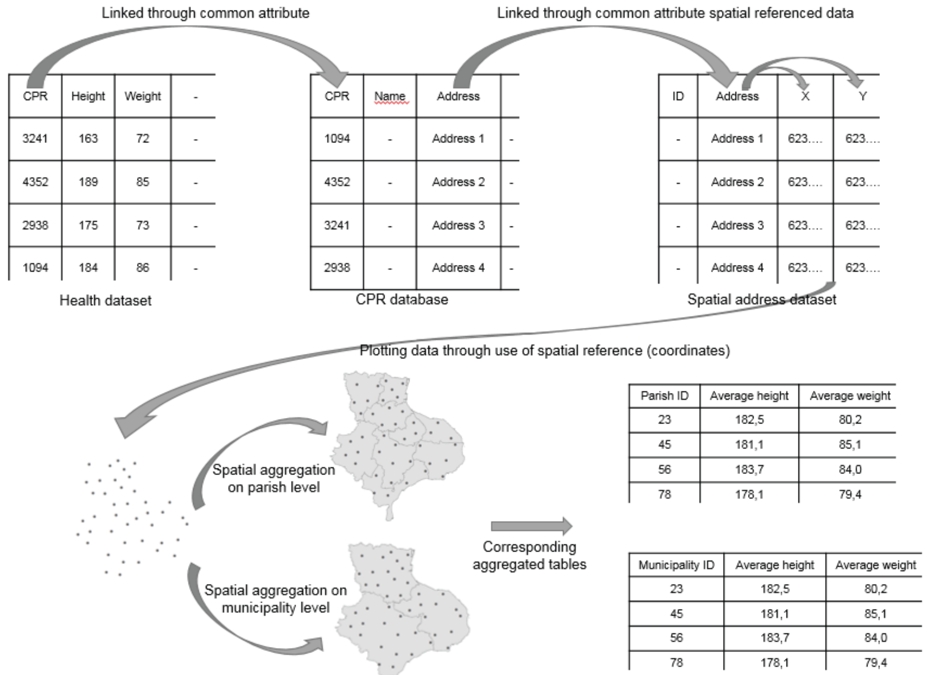
International Journal of Spatial Data Infrastructures Research, 2015, Vol.10, 84-102

Table 1: List of Health Registers in Denmark and the Responsible Institution.

Health Registers	Responsible authority
The Danish Pathology Register - 51 clinical registers	SSI SSI & SIF
Cause of death register	SSI
National Patient Register - Birth and fertility register - Psychiatric Register	SSI SSI SSI
The Children's Database	SSI
The Conscription Register	SSI
The National Health Insurance Register	SSI
Rehabilitation Register	SSI
Health Service Provider Register	SSI
Central Business Register (CVR)	SSI
Register of Medicinal Product Statistics	SSI
Register on Drug Abusers in Treatment	SSI
The National Health Profile	SIF

All of the national registers maintained at SSI and SIF contain a personal identification number (CPR number) such that all registers can be combined (Pedersen, 2011). The CPR number also provides an opportunity to geocode the data at the address level to support further spatial analysis (e.g., Storgaard et al., 2013). Data from the National Health Profile are based on addresses and are aggregated at the municipality level; the data are freely available from a web portal. The National Health Profile data can be geo-referenced through a polygon feature dataset (e.g., administrative boundaries) from the Danish spatial geodata portal, as presented in Figure 2.

Figure 2: The Process of Geocoding Health Data via the CPR Register and then Spatially Aggregating the Health Data into Administrative Units.



5. IDENTIFICATION OF HEALTH DATA IN DENMARK

The identification of relevant data for the INSPIRE geo-portal is based on the definition in the INSPIRE Directive theme of human health and safety. To identify possible health data, the definition is divided into six groups, and a detailed description follows

5.1. Spatial Distribution of Dominance of Pathologies

In Denmark, 51 clinical databases contain information on the prevalence, incidence and mortality of diseases (Green, 2011). There have been studies on the validity and coverage of these databases (Abildstrøm and Madsen, 2011; Bjerregaard and Larsen, 2011; Gjerstorff, 2011; Green, 2011; Sortsø, 2011; Thygesen et al, 2011). In general, the clinical database coverage is large (greater than 95%). The clinical registers contain data that originate from The Danish Pathology Register

International Journal of Spatial Data Infrastructures Research, 2015, Vol.10, 84-102

(Bjerregaard and Larsen, 2011). Additionally, the Cause of Death Register contains information on all deaths in Denmark. The Cause of Death Register is available in aggregated form through the INSPIRE Geoportal. In addition to the clinical registers, The National Institute of Public Health conducts a national health profile on 5% of the population every three years to determine the rate of sickness and specific diseases (Christensen et al, 2013).

5.2. Health Indicators

There are several registers in Denmark that provide data on the health conditions of individuals or population samples. The National Children's Database has information on children's exposure to smoking, duration of breastfeeding, and heights and weights in the first and 10th years in school. Second, the Conscription Register contains information on the weight and height of (mostly) men at their eligibility test for conscription at approximately the age of 18. The National Health Profile contains information on the percentage of the population that is over- or under-weight. In addition, the population sample contains information on the absence of sickness.

Birth and fertility registers, which include all births in Denmark, are kept at SSI. SSI also maintains the Psychiatric Register, with information on admission, discharge, treatment and diagnosis in psychiatric departments (Mors et al, 2011). The Birth Register and the Psychiatric Register are part of the Danish National Patient Register, which contains information on treatment at hospitals (Lynge et al, 2011).

Data on the use of health services is available from The National Health Insurance Register and the Rehabilitation Register, providing information on the population's use of health insurance services, i.e., general practitioners, physiotherapists and dentists (Andersen et al, 2011). Statistics Denmark also has information on hospital activity, hospital occupancy rate, home health care and health service use. In the National Health Profile, there is information on the proportion of the population sample that has visited their general practitioner within the last twelve months. Addresses of registered health service providers are included in the Health Service Provider Register and CVR.

The Register of Medicinal Product Statistics (MEDSTAT) contains information on the medicine user, the prescribers of the medicine and the pharmacies distributing the medicine (Kildemoes et al, 2011). The medicine user is identified through the CPR number, and the prescriber and pharmacy are identified through a code. This information can used to geocode the user, prescribers and pharmacies.

5.3. Indicators of Well-being

The National Health Profile contains information on stress, self-estimated health, physical/mental health, fatigue, headache, pain, discomfort, sleep disturbance, depression and social interactions, which all are indicators of a person's well-being.

5.4. Quality of the Environment Directly Influencing Health

Air and ground pollution, water quality, noise, the UV index and pollen are phenomena that influence health. These data are part of the INSPIRE theme of human health and safety, atmospheric conditions, environmental monitoring facilities and natural risk zones. Data on the UV index and air pollution are a part of the INSPIRE geoportal. Noise, air and ground pollution data are also available as spatial data from the Danish Natural Environment portal. The UV index and pollen data are available at the Danish Meteorological Institute. Distributions of industries can be extracted from CVR.

5.5. Quality of the Environment Indirectly Influencing Health

The National Health Profile has information on a population's habits related to alcohol, smoking, diet, physical activity and sedentary behaviour. SSI has a register on the drug abusers in treatment. Distribution of food retailers, fitness centres and sport facilities are extractable from CVR. The locations of green areas are available as spatial data from two web services: Danish Natural Environment Portal and PlansystemDK; the data are already available in the INSPIRE portal.

5.6. Events of Injury and Death

Information on traffic accidents and number of crimes aggregated at the municipality level are available from Statistics Denmark. The National Board of Industrial Injuries and Danish Working Environment Authority keeps a register of work injuries aggregated at the municipality level. The Danish Safety Technology Authority and Danish Emergency Management Agency have information on injuries and deaths caused by electricity, gas, fire and fireworks at the municipality level.

5.7. INSPIRE Compliance of Health Data

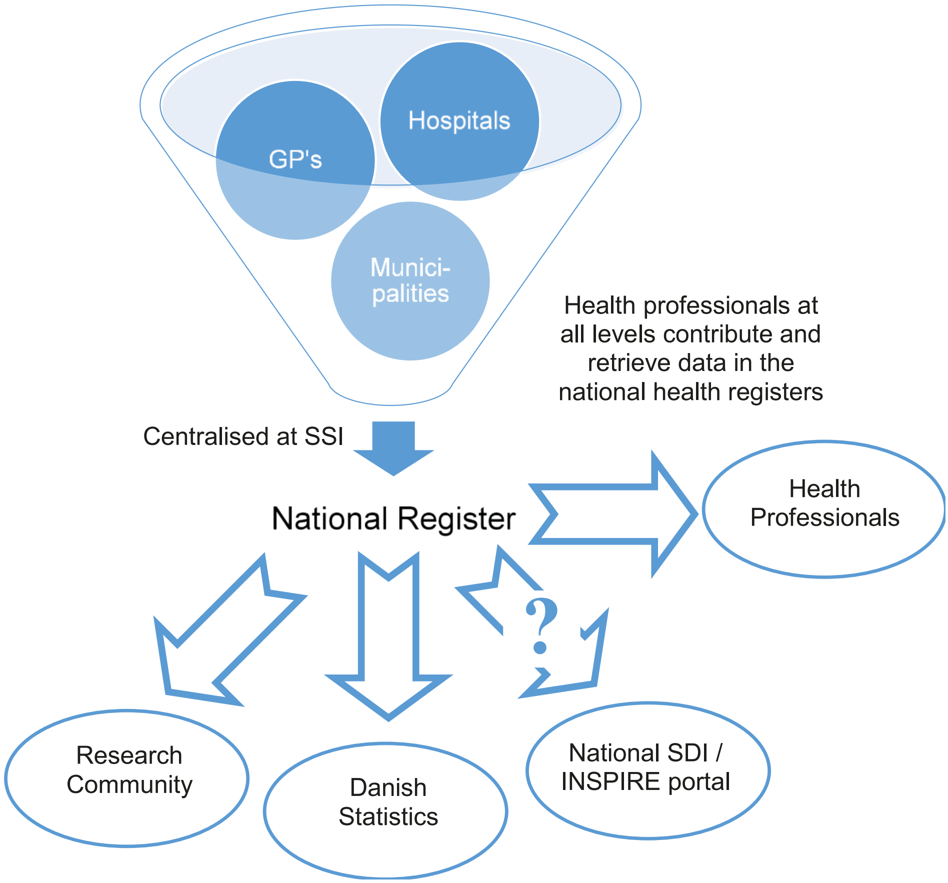
Duplicate data registrations, e.g., clinical registers and the Danish Pathology Register and the Birth Register or Psychiatric Register and the National Patient Register, contradict the INSPIRE principle of collecting and storing data only once. Similarly, for injuries and accident data, this information is collected at more than one location. The clinical registers are consistent with the requirements for the Danish SDI because the data are stored electronically, are nationwide and nationally applicable, and are statutory. Data from the National Health Profile represent a sample of the nationwide population, although the data collection is not statutory. The Conscription Register primarily contains information on the male population, i.e., it is often biased. Similarly, the statutory requirement for providing data for the Children's Database is rather new, which results in discrepant numbers of records for different municipalities.

The registers that contain the CPR number are subject to severe privacy and confidentiality issues, legally and technically. Data must be anonymised before it is compliant with INSPIRE or any other public data portal. Aggregation of the data into administrative units or the national/European grid systems (as visualised in Figure 2) is a viable option for making the data anonymous. Administrative units or grid systems make the data easily transparent and combinable with other spatial data.

Metadata is a keystone in the INSPIRE Directive to ensure easy access to information. Currently, there is no easy access to metadata on health data in Denmark. Metadata should follow the standards of INSPIRE so it can be included in the INSPIRE Geoportal and the Danish counterpart Geodata-info.

Centralisation of health data at primarily one authority (SSI) is the first step towards improving the organisation of health data in Denmark. SSI plays an important role in the ongoing digitisation process of the Danish health care system. SSI is responsible for the organisation and control of health data. Thus, SSI has initiated the development of IT systems to support efficient and standardised reporting of health data throughout the health care system. The concept is that health data are registered and added to the registers through real-time updates. This ensures that the health data are available for everybody in the health care sector. The flow of health data is presented in Figure 3. Health professionals have dual roles in the system. Professionals first register data on patients and then access information for treatment. The SSI role is to develop and maintain the data and IT systems and to promote the use of health data for research. Data from some health registers is applied at Danish Statistics, where the data are available in aggregated form. The aggregation is by region, municipality, parish or the national grid system; thus, some health data already have a spatial reference. Previously, the administration of health data was split between multiple units with different procedures. Organising the data at SSI has initiated the development of common standards for health data. This organisation will be an improvement that potentially enhances the use of health data.

Figure 3: The Process of Geocoding Health Data via the CPR Register and Spatially Aggregating the Health Data into Administrative Units.



6. DISCUSSION

Denmark is dedicated to digitising health care systems. As a result, many advanced national registers with person-identifiable health data on the entire population are available. The registers are high quality and accurate, with broad coverage. An efficient public healthcare system is now possible. In addition, researchers are able to perform high-quality analysis on the entire population or selected cohorts and patient groups without time-consuming and expensive data collection. Implementation of health data in the Danish SDI would be a major step forward in promoting spatial health research because it allows for effective access and use. Moreover, the data could help improve the quality and efficiency of the healthcare system. Digitising the healthcare system is the foundation for good data management and sharing, but a few Danish national registers are still updated manually, leading to multiple data registrations (Danish Government et al, 2013). This duplicity opposes the basic principles of INSPIRE and the Danish digitisation strategy. This issue must be addressed, as it is too expensive to perform the same task twice. Maintaining several versions of the same dataset raises concerns of data reliability and authority.

The national registers in Denmark all hold information, such as CPR numbers and addresses, for simply linkage of registers and spatial datasets through the use of geographical information systems. Therefore, environmental data and health data can be combined in analysing the causes and outcomes of diseases. However, health data and environmental data have very different requirements and restrictions with regard to spatial and privacy relevance. Health data are subject to heavy legal and ethical restrictions. Modifying data to an aggregated form is essential to bypass these restrictions. However, some degree of detail is inevitably lost in the process. Conversely, legal or sensitivity issues do not restrict environmental data, even though the data refer to a person's activities, e.g., pollution. Health data are based on individuals who are spatially dynamic. Aggregated health data are a momentary representation of the spatial distribution of incidence and prevalence rates. Environmental data are more stationary as it often only changes in the long-term. Environmental data are not a generalisation, in contrast to the aggregated health data. However, environmental data, such as UV and pollen, have continuous extension and fluctuation over a yearly cycle. This type of data is often based on the interpolation between discrete features and average values over several years. People frequently move; thus, a higher update frequency of health data in the SDI is needed. Environmental data, however, changes slowly over long periods and often does not significantly affect the spatial extent over a year.

The rapid development of technology and spatial software over the last decade has important implications for health applications. However, without proper structuring of health data, the preparation of health data for spatial analysis will continue to be tedious and time-consuming work. However, the harmonisation and implementation of health data in compliance with INSPIRE requires a huge amount of work. To overcome this hurdle, an overall policy is required for the implementation, creation of metadata, linkage to spatial data and aggregation of sensitive and confident data.

The current strategy for digitisation of the Danish healthcare system emphasises basic registration through the development of common systems for the many agencies involved and through centralisation of data agencies. As a result, more experts are involved. With the few entry points to access data, the required work for researchers is minimised significantly. In the digitisation strategy for the healthcare system, there are no initiatives for spatial health data, which is a problem for further implementation. Policies are essential for setting binding agreements; without strict requirements, there is little impetus to develop spatial health data.

Metadata is a keystone for an efficient SDI to support good data management and exchange information. Metadata creation and publication is perceived as tedious and time-consuming; however, without metadata, there is a risk that datasets become redundant. Documentation of Danish health care data is unstructured and is probably mainly a result of the previous administration. Restructuring of the metadata for health is required for consistency and organisation. There are several studies on the validation of registers with health data in Denmark (Abildstrøm and Madsen, 2011; Bjerregaard and Larsen, 2011; Gjerstorff, 2011; Green, 2011; Lyseen and Hansen, 2014; Sortsø, 2011; Thygesen et al, 2011), which could be applied for the creation of metadata information for health datasets.

Health data on the individual level holds detailed information that is subject to concerns over patient confidentiality and data security. There are numerous examples of data security breaches in Denmark in which data containing person-identifiable CPR numbers were hacked or stolen. Strong trust is put into the CPR number in Denmark; thus, the CPR number is powerful. Health registers have sensitive information, and most people would not like their personal medical history to be freely available. All parties involved with health data must ensure the confidentiality of individuals and protect the data both legally and technically. Patient-identifiable data are critical to medical research: updating, linking and validating data are impossible without person identifiable data, and the implementation of potential confounders in the analysis is difficult (Haynes et al, 2007). There is always a trade-off between the requirements of researchers and privacy concerns, despite the indisputable value of patient-identifiable data. Danish legislation limits the access to individual health information to employees within the sector with relevant needs and to researchers who were granted permission through legal channels. Person-identifiable data cannot be part of INSPIRE, but by making the data anonymous, the data can legally be a part of the national geoportal. Aggregating health data into administrative units or grid systems is a possibility for making the data anonymous. However, the spatial units must contain enough observations such that individual persons or families are not recognisable. Deciding between administrative units or a particular grid size for aggregation is difficult due to the various needs at multiple spatial and temporal scales. However, grid systems have the advantage of remaining the same over time and avoiding the modifiable area unit problem (MAUP) (Openshaw, 1983).

Denmark has progressed in the implementation of INSPIRE annex 1 and 2 data. The next couple of years are devoted to the harmonisation of the many thematic datasets in annex 3, including health data. The continuation of work with applications and the integration of the spatial components of the digital public administration of health data support the possibility of efficiency and increased use in research, public administration and the private sector.

7. CONCLUSION

Denmark has a vast amount of health data in digital databases, but there are several issues that must be addressed before the data are compliant with the INSPIRE principles. The issues relate to duplicate registrations of the same data, and the lack of access to structured metadata. Further efforts to anonymise data are required to ensure the privacy and confidentiality of health data. Spatial aggregation is the key to making the data anonymous, whereas CPR ensures the linkage between the health data and the addresses that serve as the spatial reference. The lack of initiative to include spatial applications of health data on the agenda for the digitisation of health data hinders the implementation of health data in INSPIRE. The issues of spatially aggregated health data and compliancy must be addressed. The harmonisation and implementation of health data in accordance with the INSPIRE Directive are unstructured and slowed by the lack of strategic promotion of spatial health data.

REFERENCES

Abildstrøm S. Z. and M. Madsen (2011). The Danish Heart Register, Scandinavian Journal of Public Health 39(7): 46–49.

Andersen J. S., N. De Fine Olivarius and A. Krasnik (2011). The Danish National Health Service Register, Scandinavian Journal of Public Health 39(7): 34–37.

Barrett F. A. (2000). Finke's 1792 Map of Human Diseases: The First World Disease Map?, Social Science and Medicine 50(7–8): 915–21.

Bjerregaard B. and O. B. Larsen (2011). The Danish Pathology Register, Scandinavian Journal of Public Health 39(7): 72–74.

Boulos M. N. K. (2005). Research Protocol: EB-GIS4HEALTH UK - Foundation Evidence Base and Ontology-Based Framework of Modular, Reusable Models for UK/NHS Health and Healthcare GIS Applications, International Journal of Health Geographics 4(1): 2.

Boulos M. N. K., A. J. Curtis and P. Abdelmalik (2009). Musings on Privacy Issues in Health Research Involving Disaggregate Geographic Data about Individuals, International Journal of Health Geographics 8: 46.

Christensen A. I., M. Davidsen, O. Ekholm, P. V. Pedersen and K. Juel (2013). Danskernes Sundhed – Den Nationale Sundhedsprofil 2013, Sundhedsstyrelsens publikationer, Rosendahls Distribution.

Danish Agency for Digitisation (2011). The Digital Path to Future Welfare - eGovernment Strategy 2011-2015. Rosendahls-Schultz Distribution.

Danish Agency for Digitisation (2012). Fact Sheet - Basic Data, at http://www.digst.dk/Servicemenu/English/Digitisation/~/media/Files/English/Fact_sheet_BasicData_pdf.pdf [Accessed 8 June 2014].

Danish Geodata Agency (2014). Geodata-Info, at http://www.geodata-info.dk/Portal/List.aspx [Accessed 8 June 2014].

Danish Government, Local Government Denmark, and Danish Regions (2013). Making eHealth Work, National Strategy for Digitalisation of the Danish Healthcare Sector 2013-2017, at http://www.ssi.dk/English/HealthdataandICT/~/media/Indhold/DK%20-%20dansk/Sundhedsdata%20og%20it/NationalSundhedsIt/Om%20NSI/Strategy2013-17.ashx [Accessed 8 June 2014].

Danish Ministry of the Environment (2008). Infrastructure for Spatial Information Act, at https://www.retsinformation.dk/Forms/r0710.aspx?id=122571 [Accessed 8 June 2014].

EDINA. GoGeo, The Place to Discover Geospatial Information and Services for those in Education and Research at http://www.gogeo.ac.uk/gogeo/index.htm [Accessed 8 June 2014].

EPHT. New Mexico Environmental Public Health Tracking Program at https://nmtracking.org/nm_epht/ [Accessed 1. February 2014]

European Commission. About INSPIRE. INSPIRE - Infrastructure for Spatial Information in the European Community at http://inspire.ec.europa.eu/index.cfm/pageid/48 [Accessed 8 June 2014].

Gjerstorff, M. L. (2011). The Danish Cancer Registry, Scandinavian Journal of Public Health 39(7): 42–45.

Green A. (2011). Danish Clinical Databases: An Overview, Scandinavian Journal of Public Health 39(7): 68–71.

Hansen, H. S. (2001). A quasi-four dimensional database for the built environment. In: Westort, C.Y. (ed.) DEM 2001. *Lecture Notes in Computer Science* 2181: 48–59. Springer, Heidelberg.

Hansen, H. S., L. Schrøder, L. Hvingel and J. S. Christiansen (2011). Towards Spatially Enabled e-Governance – A Case Study on SDI Implementation. International Journal of Spatial Data Infrastructure Research 6:73-96.

Hansen, H. S., L. Hvingel & L. Schrøder (2013). Open Government Data – a key element in the digital society. Lecture Notes in Computer Science 8061: 167 – 180.

Haynes C. L., G. A. Cook and M. A. Jones. Legal and Ethical Considerations in Processing Patient-Identifiable Data without Patient Consent: Lessons Learnt from Developing a Disease Register, Journal of Medical Ethics 33(5): 302–7.

HIVMapper (2014). Stat Compiler at http://www.hivmapper.com/ [Accessed 8 June 2014].

HIV Spatial Data Repository. Spatial Data Repository at http://www.hivspatialdata.net/ [Accessed 8 June 2014].

INSPIRE Thematic Working Group Human Health and Safety (2011). Data Specification on Human Health and Safety - Draft Guidelines, at http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_HH_v2.0.pdf [Accessed 8 June 2014].

INSPIRE Thematic Working Group Population Distribution – Demography (2012). Data Specification on Population Distribution - Demography – Draft Guidelines, at http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_PD_v3.0rc2.pdf [Accessed 8 June 2014].

Jensen V. M. and A. W. Rasmussen. Danish Education Registers, Scandinavian Journal of Public Health 39(7): 91–94.

Kildemoes H. W., H. T. Sørensen and J. Hallas. The Danish National Prescription Registry, Scandinavian Journal of Public Health 39(7): 38–41.

Lynge E., J. L. Sandegaard and M. Rebolj (2011). The Danish National Patient Register. Scandinavian Journal of Public Health 39(7): 30–33.

Lyseen, A. K. and H. S. Hansen (2014). Spatial and Semantic Validation of Secondary Food Source Data. ISPRS International Journal of Geo-Information 3(1): 236–53.

Mathys T. and M. N. K. Boulos (2011). Geospatial Resources for Supporting Data Standards, Guidance and Best Practice in Health Informatics, BMC Research Notes 4(1): 19.

Mors O., G. P. Perto and P. B. Mortensen. The Danish Psychiatric Central Research Register, Scandinavian Journal of Public Health 39(7): 54–57.

Murdoch T. B. and A. S. Detsky (2013). The Inevitable Application of Big Data to Health Care, The Journal of the American Medical Association 309(13): 1351–52.

Openshaw, S. (1983). The Modifiable Areal Unit Problem, Norwich: Geo Books.

Pedersen C. B. (2011). The Danish Civil Registration System, Scandinavian Journal of Public Health 39(7): 22–25.

Petersson F., M. Baadsgaard and L. C. Thygesen (2011). Danish Registers on Personal Labour Market Affiliation, Scandinavian Journal of Public Health 39(7): 95–98.

SAHSU. Small Area Health Statistics Unit. The environment and health atlas for England and Wales at http://www.envhealthatlas.co.uk/ [Accessed 1. February 2014]

Sortsø C., L. C. Thygesen and H. Brønnum-Hansen (2011). Database on Danish Population-Based Registers for Public Health and Welfare Research, Scandinavian Journal of Public Health 39(7): 17–19.

Snow J (1855). On the mode of communication of cholera. London: John Churchill.

Storgaard, R. L., H. S. Hansen, M. Aadahl and C. Glumer. (2013). Association between neighbourhood green space and sedentary leisure time in a Danish population. Scandinavian Journal of Public Health 41: 846-852.

Thompson J., S. Eagleson, P. Ghadirian and A. Rajabifard (2009). SDI for Collaborative Health Services Planning, In Global Spatial Data Infrastructures World Conference, Rotterdam, The Netherlands. At http://gsditest.opengeospatial.org/gsdiconf/gsdi11/papers/pdf/114.pdf [Accessed 8 June 2014].

Thygesen L. C., C. Daasnes, I. Thaulow and H. Brønnum-Hansen. Introduction to Danish (nationwide) Registers on Health and Social Issues: Structure, Access, Legislation, and Archiving, Scandinavian Journal of Public Health 39(7): 12–16.

World Health Organization (1). GeoNetwork - The Portal to Spatial Data and Information, at http://apps.who.int/geonetwork/srv/en/main.home [Accessed 8 June 2014].

World Health Organization (2). International Classification of Diseases (ICD), at http://www.who.int/classifications/icd/en/ [Accessed 8 June 2014].

Worldmapper. Worldmapper, at http://www.worldmapper.org/ [Accessed 8 June 2014].

APPENDIX II – PAPER II: SPATIAL AND SEMANTIC VALIDATION OF SECONDARY FOOD SOURCE DATA

ISPRS Int. J. Geo-Inf. **2014**, *3*, 236-253; doi:10.3390/ijgi3010236

OPEN ACCESS

ISPRS International Journal of
Geo-Information
ISSN 2220-9964
www.mdpi.com/journal/ijgi/

Article

Spatial and Semantic Validation of Secondary Food Source Data

Anders K. Lyseen * and Henning Sten Hansen

Department of Development and Planning, Aalborg University, A. C. Meyers Vænge 15, København SV 2450, Denmark; E-Mail: hsh@plan.aau.dk

* Author to whom correspondence should be addressed; E-Mail: alyseen@plan.aau.dk; Tel.: +45-411-780-28.

Received: 28 November 2013; in revised form: 6 February 2014 / Accepted: 12 February 2014 / Published: 28 February 2014

Abstract: Governmental and commercial lists of food retailers are often used to measure food environments and foodscapes for health and nutritional research. Information about the validity of such secondary food source data is relevant to understanding the potential and limitations of its application. This study assesses the validity of two government lists of food retailer locations and types by comparing them to direct field observations, including an assessment of whether pre-classification of the directories can reduce the need for field observation. Lists of food retailers were obtained from the Central Business Register (CVR) and the Smiley directory. For each directory, the positive prediction value (PPV) and sensitivity were calculated as measures of completeness and thematic accuracy, respectively. Standard deviation was calculated as a measure of geographic accuracy. The effect of the pre-classification was measured through the calculation of PPV, sensitivity and negative prediction value (NPV). The application of either CVR or Smiley as a measure of the food environment would result in a misrepresentation. The pre-classification based on the food retailer names was found to be a valid method for identifying approximately 80% of the food retailers and limiting the need for field observation.

Keywords: spatial; semantic; public health nutrition; food environments; geographical information systems; measurement

1. Introduction

Personal factors, such as taste preferences, nutritional knowledge, cooking culture, sensitivity to price and accessibility to food outlets, interact with the environment to influence food behavior. The food environment includes places where food can be acquired, such as supermarkets, bakeries and restaurants [1]. This physical food environment influences the types and amounts of food available and the opportunity for choosing a healthful diet [2,3]. Insights into food environments and nutritional behavior can facilitate human wellbeing and improve nutritional benefits [4]. Local food environments have proven to be an indicator of individual food behavior [1,5].

Reliable and valid measures of food environments are needed to fully understand the relationship between these environments and health [6]. Secondary food source data, including both governmental and commercial lists, are used repeatedly to measure food environments and foodscapes within health and nutritional research [4,5,7–11]. Knowledge of the validity of such secondary food source data is needed to fully understand the potential and limitations of the application of such datasets. Hence, the analysis, results and conclusions based on secondary data sources are influenced by four types of data integrity: completeness, thematic accuracy, geographical accuracy and contemporaneity. For food retailer lists, completeness refers to the percentage of the listed retailers that are actually present and whether there are missing retailers on the lists. Thematic accuracy is an expression of correctness in the classification of the food retailers. Geographic accuracy is the difference between the listed position (geocoded addresses or coordinates) and the actual position. Contemporaneity informs about the retention of outdated information. Unknown errors in the data lead to misinterpretations of the results or under- or over-estimation [12,13] of, for example, the density of food retailers or an analysis of the association of foodscapes with health or socioeconomic factors.

Previous examinations of the validity of food retailer lists have demonstrated limitations compared to direct observations, due to the lack of completeness, thematic and geographical accuracy and contemporaneity of such lists in the United States of America [13,14] and the United Kingdom [12]. However, studies have demonstrated contradicting results between the use of commercial and government lists. A study from the United Kingdom demonstrated high sensitivity between direct observations and council data, but only moderate sensitivity of commercial data sources [15]. On the contrary, a Danish study demonstrated a high positive prediction value (PPV) between commercial lists and field observations and only a moderate PPV for the government list [16]. The alternative to secondary food source data is direct observations, which are very time consuming and expensive to complete for large and/or densely populated areas. The combination of more than one source of secondary food data has been shown to improve the validity of data on individual food retailers based on the number of lists a retailer appears on [10].

Few studies [16–18] have been conducted on the validity of secondary food source data in Denmark despite the strong tradition of using register data. The studies have been limited geographically to the capital area of Copenhagen and thematically to supermarkets and fast food outlets, which made room for further development of methods for measuring the food environment [16,17].

The aim of this study is to examine the possibility of combining two government food retailer directories to achieve a higher validity though a proposed method for classifying food retailers based on a combination of retailer name and the standard classification in the directories. The purpose of the classification is to focus the time used for field observations of the retailers on the lists that may be wrongly classified or for which there is doubt about the coherence between the retailer name and classification. Previous studies have successfully applied a search for the identification of fast food outlets by combining the relevant NACE classification (the statistical classification of economic activities in the European Community) and retailer name [18]. This study expands this approach to include all retailers primarily targeted at selling food in public. Field observation is applied to evaluate the validity of the CVR and Smiley directories and also the proposed method for focusing field observations in future studies.

This paper will present the two secondary food sources examined and the method proposed to limit the time used on field observation. Furthermore, the method used for the field observations is explained. The PPV and sensitivity results are presented to evaluate the proposed method and the validity of the studied secondary food sources.

2. Methods

Forty-nine parishes in Northern Jutland were selected for the study, including both urban and rural areas. Aalborg is the largest city in the area, with a population of approximately 100,000, whereas the remaining areas consist of small villages with populations up to 7,000 and low-density housing. The study area is approximately 974 km^2, of which the city of Aalborg with the high-density housing constitutes 75 km^2 (8%). Approximately 15% of the population in the study area has an ethnicity other than Danish, and the levels of education and income are diverse across both the low- and high-density housing. Northern Jutland consists of eleven municipalities, of which five are defined as peripheral regions. Peripheral regions are defined by, among others, a lower average income than the national average, a lower amount of commuting traffic and low or negative population growth. In contrast to the peripheral regions, Aalborg attracts many young people and is the economic center of the region.

Food premises in the study area were identified using two freely available government directories (CVR and Smiley). In both directories, branch codes were used to define food premises. The branch codes are based on the European NACE classification [19]. The Smiley and CVR data were retrieved in June 2013.

2.1. Central Business Register (CVR)

The CVR is a government register that contains information about businesses in Denmark. Information about the legal unit in the companies is uniquely identified through the CVR number, and within each legal unit, production units are identified through unique P-numbers. The P-number is used for a complete list of food retailers, because each individual retailer in a chain has its own P-number. The CVR is updated once each day, 5 days a week, year-round. The database is administered and managed by the Danish Business Authority. The business owners provide the information, and it is their responsibility by law to keep the information up to date and correct. That the information about the branch and address are kept up to date through third party reporting implies that information consistency, accuracy and completeness could be doubtful. The CVR contains no information about the availability of foods, such as fresh meat or vegetables, in food selling premises or about the furnishing, business hours or payment options of food serving premises. Consequently, the NACE classification and business names are the only sources for identifying different food premises. The 15 branches listed in Table 1 were identified in the CVR as food selling or serving premises by definition [20].

Table 1. List of NACE codes applied to limit the search to food retailers in Smiley and Central Business Register (CVR).

Classification	NACE Code Used in CVR	NACE Code Used in Smiley
Grocery shops and kiosks	47.11.10	47.11.00.A
Supermarket	47.11.20	
Discount supermarket	47.11.30	
Other non-specialized shops	47.19.00	
Greengrocer	47.21.00	47.21.00.A 47.21.00.B
Butcher shops and delis	47.22.00	47.22.00.A 47.22.00.B
Fish shops	47.23.00	47.23.00
Retail with bread, confectionery and sugar products	47.24.00	47.24.00.A 47.24.00.B
Retail with beverages	47.25.00	47.25.00
Other food in specialized shops	47.29.00	47.29.00.C 47.29.00.D 47.29.00.E
Gas stations	47.30.00	-
Full service restaurants	56.10.10	56.10.00.A
Pizzeria, ice cream, *etc.*	56.10.20	56.10.00.B
Bars, cafés, *etc.*	56.30.00	56.30.00

2.2. Smiley Register

The Smiley register was introduced in 2001 and belongs under the Ministry of Food, Agriculture and Fisheries, who administers the food safety and hygiene regulations in Denmark. The register was created to register the food safety inspections of businesses and present the food safety level of each business to the public. Inspections are performed to ensure that shops and restaurants comply with the regulation. The inspection rates of the businesses are based on the health risk the branches constitute, ranging from twice a year to once every two years. Businesses with non-perishable goods are inspected as needed. Consequently, updates of the register are similar to the inspection rate, which suggests the retention of outdated data for up to two years. The register is updated every three months with the latest inspections. The lag time of three months between inspections and updates decreases the validity of the data, as it is less accurate and complete, as well as retaining outdated information. The relevant NACE classifications identified are listed in Table 1 along with the indication of aggregated and disaggregated groups in Smiley compared to the use of the NACE codes in the CVR. The NACE classification and the business names are the only indicators of type of food premise, as there is no information about merchandise, menu, business hours, table service or payment options [21].

2.3. Pre-Classification of Businesses

Pre-classification of the business records in Smiley and CVR was performed to examine the possibility of reducing or removing the field observation process, as this is a very time-consuming and expensive process. Previous literature has used a pre-classification based on a combination of business name and the NACE classifications to identify fast food restaurants [17,18]. Fast food restaurants were defined as within the NACE classification in question and with a restaurant name, including one of the following words associated with fast food: pizza, burger, sausages, barbeque (grill), kebab and falafel.

Table 2. Positive and negative words for each NACE code used to pre-classify the business records.

NACE Codes	Positive Words	Negative Words	Chain Names
47.11.10 Grocery shops and kiosks	Kiosk, convenience shop, grocery, food, marked, staple goods	Canteen, cafeteria, flowers	Spar, Brugsen, 7-Eleven, Twenty 4–7
47.11.20 Supermarket	Grocery, food, marked, staple goods	Canteen, cafeteria, flowers	Spar, Superbest, Dreisler, Brugsen
47.11.30 Discount supermarket	Convenience shop, grocery, food, marked, staple goods	Canteen, cafeteria, flowers	Rema, Fakta, Netto Kiwi, Irma
47.19.00 Other retail from non-specialized shops	Kiosk, convenience shop, grocery, food, marked, staple goods	Canteen, cafeteria, flowers	Kvickly, Bilka, Føtex, Salling
47.21.00 Greengrocer	Vegetables, green, fruit	Canteen, cafeteria, flowers	-
47.22.00 Butcher shops and delis	Slaughter, butcher, delis, delicatessen	-	-
47.23.00 Fish shops	Fish	-	-
47.24.00 Retail with bread, confectionery and sugar products	Bakery, candy, chocolate, confectionary, sweets	Sport, care home, canteen, cafeteria	Frellsen
47.25.00 Retail with beverages	Wine, beer	Canteen, cafeteria	-
47.29.00 Other retail with food in specialized shops	Cheese, nutrition, bazaar, egg, thee, coffee	Transportation, canteen, cafeteria	-
47.30.00 Gas stations	Retail, shop, 7-Eleven, service		Q8, Shell, Statoil, Haahr
56.10.10 Full service restaurants	Restaurant	Pizza, pub (bodega), rental, sport, invest, club, development, golf, kiosk, assembly room, management	-
56.10.20 Pizzeria, take away, ice cream shops, *etc.*	Sausage, hotdog, pizza, grill, sandwich, pita, barbeque, burger, shawarma, sushi, kebab, Thai, salad, pancakes, take away	Cultural center, bingo, cafeteria, sport, trader, canteen, ice cream, bar, invest, club, assembly room, pool, administration, office, hall	McDonalds, Burger King, Subway,
56.29.00 Other restaurants	-	Canteen, hall, catering, school, sport	-
56.30.00 Cafés, pub, bars, *etc.*	Bar, café, bodega, pub, nightclub, disco	Sport, club	

Pre-classifying the businesses has previously been proven to focus the search for fast food outlets in the Smiley register [18] and is applied here to all types of food retailers to evaluate the results for different food sources. The list of words used for classifying the businesses can be found in Table 2. The words are based on Danish food tradition and culture combined with empirical knowledge gathered in the fieldwork. Positive words indicate that a business is most likely selling or serving food based on the business name combined with the NACE classification. Negative words indicate that a business is not targeted at selling food, has very limited opening hours or is located in a restricted area. Positive words listed under a different NACE code than the one in question indicates that the business has been given the wrong NACE code. Any business name not associated with either a positive or a negative word is not classifiable. Based on the positive and negative words and NACE codes, the business records can be divided into four groups.

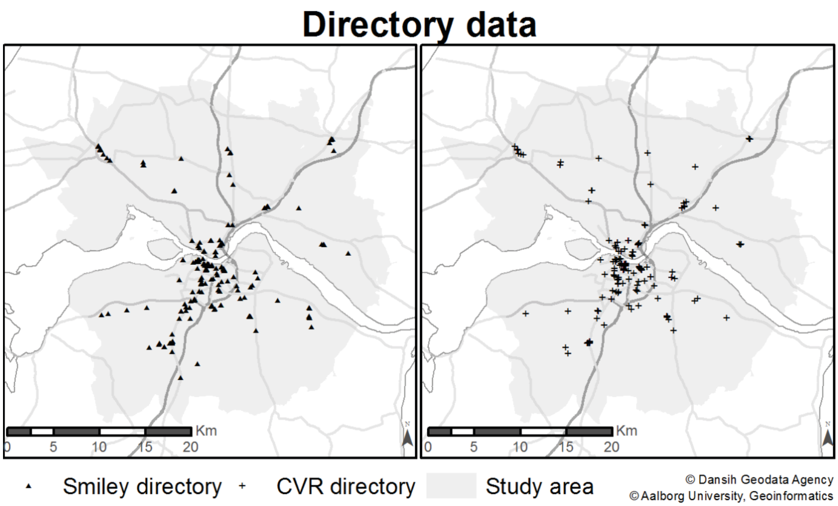
1. Most likely food businesses: the business name contains positive words associated with the NACE code.
2. Non-food targeted businesses: the business name contains negative words associated with the NACE code.
3. Wrongly classified businesses: the business name contains positive words associated with a different NACE code.
4. The business's relevance is not possible to categorize based on the name.

If the pre-classification proves successful, the application thereof to the registers in other parts of the country would reduce the field observation process to include only group four.

2.4. Geo-Coding

The addresses in CVR were geocoded based on address reference data in the Universal Transverse Mercator (UTM) projection obtained from the Danish Geodata Agency. The Smiley register contains WGS84 (World Geodetic System 84) coordinates for approximately 95% of entries, which are transformed to UTM and used as their locations. The remaining records are geocoded through the use of the address and reference data from the Danish Geodata Agency. The distribution of the Smiley and CVR directory entries is visualized in Figure 1.

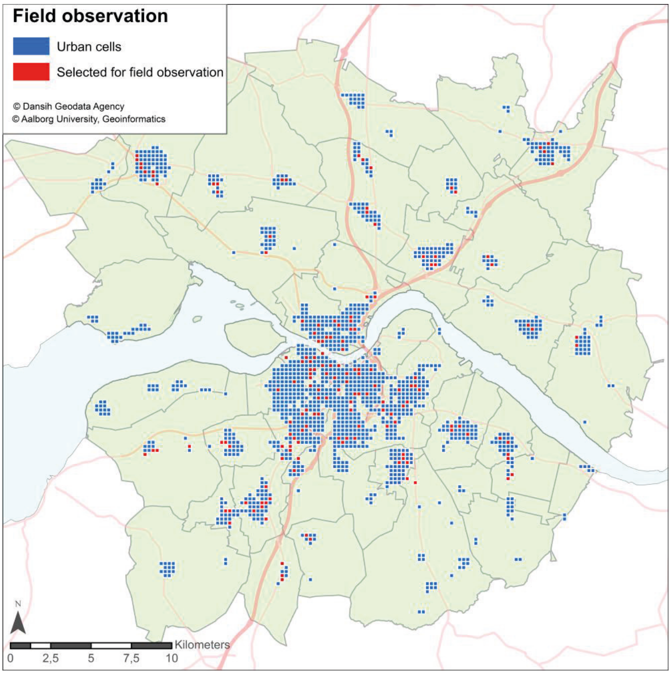
Figure 1. Map of the records in Smiley and CVR.



2.5. Field Observation

The method for field observation was adopted from Toft *et al.* [18]. The study area was divided into cells of 250 × 250 m through the use of the standard Danish Grid Cell system. Four hundred and ninety-seven grid cells contain records from the Smiley register and CVR. A random sample of 125 cells was selected from the 497 cells for field observation. An additional 35 cells were selected to search for unlisted food retailers in cells that, based on population, could possibly support the existence of a food retailer. To fulfill this, the 35 cells had to follow these three criteria: the cell contains no records in Smiley or CVR; a minimum of 10 addresses from the address reference data are located in the cell; and a minimum of two neighboring (queen's rule) cells have a minimum of 10 addresses. The selected and populated (following the criteria) cells are illustrated in Figure 2. The selected cells were approximately divided into 50% located in the metropolitan area of Aalborg and 50% located in the rural areas surrounding Aalborg.

Figure 2. Map of the 160 randomly selected grid cells located within the 60 parishes in the region around Aalborg.



Two surveyors performed the field observations in July 2013, visiting the 160 grid cells. Every street in the cells was systematically searched for food retailers listed in Smiley or CVR, as well as unlisted food retailers. A real-time kinematic global navigation satellite system (RTK GNSS) was used to measure every observed food retailer, and the characteristics of the retailer were identified to classify the food retailer by type. The characteristics of the businesses used to classify the food retailers were drawn from previous literature used for classifying food stores [22] and restaurants [16,23], but modified to fit Danish standards. The definitions of the food retailers are based on the businesses' characteristics as listed in Table 3. In the field observations, food retailers listed in CVR and Smiley were omitted from being measured if they belonged to one of the three following definitions: retailers not targeted at selling food, retailers located within a restricted area and nonexistent retailers.

Table 3. Characteristics used to classify food stores and restaurants.

Food Retailer Type	Characteristics
Supermarket	Supermarkets that are part of a large chain, a minimum of three cash registers, fresh meat, a large selection of fresh vegetables and fruit and often one or more of the following features: butcher, deli or bakery
Discount supermarket	Supermarkets that are part of a chain, a maximum of two cash registers, a small selection of fresh meat and vegetables and fruit
Grocery shops and kiosks Gas stations	Small independent convenience and grocery stores, kiosks and gas stations with a limited selection of food items
Specialty food stores: fish, greengrocers, butchers, delis, bakers, beverages, *etc.*	Specialized in the trade of one food (meat, vegetables, beverages, fish, *etc.*) with little or no other food types in store
Full service restaurants and cafés	Fine dining, sit down (eat-in) with service at tables
Pizzeria, take away, ice cream shops, *etc*. (fast food)	Fast food chains and independent retailers with two or more of the following features: expedited food service, counter service only, takeout business and payment tendered prior to receiving food
Bars, pubs, *etc*.	Limited food serving with a focus on serving alcohol and late-night opening hours

2.6. Statistical Analysis

Sensitivity and PPV were calculated to establish the level of agreement between the two food directories and the field observations. The results from the field observations were treated as the "gold standard". The calculation was performed using the 2 × 2 shown in Table 4.

Table 4. Illustration of the relationships between true and false field observations and food directories.

		Field Observation	
		Present	Absent
Food directories	Present	True positive (TP)	False positive (FP)
	Absent	False negative (FN)	True negative (TN)

Sensitivity is the proportion of food retailers observed through the field observations that were listed in the food directories. Sensitivity is a measure of the completeness of the food directories calculated using Equation (1).

$$Sensitivity = \frac{TP}{TP + FN} \tag{1}$$

PPV is the proportion of food retailers listed in the directories that were observed in the field observations and was calculated using Equation (2).

$$PPV = \frac{TP}{TP + FP} \tag{2}$$

Sensitivity and PPV were also calculated for the NACE classification, including both non-exact and exact classification matches between the NACE classification and the field observations. This presents a measure of the thematic accuracy of the government directories.

The pre-classification of the food retailers was evaluated through sensitivity, PPV and negative predictive value (NPV). NPV is the proportion of observations pre-classified as not targeted at selling food and observed in the field observation as not selling food. NPV was calculated using Equation (3).

$$NPV = \frac{TN}{TN + FN} \tag{3}$$

The categorization of sensitivity, PPV and NPV was as follows [24]: <0.30 (poor), 0.31–0.50 (fair), 0.51–0.70 (moderate), 0.71–0.90 (good) and >0.91 (excellent). The standard deviation (σ) between the food directory's location and the RTK GNSS measurements collected in the field observations was calculated as a measure of the geographical accuracy. This standard deviation was calculated using Equation (4) [25], where d_i is the Euclidean distance between a retailer's observed location and the location in the food directory and $\bar{d}$ is the mean value of all the distances, d_i.

$$\sigma = \sqrt{\frac{\sum\left(d_i - \bar{d}\right)^2}{n}} \tag{4}$$

The standard deviation is an indicator of the dispersion from the expected or "true" value. The observations measured by a real-time kinematic global navigation satellite system (RTK GNNS; advanced GPS) have an accuracy of 1–2 cm in the plane [26], and hence, the coordinates measured by the RTK GNSS receiver were considered the "true value".

3. Results

3.1. Completeness

In Table 5, the comparison between the retailers listed in Smiley and CVR and the field observations is summarized. From Smiley and CVR, 285 and 199 retailers, respectively, were selected for field observation. In the field observations, 272 retailers from the Smiley directory and 164 retailers from the CVR directory were present. Thirteen of the retailers listed in Smiley were not observed in the field. This was primarily because either the retailer was listed at the owner's address (n = 5) or the listing was for a mobile retailer (n = 4). The PPV calculated for the retailers listed in Smiley that were present in the field observations was excellent (0.95). Thirty-five of the retailers listed in CVR were not observed in the field. The majority were retailers listed at the address of the owner (n = 25) or mobile retailers (n = 2), similar to Smiley. The PPV for the retailers listed in CVR was good (0.82).

Table 5. Identification of retailers in CVR and Smiley in relation to the field observations.

		Field Observation	
		Present	**Absent**
Smiley	Present	272	13
	Absent	-	-
CVR	Present	164	35
	Absent	-	-

3.2. Thematic Accuracy

The retailers present in Smiley and CVR did not all fit the characteristics of one of the food retailer types in Table 3. Table 6 presents the comparison between the food retailers listed in Smiley and the food retailers found in the field observation. A total of 187 food retailers were observed in the field observations and also listed in Smiley, and 41 (21.93%) were observed that were unlisted in Smiley. One third of the retailers listed in Smiley were not located in the field observations (*n* = 98), including those omitted because they were not targeted at selling food (*n* = 15), or were located in a restricted area (*n* = 76). This primarily included canteens (*n* = 11), institutions for children and the elderly (*n* = 29) and sports venues (*n* = 34). The PPV calculated for the food retailers in Smiley that were present in the field observations was moderate (0.66), and the sensitivity for food retailers in the field observations listed in Smiley was good (0.82). The individually calculated sensitivities for each food retailer classification were good and ranged from 0.77–0.86. PPVs were also calculated for the individual classifications, but with a larger dispersion from fair to excellent (0.50–0.93).

Table 6. Comparison of the food retailers listed in Smiley with those found in the field observations for each classification of food retailers and the total number (* incorrectly classified retailers). PPV, positive prediction value.

		Supermarket		**Specialty Food Stores**		**Restaurants**		**Bars, Cafés, *etc.***		**Total**	
		Present	**Absent**	**Present**	**Absent**	**Present**	**Absent**	**Present**	**Absent**	**Present**	**Absent**
Field observation	Present	40 (1 *)	12	36 (2 *)	6	99 (25 *)	20	12 (0 *)	3	187 (28 *)	41
	Absent	3	-	8	-	75	-	12	-	98	-
Sensitivity		0.77		0.86		0.83		0.80		0.82	
PPV		0.93		0.82		0.57		0.50		0.66	

Of the 187 food retailers present in both the field observations and Smiley, 28 (14.97%) were incorrectly classified based on the characteristics from Table 3, though 17 of these were cafés listed as restaurants, which in terms of their characteristics are much more similar than bars and cafés according to the NACE classification. The remaining misclassified retailers were fast food retailers listed as supermarkets (*n* = 1) or specialty food stores (*n* = 2), bars listed as restaurants (*n* = 3) and kiosks listed as restaurants (*n* = 5).

In Table 7, the comparison between the food retailers listed in CVR and the food retailers found in the field observations is presented. One hundred and forty-three of the food retailers in CVR were found in the field observations and 55 were absent. Of those 55, nine were not located, 25 were located at the owner's home address and 14 were in restricted areas, such as canteens (*n* = 5) and sport venues

(n = 4). The PPV and sensitivity for the comparison of CVR and field observations were, respectively, good (PPV = 0.72) and moderate (sensitivity = 0.63). The sensitivity for the individual food retailer classifications ranged from fair to good (0.34–0.81). PPV ranged from moderate to excellent (0.54–0.91).

Table 7. Comparison of the food retailers listed in CVR with those found in the field observations for each classification of food retailers and the total number (* incorrectly classified retailers).

		Supermarket		Specialty Food Store		Restaurant		Fast Food		Bar, Cafés, *etc.*		Total	
		Present	Absent	Present	Absent	Present	Absent	Present	Absent	Present	Absent	Present	Absent
Field observation	Present	42 (2 *)	10	13	25	22 (15 *)	10	48 (2 *)	33	18 (1 *)	7	143 (20 *)	85
	Absent	4	-	11	-	11	-	20	-	10	-	56	-
Sensitivity		0.81		0.34		0.69		0.59		0.72		0.63	
PPV		0.91		0.54		0.67		0.71		0.64		0.72	

In the comparison of food retailer classifications between CVR and the field observations, 20 of the 143 retailers (13.99%) found in the field observations were incorrectly classified. These included fast food retailers listed as supermarkets (n = 2) or restaurants (n = 9), cafés listed as fast food (n = 4), bars listed as restaurants (n = 2) and restaurants listed as bars in CVR (n = 1).

In Table 8, rural and urban areas are compared based on the number of food retailers listed in CVR or Smiley and the field validation. The PPV for Smiley ranged from 0.62 in rural to 0.67 in urban areas and for CVR from 0.73 in rural to 0.71 in urban areas. The sensitivity for Smiley ranged from 0.88 in rural to 0.95 in urban areas and for CVR from 0.85 in rural to 0.93 in urban areas. Only small differences were found in both PPV and sensitivity between the rural and urban areas for both CVR and Smiley. However, there was a small tendency that retailers found during field observations in urban areas were a bit more likely to be present in Smiley and CVR.

Table 8. Comparison of food retailers divided into urban and rural areas.

		Urban Area				Rural Area			
		Smiley		CVR		Smiley		CVR	
		Present	Absent	Present	Absent	Present	Absent	Present	Absent
Field observation	Present	126	7	99	7	61	8	44	8
	Absent	61	-	40	-	37	-	16	-

A comparison of Smiley with CVR is presented in Table 9. In the field observations, 228 food retailers were identified, but only 117 (51.32%) of these were listed in both CVR and Smiley. Additionally, 15 observations from the field observations were not found in either CVR or Smiley. The probability of a food retailer found in the field observations being listed in either CVR or Smiley is excellent (sensitivity = 0.93).

Table 9. Comparison of the food retailers found in the field observations being listed in Smiley and CVR.

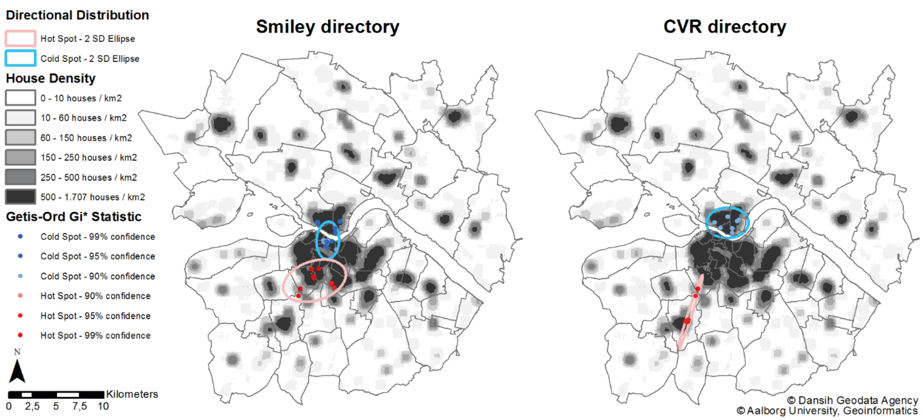
		Smiley	
		Present	**Absent**
CVR	Present	117	26
	Absent	70	15

3.3. Geographic Accuracy

The field observation coordinates collected with the RTK GNSS receiver and those from Smiley (few geocoded) and CVR (all geocoded) were compared based on joint Euclidian distance. The mean and standard deviation for Smiley and CVR are 23.74 ± 23.04 m and 18.74 ± 19.83 m, respectively. For Smiley, 97.33% of the records measured in the field were within 100 m of the listed coordinates and 87.70% were within 50 m. For CVR, all records measured in the field were within 100 m and 92.31% were within 50 m. For the 250 × 250 m cells, 12.30% of the records in Smiley and 12.59% of the records in CVR were found outside the cell in which the listing was registered. None of the records in either Smiley or CVR were found outside the parish in which the retailer was registered.

The errors between the locations in the registers and the measured locations were analyzed for spatial patterns through the measurement of spatial autocorrelation (Moran's I) and high/low clustering (Getis-Ord General G). The results of the analysis were high positive z-scores for both spatial autocorrelation (Smiley 15.74; CVR 15.96) and high/low clustering (Smiley 8.66; CVR 11.18), indicating clustered results. The p-value was, on all occasions, below 0.001, indicating significant results.

Figure 3. Map of hot/cold spot Getis-Ord Gi* statistical analysis of the Euclidean distances between "true" locations and the locations derived from the registers. Two standard deviational ellipses are visualized for the hot and cold spots.



The distribution of the clusters was analyzed to determine whether the clusters are located in urban or rural areas. The analysis was conducted in the software ArcGIS Desktop 10.2 by ESRI using optimized hot spot analysis (Getis-Ord Gi* Statistic) from the Spatial Statistics package. In Figure 3, the results are visualized. The clusters with low values (cold spots) are for both Smiley and CVR located in the central part of Aalborg, whereas the clusters with high values are located in the sub-urban/rural areas for Smiley and in rural areas for CVR.

3.4. Pre-Classification

The pre-classification divided the food retailers listed in CVR and Smiley into four groups based on the retailers' names. In CVR and Smiley, respectively, 109 and 124 retailers were classified as "most likely food business", 26 and 85 retailers as "non-food targeted business", 20 and 29 retailers as "wrongly classified business" and 44 and 47 as "business classification not possible". The field observations were compared to each group in the pre-classification, as shown in Table 10, and the proportion of correctly classified retailers in each group was calculated as PPV for three of the groups and as NPV for the group "non-food-targeted business". The PPVs for the classifications "most likely food business" (0.98) and "wrongly classified business" (0.97) were both excellent for Smiley, as was the NPV for the classification "non-food-targeted business" (0.98). The PPV for the classification "business classification not possible" in Smiley was good (0.74). Similarly excellent results were calculated for CVR when comparing the pre-classification and the field observations for the classes "most likely food business" (0.95), "wrongly classified business" (0.95) and "non-food-targeted business" (1.00), but only a fair PPV for the class "business classification not possible" (0.45). Based on the pre-classification, 47 retailers in Smiley and 44 retailers in CVR would be selected for field observation, thereby reducing the amount of field observation needed, with 83.51% for Smiley and 77.89% for CVR. The remaining retailers in Smiley (n = 238) and CVR (n = 155) have excellent PPVs of, respectively, 0.98 and 0.93 as a measure of being classified correctly. The combination of CVR and Smiley results in a total of 224 food retailers, including 11 errors, where only 23.15% were selected for field observation. Additionally, 15 retailers are missing, as they were not found in the field observations. This results in an excellent PPV (0.95) and sensitivity (0.93).

Table 10. Comparison of the pre-classification method, where the retailers are classified based on their name and the field observations.

		Most Likely Food Business		**Non-Food-Targeted Business**		**Wrongly Classified Business**		**Business Classification not Possible**	
		Present	**Absent**	**Present**	**Absent**	**Present**	**Absent**	**Present**	**Absent**
				Pre-classification Smiley					
Field observation	Present	122	-	-	2	28	-	35	-
	Absent	2	-	-	83	1	-	12	-
				Pre-classification CVR					
Field observation	Present	104	-	-	0	19	-	20	-
	Absent	5	-	-	26	1	-	24	-

4. Discussion

The identification of food retailers in the public space using individual lists from secondary sources has limited utility as a measure of the food environment. This is because the thematic accuracy for the directories are represented by a PPV of 66% for Smiley and 72% for CVR, indicating the proportion of food retailers listed in the directories that are actually a food retailer in reality; likewise for the sensitivity values of 82% for Smiley and 63% for CVR, indicating the proportion of food retailers found through the field observations that were listed in the directories. The results have similarities to previous studies of Smiley [18], where an identical sensitivity of 82% was achieved, though the PPV was a great deal higher at 92%. The higher PPV obtained was most likely the result of that study being limited to fast food retailers. Previous studies of the CVR directory [17] reached higher values for PPV (81% *vs.* 72%) and sensitivity (75% *vs.* 63%) compared to this study. Both studies included all food retailers and had the same sample size and applied field observations as the validation method. The only difference is in the geographical extents of the studies; while the previous study was limited to Copenhagen (high-density housing), this study included Aalborg, a city somewhat comparable to Copenhagen, but also included rural areas as approximately 50% of the areas for field observation.

The differences between urban and rural areas in the identification of food retailers are hard to establish if present. The difference found in this and in previous studies was a slightly higher sensitivity in urban areas. This includes the Smiley directory (93% *vs.* 85%), the CVR (95% *vs.* 88%) and a previous study of the Smiley directory (84% *vs.* 76%) [18]. The PPV is contradictory between CVR and Smiley in this study, as urban is highest in Smiley (67% *vs.* 62%) and rural highest in CVR (73% *vs.* 71%). The previous study of Smiley found the PPV to be highest in rural areas (94% *vs.* 90%), which contradicts the results found for Smiley in this study. Hence, there is no clear indication of better or worse PPV between urban and rural areas, with only a marginally better sensitivity for urban areas. These contradictions and small differences make no positive indications as to the possibility of significantly improving the accuracy of the directories.

Previous studies have stated that individual lists of food retailers have limited utility for identifying food stores, but combining the lists improves the likelihood of a retailer being a food store [27]. Combining CVR and Smiley produced the same results, as sensitivity increased to 93%, but still fell short of getting a high PPV. A combination of the two directories is not a method for reaching a valid list of food retailers without field observation or another method.

The geographic accuracy of the Smiley directory (23.74 ± 23.04 m) is comparable to previous studies (15 ± 24 m) [18]. The CVR is slightly better than Smiley with an accuracy of 18.74 ± 19.83 m. With 87.70% of the retailers in Smiley and 92.31% in CVR registered within 50 m of the true GPS position, the directories are accurate compared to other studies yielding results of 53%–56% within 100 m in the United States of America [13]. Whether the errors are larger in urban or rural areas is uncertain based on the analysis, though with a small tendency towards smaller errors being in the most populated areas.

The geographic accuracy clearly influences the applicability of the data. Analyses aggregating retailers over large areas or analyzing distances to the nearest food retailer are less affected by geographical inaccuracy, particularly if the food environment is dense with retailers. On the other hand, areas with few food retailers and analyses at small scales are vulnerable to geographic inaccuracy. In areas with a high density of food retailers, the distance in the analysis will theoretically have no impact, as the direction of the errors should be random. Whether this holds true is doubtful, but it calls for further research to fully understand the nature of the errors. The aggregation of retailers over small areas will create errors, as exemplified by the CVR directory. In CVR, 92.31% of the records were within 50 m, and according to the standard deviation, 95% should be within 58 m, but when aggregated into 250 × 250 m cells, more than 12% were aggregated incorrectly.

The completeness and thematic accuracy of the data demonstrates that if the raw data were used in research, there would exist a huge overrepresentation of food retailers similar to other studies [13]. The misclassification of retailers poses a major problem if analyzing small retailer groups, such as specialty stores, whereas the errors have less of an impact on large groups, such as restaurants or supermarkets. The completeness of both CVR and Smiley are poor in their raw state, as they are both missing retailers and have retailers that are in restricted areas, misclassified and nonexistent. We have not managed to identify the contemporaneity of the data, as there are several problems in measuring this completely. There are obvious problems with the retention of old data and the lack of new data in Smiley. The extent of these problems differs, as retailers closing down may only be visited once every second year, whereas retailers opening a shop need to enroll in the Smiley register within two weeks. This could indicate an overrepresentation of retailers in Smiley. The CVR directory has different issues, as this is updated on a daily basis, but requires input from the retail owners about address and classification. Based on the field work, the accuracy of the addresses is good, but the classifications include many errors, especially in regard to combined retailer classifications, *i.e.*, gas stations often have a small kiosk, but are only classified as a gas station.

The Danish government has made basic data freely available to all, by which action the data are usable by a much larger crowd. Hence, there are obvious applications for this information in research, but the data were not collected for the purpose of research and, therefore, have limitations in term of completeness and thematic accuracy. In the Smiley directory, all units serving food are listed, which include limited access retailers that are not relevant in a measure of the public food environment. Similarly, for CVR, many mobile stands are included as being located at the owners address, but during business hours are located at more central spots in the city. Consequently, knowledge about the data's accuracy, completeness, *etc.*, is essential when basing analysis and conclusions on such directories.

The pre-classification method based on business names was earlier proven to be a good method for improving PPV and sensitivity for the identification of fast food outlets in Copenhagen [18]. The results of applying the pre-classification in this study were excellent, with a greatly improved PPV and sensitivity of the directories. The method demands knowledge about the tradition and culture of the food retailers, as well as the language to determine which words the classification should be based on. In a Danish context, the study confirms the results of a previous study by Toft and colleagues for both CVR and Smiley. The pre-classification limits the time and cost of field observations, which is most needed, as fieldwork can be a very expensive affair if the area and the number of food retailers in question are large [6]. Based on a study including five secondary sources [17] and another combining nine secondary sources of food retailers [27], the inclusion of more sources is believed to improve the identification of food retailers in the directories and, hence, the measure of the food environment. The application of the pre-classification method followed by the use of additional food retailer directories to further limit the needed amount of field observation is considered to improve the measure of the food environment even more in terms of time and finances needed.

5. Conclusions

The completeness of the listings of retailers in Smiley and CVR were excellent and good, respectively, but a large proportion of the retailers (34% in Smiley and 28% in CVR) were not targeted to selling food in the public space or were limited to a confined area. This was the result for all of the NACE classifications, though most pronouncedly for restaurants (PPV = 0.57) and bars (PPV = 0.50) in Smiley and for specialty food shops in CVR (PPV = 0.54). Both CVR and Smiley were missing retailers, which were found in the field observations with sensitivities of, respectively, 0.63 and 0.82. As neither CVR nor Smiley has a combination of excellent PPV and sensitivity, the direct application of either directory would result in a misrepresentation of food retailers.

There were found to be no clear differences between food retailers in urban *vs*. rural areas, with differences of 0.02–0.08 for sensitivity and PPV.

Combining CVR and Smiley resulted in an excellent sensitivity (0.93), with only 15 retailers missing from both directories, but without field observation, the retailers not targeted at selling food in the public space cannot be removed from the directories, again leading to a misrepresentation of food retailers.

The pre-classification resulted in an excellent PPV and sensitivity, but is limited to the specific classification characteristics and application in CVR and Smiley. Adaption to other Danish and possibly Scandinavian directories is plausible with the current characteristics of the pre-classification, due to the similarity in languages, tradition and culture. Application of the pre-classification to other countries' directories is believed to be possible if the criteria for classifying the food retailers are modified to the culture and tradition of the country's language and food environment.

Acknowledgments

This paper was founded by an internal research grant from Aalborg University, Copenhagen, Denmark. The authors acknowledge Mette Lund Jensen for her technical assistance during the field observation.

Author Contributions

Both authors contributed to the conceptualization of the study. Anders Knørr Lyseen led the field observation, data analysis and writing of the article. Henning Sten Hansen reviewed and revised all drafts of the article.

Conflicts of Interest

The authors declares no conflict of interest.

References

1. McKinnon, R.A.; Reedy, J.; Morrissette, M.A.; Lytle, L.A.; Yaroch, A.L. Measures of the food environment: A compilation of the literature, 1990–2007. *Am. J. Prev. Med.* **2009**, *36*, 124–133.

2. Pearce, J.A.; Hiscock, R.A.; Blakely, T.B.; Witten, K.C. A national study of the association between neighbourhood access to fast-food outlets and the diet and weight of local residents. *Health Place* **2009**, *15*, 193–197.
3. Thornton, L.E.; Pearce, J.R.; Macdonald, L.; Lamb, K.E.; 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*, doi:10.1186/1476-072X-11-29.
4. Mikkelsen, B.E. Images of foodscapes: Introduction to foodscape studies and their application in the study of healthy eating out-of-home environments. *Perspect. Public Health* **2011**, *131*, 209–216.
5. Moore, L.V.; Diez Roux, A.V. Associations of neighborhood characteristics with the location and type of food stores. *Am. J. Public Health* **2006**, *96*, 325–331.
6. Kelly, B.; Flood, V.M.; Yeatman, H. Measuring local food environments: An overview of available methods and measures. *Health Place* **2011**, *17*, 1284–1293.
7. Neckerman, K.M.; Bader, M.D.M.; Richards, C.A.; Purcial, M.; Quinn, J.W.; Thomas, J.S.; Warbelow, C.; Weiss, C.C.; Lovasi, G.S.; Rundle, A. Disparities in the food environments of New York city public schools. *Am. J. Prev. Med.* **2010**, *39*, 195–202.
8. Sturm, R. Disparities in the food environment surrounding US middle and high schools. *Public Health* **2008**, *122*, 681–690.
9. Lytle, L.A. Measuring the food environment. *Am. J. Prev. Med.* **2009**, *36*, 134–144.
10. Glanz, K. Measuring food environments: A historical perspective. *Am. J. Prev. Med.* **2009**, *36*, 93–98.
11. Wang, M.C.; Kim, S.; Gonzalez, A.A.; MacLeod, K.E.; Winkleby, M.A. Socioeconomic and food-related physical characteristics of the neighbourhood environment are associated with body mass index. *J. Epidemiol. Commun. Health* **2007**, *61*, 491–498.
12. Cummins, S.; Macintyre, S. Are secondary data sources on the neighbourhood food environment accurate? Case-study in glasgow, UK. *Prev. Med.* **2009**, *49*, 527–528.
13. Liese, A.D.; Colabianchi, N.; Lamichhane, A.P.; Barnes, T.L.; Hibbert, J.D.; Porter, D.E.; Nichols, M.D.; Lawson, A.B. Validation of 3 food outlet databases: Completeness and geospatial accuracy in rural and urban food environments. *Am. J. Epi* **2010**, *172*, 1324–1333.
14. Lanvin, M.R. A clash of the titans: Comparing America's most comprehensive business directories. *Database* **1998**, *21*, 44–48.
15. Lake, A.A.; Burgoine, T.; Greenhalgh, F.; Stamp, E.; Tyrrell, R. The foodscape: Classification and field validation of secondary data sources. *Health Place* **2010**, *16*, 666–673.
16. Svastisalee, C.M.; Holstein, B.E.; Due, P. Validation of presence of supermarkets and fast-food outlets in copenhagen: Case study comparison of multiple sources of secondary data. *Public Health Nutr.* **2012**, *15*, 1228–1231.
17. Svastisalee, C.M.; Nordahl, H.; Glümer, C.; Holstein, B.E.; Powell, L.M.; Due, P. Supermarket and fast-food outlet exposure in copenhagen: Associations with socio-economic and demographic characteristics. *Public Health Nutr.* **2011**, *14*, 1618–1626.
18. Toft, U.; Erbs-Maibing, P.; Glümer, C. Identifying fast-food restaurants using a central register as a measure of the food environment. *Scan. J. Public Health* **2011**, *39*, 864–869.

19. Eurostat-European Commision. *NACE Rev. 2—Statistical Classification of Economic Activities in the European Community*; Eurostat Methodologies and Working Papers; Office for Official Publications of the European Communities: Luxembourg, Luxembourg, 2008.
20. Danish Business Agency. CVR.dk. 2013. Available online: http://www.cvr.dk (accessed on 17 September 2013).
21. Ministry of Food, Agriculture and Fisheries, Danish Veterinary and Food Administration. 2013. The Smiley System. Available online: http://www.findsmiley.dk/en-US (accessed on 17 September 2013).
22. Powell, L.M.; Han, E.; Zenk, S.N.; Khan, T.; Quinn, C.M.; Gibbs, K.P.; Pugach, O.; Barker, D.C.; Resnick, E.A.; Myllyluoma, A.; *et al.* Field validation of secondary commercial data sources on the retail food outlet environment in the US. *Health Place* **2011**, *17*, 1122–1131.
23. Bovell-Benjamin, A.C.; Hathorn, C.S.; Ibrahim, S.; Gichuhi, P.N.; Bromfield, E.M. Healthy food choices and physical activity opportunities in two contrasting Alabama cities. *Health Place* **2009**, *15*, 429–438.
24. 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*, doi:10.1186/1479-5868-5-58.
25. De Smith, M.J.; Goodchild, M.F.; Longley, P. *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*; Matador: Leicester, UK, 2007.
26. Geoteam. Om GPSNET.dk (About GPSNET.dk). 2013. Available online: http://www.geoteam.dk/produkter/gpsnetdk/om-gpsnetdk.html (accessed on 17 September 2013).
27. Hosler, A.S.; Dharssi, A. Identifying retail food stores to evaluate the food environment. *Am. J. Prev. Med.* **2010**, *39*, 41–44.

APPENDIX III – PAPER III: BEHAVIOURAL NUTRITION & BIG DATA: HOW GEODATA, REGISTER DATA & GPS, MOBILE POSITIONING, WI-FI, BLUETOOTH & THERMAL CAMERAS CAN CONTRIBUTE TO THE STUDY OF HUMAN FOOD BEHAVIOUR

Behavioural Nutrition & Big Data: How Geodata, Register Data & GPS, Mobile Positioning, Wi-Fi, Bluetooth & Thermal Cameras Can Contribute to the study of human food behaviour

B.E. Mikkelsen[1], A.K. Lyseen[1], M. Dobroczynski[2], and H.S. Hansen[1]

[1]Department of Development and Planning, University of Aalborg, Copenhagen, Denmark. bemi@plan.aau.dk

[2]sysCore ApS, Copenhagen, Denmark.

Introduction

Place has traditionally been providing the conceptual and analytic platform for the studying physical environment and its relation to food behaviour [1, 2]. Where human activities previously were closely attached to the locations of home or work perspective of human activities has shifted towards a people-based view [3, 4]. Complexity and heterogeneity of human mobility no longer appear to correspond the use of residential neighbourhoods, but stress the need for methods and measures of individual activity and exposure, along with a change in awareness of researchers that human behaviour and activity is individual/people-based and not place-based [2, 4]. Portable intelligent devices has created a series of new opportunities for convenient assessment of food choice and as a result, food and lifestyle related behaviour is increasingly becoming the subject of measurement through application of such mobile devices [5, 6]. Intelligent devices such as smartphones, touch pads, etc. increasingly becomes used by consumers not only to get online using different wireless technologies but also for self-monitoring of lifestyle. For the research community this development also offers new opportunities since the devices can also be used in a reverse mode to track the behaviour of individuals applying GPS, mobile positioning, Wi-Fi and Bluetooth. Furthermore, thermal cameras offer new possibilities to track human behaviour, contrary to regular RGB-cameras where individuals are recognisable and subject for ethical issues. Application of GPS tracking within behavioural nutrition research [7] has been limited, but other research areas has embraced the technology and used it for task such as measuring travel patterns [8, 9], wildlife movement and habitats [10], exposure to toxics, pesticides or air pollution [11], following elderly with Alzheimer’s and other dementia [12], and within health research to measure physical activity [13]. Use of mobile positioning, Wi-Fi and Bluetooth signals for tracking is more novel technologies in terms of tracking and less used in research for tracking than GPS. Application potentials of Bluetooth and Wi-Fi are comparable and has been used for measuring travel time through airport security [14], travel time on freeways [15] or mapping large crowds at mass events [16]. GSM has been sparsely used in for tracking with the exception of Ahas et al. [17], who used it for GSM tracking of tourists.

Use of the smartphones available in a study sample are believed to have large potential within social science, due to the use of apps and the large population penetration [18]. The penetration of smartphones worldwide has gone from 35% to 56% in two years and for cell phones in general to 91% [19] making the population basis for passive tracking with Wi-Fi, Bluetooth or mobile positioning profound. Normal RGB-cameras has long been used for surveillance but the ability to recognize people often poses a problem. Thermal cameras can replace RGB-cameras for tracking, as it is anonymized data. Traditionally nutritional research has used food diaries, questionnaires and interviews to analysis of food behaviour [20]. This is not uniquely to food studies, also transport research has previously frequently applied travel diaries, but implementation of GPS has reduced some of the shortcomings such as poor data quality, lack of reporting short trips, total trip times and destination locations [4, 8]. Functionalities related to Geographical Information Systems (GIS) offers simple representations of the physical environment including its opportunities for physical activity and food. The collection of behaviour information leads to the creation of very huge dataset, which induce problems for analysis. Hence, changing from place-based polygon features to people based point features requires different analysis and an increased focus on data administration and cleaning.

The contribution of this paper is on monitoring activity to improve the comprehension of food behaviour, which has numerous examples of placed-based studies [21, 22] and only a few people-based [7, 20]. This paper aims at giving an overview of the options available through these new technologies. The paper aims at an assessment of pros and cons for different type of tracking technologies and application setups. It gives examples of combinations of the technologies, and finally discusses the reach of these new opportunities along with a discussion of the ethical dimensions of such tracking.

Comparison of technologies – strengths and weaknesses

Deciding on which technology to use for a given research problem can be a straining problem, but very crucial to get right the first time. Knowing and understanding the strengths and weaknesses of the tracking technologies are the keys to selecting one or more technologies fitting the scale and environment of the research. The best fit of a technology in a study is influence by the environment, the extent of the study area, the required accuracy in the positioning, the need for respondents (active tracking) or for a only the movement (passive tracking) and the availability and pricing of the hardware and data. The environment for a study is often indoor, outdoor, or a combination. The GNSS and A-GPS technologies are only suitable for outdoor tracking. The thermal cameras require the environment to have open spaces with limited objects that block the view, whereas BT, Wi-Fi and mobile positioning copes with both outdoor and indoor tracking, regardless of the building layout and design. Only exception to this is, if the walls inside the buildings are blocking the short-wavelength radio waves and microwaves in for example BT and Wi-Fi, which is solved by adding additional BT sensors or Wi-Fi hotspots. The accuracy in estimated positions of the devices spans widely from approximately 1 metre to 20 kilometres dependent on the environment, hardware and signals. The accuracy of GNSS depends heavily on the environment of the study. With high buildings and narrow streets, the accuracy can easily be as bad as 20-30 meters dependent on the equipment, time of day and whether the equipment can use satellites from more than one system. In a bare field the accuracy can be as good as 1-5 meters again influenced by the same elements. The accuracy of BT and Wi-Fi can vary greatly based on two parameters, the range of the sensors and the possibility to triangulate between several sensors. The accuracy with no overlapping sensors will never be better than the scanning range of the sensors and due to the nature of the signals; the range of the sensors is a bit fuzzy making the exact precision of the positioning a bit uncertain. The accuracy of mobile positioning is the worst among the technologies mentioned in this paper. As mentioned in Ahas [17] there are several methods and network standards to base the tracking upon, of which A-GPS is the most accurate with 3 meters or better in open spaces and 20 meters in urban high-rise environments. Tracking solely on the cellular network would yield accuracies from under 100 meters to 20 kilometres, very much dependent on the environment and the density of the cellular network. Thermal cameras have proven to have a high position accuracy of 1-2 meters.

The size of the study area combined with the environment and the needed accuracy has a large impact on choosing the technology, the amount of devices needed and thereby the cost of the tracking. GNSS, A-GPS and mobile positioning are generally best suited for large areas as the individuals are carrying the technology necessary for tracking, while using BT, Wi-Fi and thermal cameras stationary sensors are needed. BT, Wi-Fi and thermal cameras are in theory plausible to use for large scale tracking at a very high price for equipment proportional to the amount needed or a low density of sensors only covering certain zones in the study area. Only covering all the supermarkets in an area could be an option if only the presence in a shop is relevant for the aim of the study. GNSS and A-GPS are the most accurate technologies to capture movement over a large area, but both requires respondents acceptance, while mobile positioning covers over 90% of the population worldwide without their permission is required. Choosing a passive or an active tracking technology both has consequences. Passive tracking technologies such as thermal cameras and mobile positioning have the potential for following close too every person. However, if BT and Wi-Fi are used for passive tracking, only a proportion of the population is registered, as only some of the customers will have either or both of the signals turned on. There are methods to bypass this and get a larger part of the customers tracked through inserting tags in shopping baskets and trolleys, encourage customers to turn on the signals with prizes or other benefits for the customer or combine BT and Wi-Fi to track on both signal types. If the goal of the study is to count the customers or measure the time spent at the shop, one or two sensors might be enough dependent on the amount of entrances and exits. With passive tracking only the movement patterns of individuals are the output, whereas with active tracking is it possible to join additional information about each person's health status, socioeconomic status, etc. However, active tracking requires the consent of each respondent. Willingness to participate in tracking varies across the population and the proportion that completes the study satisfactory are often in the range of 50-60%. GNNS and A-GPS are active tracking technologies only, whereas thermal cameras are unsuited for active tracking.

The technical knowledge needed to apply the technologies for tracking differ as GNSS have a huge amount of off-the-shelf solutions and delivers an output that needs little or no processing to be implemented in GIS. Mobile positioning, BT and Wi-Fi are all based on a cell structure, from which a processing is needed to change the data to point features with coordinates. A-GPS requires an app to handle the tracking, which needs development or buying and existing that fits the purpose and hardware. Thermal cameras require the most processing as movement detection is still far from commercial use. The technologies are in theory applicable worldwide, but in reality are GNSS, A-GPS and mobile positioning the only technologies, as the rest requires power supply or changing of batteries at regular basis. Mobile positioning is limited by the goodwill of the operating companies and the legal clauses in each country.nThe prize of a complete tracking setup is influenced by the amount of devices, the type of technology and the accuracy. In many cases, A-GPS probably is the cheapest option, as a large segment of populations already owns a smartphone and the technology is well suited for combination with BT and Wi-Fi tracking inside selected shops. Combination of the technologies is preferred, to utilize the strengths of each technology in different environment settings.

Discussion

Developments in wearable positioning technologies and GIS provide an opportunity for understanding and controlling many phenomena occurring in urban areas. The position technologies offer quantifiable measures of individual's movement and exposure as they make decisions in real-time. The degrees autonomy of people varies when making decisions about residence and work, who to socialize with and where to do that, as neither the individuals nor the living environment is static. It is unrealistic to assume that the majority of people spend all or most of their time in pre-defined geographical areas. The high levels of mobility in a population requires methods to measure this without limitations, as the residential or administrative boundaries, and the residential and working addresses not necessarily are the best identifier for our dietary behaviour. Residential and administrative boundaries may not provide the most adequate basis for analysis of the impact of place on health. Determination of the most appropriate scale for analysis of places influence on dietary behaviour or preferably, to apply flexible scales to fit the patterns of every person, would increase the understanding of individual and group behaviour. Potentially tracking technologies could help loosening the dependency of residential and administrative boundaries for representing place, by an individual measure of behaviour.

Traditionally dietary research has focused on questionnaires and interviews for analysis, but it is known that people have a tendency to embellish the reality about information on food intake. Likewise have self-reported information about trips and whereabouts been incomplete in especially short trips, start and end time of trips, and the addresses visited. The development of wearable positioning provides a more objective measure for the behaviour in terms of space-time information. The potential of the tracking technologies for research on dietary behaviour are great for measurement of exposure to example healthy and unhealthy food options. GIS provides the tools for combining the individual behaviour patterns with personal information on health and socioeconomic status and conduct statistical and spatial analysis. The purpose of implementing the tracking technologies in dietary research is not to replace the questionnaires and interviews on dietary behaviour, but perceived to add to a growing toolbox for researchers. Combining the data from questionnaires and interviews with tracking data increases the potential analysis potential.

Questionnaires, interviews and tracking are all invasive on the persons followed as the question presented by researchers and knowledge of being followed possibly affects the person's behaviour. Researchers ask for information that is often regarded personal or confidential as the whereabouts or food intake. This is active tracking and often requires a lot effort from the respondents, which also have a great influence on which segments of the population that are willing to participate. Passive tracking, on the contrary, removes the demands on the respondents and provides a measure of human movement and behaviour in a non-invasive manner. Passive tracking often have the ability to capture the behaviour of large groups of the populations i.e. because of the high cell phone penetration in the population, but at the cost of losing the supplementary information of individuals. Both active and passive tracking offers great possibilities for research in term of tracking the general populations' movement patterns and shopping patterns or smaller samples of the populations' behaviour and its impact on health.

The choice of technology should not necessarily be limited to selecting one, as several of the technologies combined would enable tracking at several scales and accuracies to serve the objective of the study. Example, use of GNNS or A-GPS to track the overall movement of individuals in a city and then in all the supermarket and grocery shops set up Wi-Fi and BT tracking to track the movement inside the buildings. Then the tracking provides information of both the person's choice in food retailer and the choices made inside each individual food shop.

Use of wearable position technologies to follow the behaviour of individuals is an intrusion into people's privacy and violates many people's boundaries. Surveillance is in many countries frowned upon and raises a whole host of ethical issues for many people and governments. Privacy is a public concern, which causes debate about personal freedom and scientific ethics. Privacy, surveillance and data security are key aspects in both passive and active tracking. In active tracking, the person's identity is known to the researchers, who have the key responsibility, when storing, analysing and publishing results of tracking, to ensure the confidentiality of the participating persons. European and national legal regulations for data collection must be followed to ensure the persons and their locations are not identifiable. Privacy and ethical concerns influences the type of people willing to participate in studies involving tracking. Some of the concerns revolve around the fear of being subject to surveillance, and followed and listened to everywhere one goes. The growing ICT generation may be more open to the technologies, and the accustoming of positioning technologies may lower the concerns. Explanation of the aspects and demonstration of the results of tracking could decrease the concerns of privacy violations and data security. Private tracking have to treat the people in way that ensures their complete confidentiality also in the data. Thermal cameras ensure this compared to RGB-cameras, while application of other methods are needed for BT, Wi-Fi and mobile positioning. There are several methods [4] i.e. assigning 'dummy' variables to ID a device instead of using the MAC-address.

References

1. Leal, C. and B. Chaix. (2011). The Influence of Geographic Life Environments on Cardiometabolic Risk Factors: A Systematic Review, a Methodological Assessment and a Research Agenda. *Obesity Reviews* **12**, no. 3: 217–230.
2. Matthews, S.A. and Tse-Chuan Y. (2013) Spatial Polygamy and Contextual Exposures (SPACEs) Promoting Activity Space Approaches in Research on Place And Health. *American Behavioral Scientist* **57**, no. 8: 1057–1081.
3. Miller, H. (2007). Place-Based versus People-Based Geographic Information Science." *Geography Compass* **1**: 503–535.
4. Rainham, D., McDowell I., Krewski D. and Sawada M.. (2010). Conceptualizing the Healthscape: Contributions of Time Geography, Location Technologies and Spatial Ecology to Place and Health Research. *Social Science & Medicine* **70**, no. 5: 668–676.
5. Ngo J., Engelen A., Molag M., Roesle J., García-Segovia P. and Serra-Majem L. (2009). A review of the use of information and communication technologies for dietary assessment. *British Journal of Nutrition* **101**: 102-112.
6. Thompson F.E., Subar A.F., Loria C.M., Reedy J.L. and Baranowski T. (2010). Need for technological innovation in dietary assessment. *Journal of American Diet Associasion* **110**: 48.
7. Zenk, S.N., Schulz A.J., Matthews S.A., Odoms-Young A., Wilbur J., Wegrzyn L., Gibbs K., Braunschweig C. and Stokes C. (2011) Activity Space Environment and Dietary and Physical Activity Behaviors: A Pilot Study. *Health & Place* **17**, no. 5: 1150–1161.
8. Murakami, E., and Wagner, D.P. (1999). Can Using Global Positioning System (GPS) Improve Trip Reporting? *Transportation Research Part C: Emerging Technologies* **7**, no. 2–3: 149–165.
9. Stopher, P., FitzGerald, C., and Zhang, J. (2008). Search for a Global Positioning System Device to Measure Person Travel. *Transportation Research Part C: Emerging Technologies* **16**, no. 3: 350–369.
10. Markham, A.C. and Altmann J. (2008). Remote Monitoring of Primates Using Automated GPS Technology in Open Habitats. *American Journal of Primatology* **70**, no. 5: 495–499.
11. Phillips, M.L., Hall, T.A., Esmen, N.A., Lynch, R., and Johnson, D.L. (2001). Use of Global Positioning System Technology to Track Subject's Location during Environmental Exposure Sampling. *Journal of Exposure Analysis & Environmental Epidemiology* **11**, no. 3: 207.
12. McShane, R. and Skelt L. (2009) GPS Tracking for People with Dementia. *Working With Older People* **13**: 34–37.
13. Rodriguez, D.A., Brown, A.L. and Troped, P.J. (2014) Portable Global Positioning Units to Complement Accelerometry-Based Physical Activity Monitors. *Medicine and Science in Sports and Exercise* **37**: 572–581.
14. Bullock, D., Haseman, R., Wasson, J. and Spitler, R. (2010) Automated Measurement of Wait Times at Airport Security. *Transportation Research Record: Journal of the Transportation Research Board* **2177**: 60–68.
15. Haghani, A., Hamedi, M., Sadabadi, K., Young, S. and Tarnoff, P. (2010) Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors. *Transportation Research Record: Journal of the Transportation Research Board* **2160**: 60–68.
16. Versichele, M, Neutens, T., Delafontaine, M. and Van de Weghe, N. (2012) The Use of Bluetooth for Analysing Spatiotemporal Dynamics of Human Movement at Mass Events: A Case Study of the Ghent Festivities. *Applied Geography* **32**: 208–20.
17. Ahas, R, Aasa, A., Roose, A., Mark, Ü. and Silm, S. (2008) Evaluating Passive Mobile Positioning Data for Tourism Surveys: An Estonian Case Study. *Tourism Management* **29**: 469–86.
18. Raento, M., Oulasvirta, A. and Eagle, N. (2009) Smartphones An Emerging Tool for Social Scientists. *Sociological Methods & Research* **37**: 426–454.
19. Infographic. (2013) Mobile Growth Statistics | Digital Buzz Blog. <http://www.digitalbuzzblog.com/infographic-2013-mobile-growth-statistics/>. Accessed 28 February 2014.
20. Kestens, Y., Lebel, A., Chaix, B., Clary, C., Daniel, M., Pampalon, R., Theriault, M. and Subramanian, S.V.P. (2012) Association between Activity Space Exposure to Food Establishments and Individual Risk of Overweight. *PLoS ONE* **7**.
21. Leung, C.W., Laraia, B.A., Kelly, M., Nickleach, C., Adler, N.E., Kushi, L.H., and Yen, I.H. (2011) The Influence of Neighborhood Food Stores on Change in Young Girls' Body Mass Index. *American Journal of Preventive Medicine* **41**, no. 1: 43–51.
22. Liu, G.C., Wilson, J.S., Qi, R. and Ying, J. (2007) Green Neighborhoods, Food Retail and Childhood Overweight: Differences by Population Density. *American Journal of Health Promotion* **21**: 317–25.

APPENDIX IV – PAPER IV: DEFINING NEIGHBOURHOODS AS A MEASURE OF EXPOSURE TO THE FOOD ENVIRONMENT

Int. J. Environ. Res. Public Health **2015**, *12*, 8504-8525; doi:10.3390/ijerph120708504

OPEN ACCESS

International Journal of
Environmental Research and Public Health
ISSN 1660-4601
www.mdpi.com/journal/ijerph

Article

Defining Neighbourhoods as a Measure of Exposure to the Food Environment

Anders K. Lyseen [1,*], Henning S. Hansen [1], Henrik Harder [2], Anders S. Jensen [3] and Bent E. Mikkelsen [4]

[1] Department of Development and Planning, Aalborg University, Copenhagen, A.C. Meyers Vænge 15, 2450 Copenhagen, Denmark; E-Mail: hsh@plan.aau.dk
[2] Department of Architecture, Design and Media technology, Aalborg University, Rendsburggade 14, 9000 Aalborg, Denmark; E-Mail: hhar@create.aau.dk
[3] LE34, Energivej 34, 2750 Ballerup, Denmark; E-Mail: asj@le34.dk
[4] Department of Clinical Medicine, Aalborg University, Copenhagen, A.C. Meyers Vænge 15, 2450 Copenhagen, Denmark; E-Mail: bemi@plan.aau.dk

* Author to whom correspondence should be addressed; E-Mail: alyseen@plan.aau.dk; Tel.: +45-99-408-356.

Academic Editor: Paul B. Tchounwou

Received: 6 April 2015 / Accepted: 15 July 2015 / Published: 21 July 2015

Abstract: Neighbourhoods are frequently used as a measure for individuals' exposure to the food environment. However, the definitions of neighbourhoods fluctuate and have not been applied consistently in previous studies. Neighbourhoods defined from a single fixed location fail to capture people's complete exposure in multiple locations, but measuring behaviour using traditional methods can be challenging. This study compares the traditional methods of measuring exposure to the food environment to methods that use data from GPS tracking. For each of the 187 participants, 11 different neighbourhoods were created in which the exposure to supermarkets and fast food outlets were measured. ANOVA, Tukey's Honestly Significant Difference (HSD) test and *t*-tests were performed to compare the neighbourhoods. Significant differences were found between area sizes and the exposure to supermarkets and fast food outlets for different neighbourhood types. Second, significant differences in exposure to food outlets were found between the urban and rural neighbourhoods. Neighbourhoods are clearly a diffused and blurred concept that varies in meaning depending on each person's perception and the conducted study. Complexity and heterogeneity of human mobility no longer appear to correspond to the use of residential neighbourhoods but rather emphasise the need for methods, concepts and measures of individual activity and exposure.

Keywords: food environment; neighbourhood; exposure assessment; geographic information systems; Global Positioning System; activity spaces

1. Introduction

Studies of nutrition and physical activity behaviour in the past decade have recognised the importance of the environment in understanding health and health related behaviour [1–4]. Within nutritional research, an increased focus has been placed on measuring the impact of the food environment on health outcomes such as Body Mass Index (BMI) [5–8], body weight [9,10], obesity [11,12] and diet [3,10,13]. The environmental exposure is often conceptualised through and measured within neighbourhoods. However, the spatial extent of neighbourhoods has proven difficult for researchers to define, and the result is a great variation in the definitions of neighbourhood used to study the environmental exposure [2].

The method used to define a neighbourhood is essential for researchers to ensure that measured exposure reaches optimal agreement with the actual exposure. However, for researchers to achieve this result, they must scrutinise the behaviour carefully to fully understand the phenomenon. The way a neighbourhood is defined should reflect the context of its application [14]. Therefore, when measuring the food environment, researchers must make qualified assumptions about where people shop or dine, the distance people are willing to travel for shopping or dining and other individual preferences [2].

Applying neighbourhoods to measuring food exposure creates a manageable concept to analyse the effect of the exposure. However, variations in neighbourhood definitions indicate that not all definitions manage to conceive and measure the actual exposure equally well [3,15]. Giles-Corti *et al.* found little agreement among previous studies on the appropriate distance from home, work or school to search for a relationship to physical activity [16]. A study in Seattle found that 49% of participants had greater exposure to supermarkets outside their home neighbourhood [17]. Similar results were found in Minnesota, where the participants had more than twice the exposure at work than at home [11].

That defining neighbourhoods presents challenges seems evident, and several studies appear to agree on several suggested challenges [9,15,18–20]. Ball *et al.* [1] explain that (1) people live and function in multiple contexts and settings; (2) people live and work in multiple geographic areas; and (3) different types of environmental influences exist, including built, natural, social, cultural and policy environments. Consequently, methods used for defining neighbourhoods must comply with individual behavioural characteristics. Focus on the individual is conceptualised by Rainham *et al.* through the change from a place-based to a people-based perspective with individual-based measures [21].

Previous studies reveal numerous examples that contradict the people-based approach through application of administrative divisions as the spatial extent for a neighbourhood [18,22]. Census tracts [23–26], zip codes [22] or parishes are used as a spatial representation of a neighbourhood for analysis of exposure to the food environment.

Neighbourhoods based on buffers also rely strongly on the location of the home but also offer an individual measurement. However, the difference is small for people living close to one another. The buffer method is widely used [16] to create neighbourhood definitions for residences [5,9,24,27], schools [13,28–33] and work locations [9]. The buffer distances and methods varies between fixed distances or a travel time constraint and either Euclidian or network distances [27].

Administrative divisions and buffers applied to the residential location adhere to a conceptual and analytic platform, where place is the central element in studying human behaviour. From the place-based perspective, all behaviour is located and centralised around the home. The importance of people's closeness and sense of belonging to a certain community and place is challenged by today's society. No matter what one believes, human mobility has increased substantially in the last century, and connectivity now makes activities and places more dynamic.

The problem is that each individual is unique and consequently must be assumed to have their own concept of neighbourhood. Complexity and heterogeneity of human mobility no longer appear to correspond to the use of residential neighbourhoods. Exposure to the food environment occurs in multiple environments, but to measure the impact of people's individual exposure in multiple environments is challenging.

Technologies for tracking individuals' behaviour have been available for more than a decade. However, development of lightweight, low-cost and accurate Global Position System (GPS) devices and assisted GPS in smartphones has boosted the use of tracking within behavioural nutrition research. GPS provides an individual measurement of space-time information about people's behaviour. The outcome of GPS tracking can potentially consist of millions of data entries, which must be handled and conceptualised to resemble a neighbourhood. Common methods for simplifying neighbourhoods (or activity spaces) from GPS data are standard deviational ellipses (SD ellipses) and home range (minimum convex polygon) [21,34]. The derived activity spaces are individual and not dependent on a fixed location. Commuting routes and leisure time activities are therefore also included.

Although many studies utilise neighbourhood as a concept, few studies explore how neighbourhoods are defined or which definition is most suitable for the study. A variety of neighbourhood definitions are applied in relation to measuring the impact of the food environment.

Therefore, the aims of this study are (1) to compare different definitions of neighbourhoods for analysis of exposure to healthy/unhealthy food options, where supermarket exposure is perceived to be healthy and fast food exposure to be unhealthy; (2) to investigate the differences in neighbourhood area size and in the number of food outlets by type within neighbourhoods; and (3) to discuss the influence of the neighbourhood definition on the measure of exposure.

2. Methods

2.1. Study Area and Sample

The study area consists of 65 parishes (15 urban and 50 rural) in Northern Jutland (Denmark) centralised around Aalborg as the largest city in the region. The population in the study area is approximately 230,000, and of that number, approximately 120,000 live in Aalborg. The study area is approximately 1552 km^2, of which Aalborg, with its high-density housing (mean ≈ 1700 people/km^2) only comprises 68.3 km^2 (≈4.4%). The remaining areas consists of small villages with populations up to 7000 and low-density housing (mean ≈ 85 people/km^2). The study area's spatial extent, relative location in Denmark and the divide in urban and rural areas are presented in Figure 1. Northern Jutland consists of 11 municipalities, five of which are defined as peripheral regions. Peripheral regions are characterised by, among other factors, a lower average income than the national average, a lower amount of commuting traffic and low or negative population growth. However, Aalborg attracts many young people and is the economic centre of the region. In Northern Jutland, approximately 50% of all people aged 16 to 25 lives in Aalborg, whereas these people are only approximately 17% of the entire population.

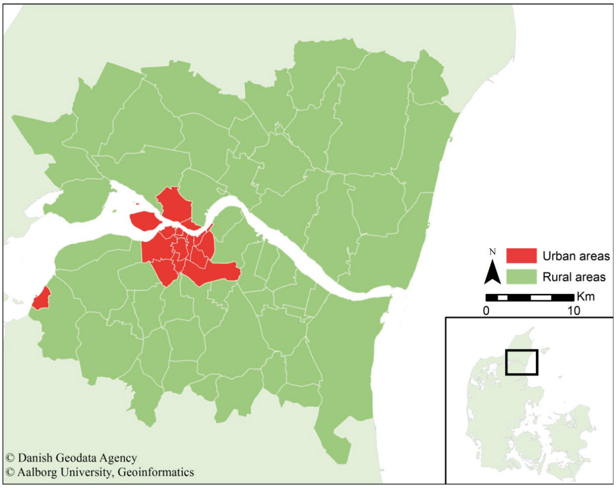


Figure 1. Presentation of the study area, the relative location in Denmark and the division between urban and rural areas.

The study involves a random sample of 223 people selected from a population of 7277 people enrolled in school in Aalborg. Respondents were distributed between six school locations. The sample has a higher proportion of female (57%) than male (43%) participants. The participants' ages range from 16 to 23 years old, with an average age of 17.7 years. Each person was tracked by the Global Positioning System (GPS) for one week of their typical school schedule. The GPS devices used in this study are the Lommy Phoenix and are approximately the same size as a mobile phone. The participants were asked to carry the device at all possible times during the week. All subjects provided their informed consent for inclusion before they participated in the study and could opt out at any time by turning off the GPS device. The tracking resulted in 8.22 million records for the 223 participants. The number of loggings registered for each person varied from 579 to 128,679, with an average of 36,523.

A threshold of 30 h (equal to waking hours for two days) of tracking was set as a minimum for the participants to be included in the study. The final sample consists of 187 people (36 were excluded). The final sample population includes 110 women (58.8%) and 77 men (41.2%) from 16 to 23 years old (the mean age is 17.3 years old). The final sample includes 93 people who live in a rural area and 94 people who live in an urban area.

2.2. GPS Data Preparation

GPS tracking is subject to several technical limitations when measuring space-time data [19,35]. Connection to an adequate amount of satellites is critical because lack of such a connection can result in inaccurate position data or complete loss of data for a period. The errors can be categorised as (1) outliers, either in attribute values for number of satellites, horizontal delusion of precision (HDOP) and time to fix (TTF), or extreme positions (e.g., on equator); or (2) scatter, in the form of unnatural linear point patterns [35]. The unnatural linear point patterns are detected by little or no change in the direction between three or more subsequent loggings, and the location of these loggings are outside a 50 m buffer on the road network. Detection of outliers and scatter found 341,741 loggings that were perceived as erroneous data.

The GPS devices were set to register the location at 7 s intervals, which was the lowest interval possible for the devices used. However, due to external conditions (*i.e.*, visibility to satellites and time to establish a fix), the logging interval varies up to 60 s. Calculation of several neighbourhood definitions assumes an even time interval between loggings (e.g., SD ellipses) because they are based on statistical assumptions. Spatial linear interpolation between subsequent loggings was applied to create an even time interval of 1 s between each logging. However, a 60 s threshold is set because the GPS creates a duplicate of the previous logging if it cannot obtain three consecutive measurements with a HDOP less than 30 in 60 s. The consequence can be large time leaps, for which it is difficult to estimate or guess the location. The interpolation results in a data set consisting of 60.18 million loggings, which corresponds to an average of three days and 17.4 h of active tracking for each participant.

2.3. Neighbourhood Definitions

2.3.1. Administrative Divisions

Division of the land into smaller areas is used administratively on several levels in most countries, and previous studies refer to census tracts and zip codes used for spatial analysis. The purposes of the administrative division vary, but none were created for research purposes. The consequence of using administrative divisions as measures of exposure to the food environment implies that all individuals within these divisions will be exposed solely to the food outlets within those boundaries. Thus, it relies on people to have a strong residential connection.

This study uses parishes because they are the smallest official administrative division in Denmark. The area size of parishes within the study varies from 0.65 to 110.49 km^2 (mean = 23.85 km^2), the population ranges from 98 to 12,544 people and the population density varies from 14.39 to 9097 people/km^2. People were assigned to the parish in which their residence is located.

2.3.2. Buffers

Buffers are used to create a circular area at a specified distance, and they are quick to calculate, easy to understand and easy to compare because the area size is equal for all study subjects. Simple buffers are based on Euclidian distances, whereas buffers that are more complex are based on network analysis. The buffer distance should be appropriate for examining nutrition-related behaviours for the target group involved. Little agreement exists on the appropriate distance, and multiple distances are applied in research [16]. This study applies two distances for defining the buffer size. A distance of 800 m was selected because it is approximately equal to a 10 min walk (5 km/h). Second, a distance of 1600 m (≈1 mile) was selected because it is frequently used in other studies [5,9,13,16,24,28,29,32]. A study of adults in England demonstrated that more than 95% of usual walking destinations were within 1600 m of the home [36]. This study calculates buffers on the home and school addresses. A third neighbourhood definition is defined by combining the buffers for home and school.

2.3.3. Convex Hull (Minimum Bounding Geometry)

The convex hull area is created to represent the minimum bounding geometry enclosing all the GPS loggings for each individual. The convex hull represents the maximum area in which the individuals engaged in activities.

2.3.4. Standard Deviational Ellipses

The standard deviational (SD) ellipses are created by calculating the standard deviation in the x-coordinates and y-coordinates from the mean centre of the coordinates. The ellipses do not represent the maximum area in which the individual could engage in activities but rather the area in which the individual is likely to be regularly involved in activities. This study applies one and two SD ellipses, which implies that approximately 68% and 95% or more of the GPS loggings are positioned within the one or two SDs, respectively. The position of each GPS logging is a weight in calculating the ellipses extent. The GPS loggings therefore must represent an individual's whereabouts, which is performed through interpolation on the space-time data.

2.3.5. Path Area

The GPS loggings are used to create the path area represents the participants' travel patterns. For each GPS logging, the nearest road or path segment was determined through a near analysis. On the road and path segments, a 50 m buffer was applied. The buffer is needed to capture the exposure to food outlets, for which spatial location often has an offset of 5–30 m from roads.

Figure 2 presents a spatial comparison of the neighbourhood definitions.

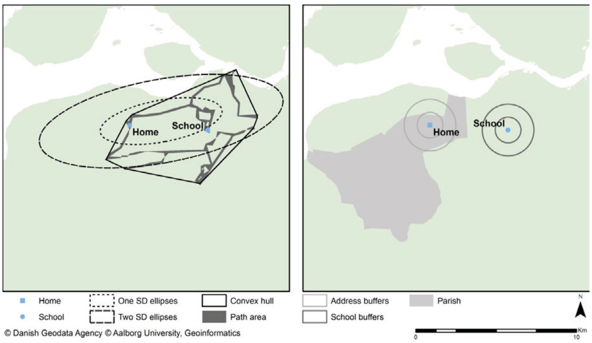
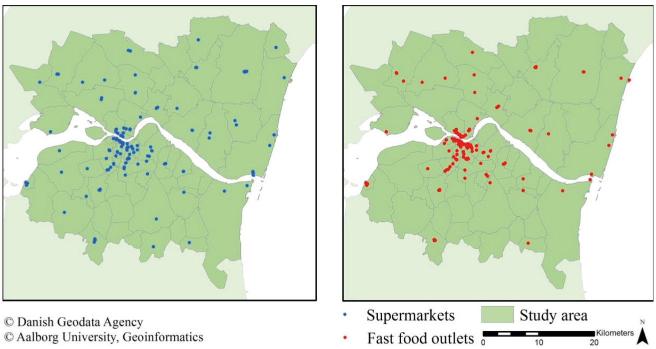


Figure 2. Visual representation of neighbourhood spatial extent and definition.

2.4. Food Outlet Data

Data on fast food outlets and supermarkets were retrieved from the national business register (CVR) and the national food safety and hygiene regulation register (Smiley). The spatial and semantic validity has been described in previous research [37]. A pre-classification method of the business type based on the outlets name was applied as described in [37]. This resulted in 144 supermarkets (including discount) and 154 fast food outlets in the study area. The addresses in CVR were geocoded based on address reference data in the Universal Transverse Mercator (UTM) projection obtained from the Danish Geodata Agency. The Smiley register contains World Geodetic System 84 (WGS84) coordinates for approximately 95% of entries, which were transformed into UTM and used as their locations. The remaining records are geocoded by the address using reference data from the Danish Geodata Agency. The distribution of the supermarkets and fast food outlets is depicted in Figure 3.



(**a**) (**b**)

Figure 3. The spatial distribution of (**a**) supermarkets; and (**b**) fast food outlets within the study area.

2.5. Statistical Analysis

This study compares the mean values for food outlets exposure in each neighbourhood to analyse differences. Consequently, the null hypothesis is that any difference between the groups is a result of sampling error, and the actual differences between the means are effectively zero. The Welch two-sample *t*-test is applied to compare two groups, and the one-way ANOVA (*F*-test) is applied for comparing three or more groups.

One-way ANOVA assumes that the data are sampled from populations that follow a Gaussian distribution. Although this assumption is not very important with large samples, it is important with small sample sizes and particularly with unequal sample sizes. One-way ANOVA assumes that all the groups have the same standard deviation. This assumption is not very important when all the groups have the same or almost the same number of individuals. The sample sizes in this study are equal for all one-way ANOVA tests.

The one-way ANOVA compares several groups but does not inform about groups having significantly different means. The differences between groups might be due to errors in the sampling whereas others might not be. Therefore, a post hoc comparison test is conducted to examine the differences between pairs of each of the neighbourhood types. This identifies pairs of neighbourhoods that have significantly large differences, which are not the result of sampling errors. This is calculated using Tukey's HSD (honest significant difference) test. Tukey's HSD test is weak, meaning it is less likely to detect significant results. The test assumes normality for each group of data, the observations are independent within and among groups and there is homogeneity of variance. The test is quite robust to violations of normality and to some extent violations of homogeneity of variance for large samples. Tukey's HSD test requires previous calculation of one-way ANOVA and is calculated using Equation (1). M_1 and M_2 are the means of the neighbourhood groups, MS_w is the mean square within groups from the one-way ANOVA and n is the number per group.

$$HSD = \frac{M_1 - M_2}{\sqrt{MS_w\left(\frac{1}{n}\right)}} \tag{1}$$

The Welch t-test is used to test the hypothesis that two independent or unpaired groups of data have equal means. The test is an adaption of the students' t-test, but it is used when the variance possibly is unequal. The test compares urban and rural samples, which are non-overlapping. The test assumes the data are independent. The Welch t-test is calculated using Equation (2), where $\bar{X}_i$ is the group means, S_i is the group variance and N_i is the group sample size.

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{2}$$

All statistical analyses are calculated using R [38].

3. Results

3.1. Comparison of Neighbourhood Area Sizes

There are 11 different definitions of neighbourhoods in this study with different spatial characteristics and extents as illustrated in Figure 2. Descriptive statistics for neighbourhood area sizes are presented in Table 1. The mean areas vary from 2.01 to 51.39 km^2. Significant dispersions occur for the neighbourhood type's parish, convex hull, one SD ellipses and two SD ellipses, which is reduced slightly by dividing the sample into urban and rural areas. No variance exists between buffers around schools or addresses due to equality of area sizes for all participants.

Table 1. Mean area and standard deviation for neighbourhoods for total sample (n = 187), urban (n = 94) and rural (n = 93) areas. Lower portion of table presents results of ANOVA for neighbourhoods.

Neighbourhood		Area (km^2)		Urban Area (km^2)		Rural Area (km^2)	
		Mean	**σ**	**Mean**	**σ**	**Mean**	**σ**
Place-based neighbourhoods							
Parish		17.80	20.05	5.70	3.94	30.02	22.28
Address 800 m buffer		2.01	-	2.01	-	2.01	-
Address 1600 m buffer		8.04	-	8.04	-	8.04	-
School 800 m buffer		2.01	-	2.01	-	2.01	-
School 1600 m buffer		8.04	-	8.04	-	8.04	-
Combined 800 m buffer		3.91	0.34	3.80	0.46	4.02	0
Combined 1600 m buffer		15.11	1.98	13.99	2.29	16.06	0.32
Person-based neighbourhoods							
Convex hull		51.13	82.30	21.14	34.93	81.45	103.02
One SD ellipses		17.78	40.55	4.53	6.26	31.17	54.08
Two SD ellipses		51.39	89.99	16.69	21.48	86.46	115.90
Path area		4.76	2.96	3.45	2.40	6.08	2.91
ANOVA	*F*-test values	39.83		24.34		35.48	
	Significance level	<0.001		<0.001		<0.001	
Tukey's HSD test		26 of 55 pairs have significant different means		27 of 55 pairs have significant different means		26 of 55 pairs have significant different means	

Table 1 suggests that the areas of rural neighbourhoods are noticeably larger than those in urban neighbourhoods. The Welch *t*-test compares the area sizes for urban and rural neighbourhoods, and the results are presented in Table 2. All *t*-values from the test are positive, indicating that the rural areas are larger than the urban areas. The *t*-values range from 4.671 to 10.369, and the significance levels for all neighbourhoods are below 0.001. The significance levels indicate that the null hypothesis of no difference in area sizes is rejected. Hence, the differences between urban and rural area sizes are most likely not due to sampling error.

Results of the one-way ANOVA are presented in Table 1. *F*-test values in the interval from 24.34 to 39.83 and significance levels below 0.001 indicates that the differences between mean area sizes of the 11 neighbourhoods has almost no chance of being caused by sampling error. The null hypothesis of no difference between mean area sizes is rejected. The results of Tukey's HSD test presented in the Supplementary Material provide information about which pairs of neighbourhood area sizes have significantly different means.

Table 2. The results of the Welch *t*-test comparing urban and rural neighbourhood area sizes. Buffer around school and address is omitted due to no difference.

Neighbourhood	t	df	Sig. (Two-Sided)	95% Conf. Interval of the Differences	
				Lower	Upper
Place-based neighbourhoods					
Parish	10.369	97.691	** <0.001	19.667	28.977
Address & school buffer 800 m	4.671	93.000	** <0.001	0.126	0.313
Address & school buffer 1600 m	9.427	96.655	** <0.001	1.773	2.718
Person-based neighbourhoods					
Convex hull	5.349	112.674	** <0.001	37.973	82.646
1 standard deviational ellipses	4.721	94.439	** <0.001	15.439	37.853
2 standard deviational ellipses	5.709	98.246	** <0.001	45.521	94.022
Path area	6.731	177.809	** <0.001	1.856	3.395

Note: ** Statistically significant below the 0.01 level.

Table 3 illustrates the average overlaps between neighbourhood types presented. Because the neighbourhoods vary in extent and location, the overlap between two neighbourhoods is not equal. The path area is the tightest measure of the participants' behaviour. Comparison of the overlaps between path area and the other 10 neighbourhoods reveals overlaps from 11.9% to 27.2%, which indicate that the other 10 neighbourhoods are only partially used.

3.2. Comparison of Neighbourhoods' Ability to Capture measured GPS Activity

Each participant's activity was measured using GPS and the neighbourhood types' convex hull and path area by definition captured 100% of the activity. The mean amount of loggings within each neighbourhood type is presented in Table 4. The neighbourhood types, which most poorly captured the GPS-measured activities, were the 800 and 1600 m buffers around schools. The remaining mean values range from 72.93% to 94.35% GPS loggings within the neighbourhoods.

Table 4 presents the results of the one-way ANOVA. The large *F*-test value of 509.8 and a significance level of <0.001 denote that the neighbourhood's ability to capture the measured GPS activity has almost no chance of demonstrating equal means for all 11 neighbourhoods. The null hypothesis of no difference between each neighbourhood's ability to capture human activity is rejected. Tukey's HSD test was calculated to compare the individual pairs and 47 out of 55 pairs were significantly different in mean amount for loggings located within the neighbourhood boundaries. The results of Tukey's HSD test are available in the Supplementary Material.

Tests were conducted by dividing the data into urban and rural areas. The Welch *t*-test reported significant differences for the school 1600 m buffer ($t = -3.220$, sig = 0.001) and combined 800 m buffer ($t = -4.894$, sig < 0.001). In both cases, the urban neighbourhoods captured a significantly larger proportion than the rural sample.

Table 3. Percentage of the column neighbourhood type that the row neighbourhood type overlaps, on average.

Neighbourhood	Address Buffer 800 m	Address Buffer 1 mile	Convex Hull	Address & School Buffer 800 m	Address & School Buffer 1 Mile	Path Area	School Buffer 800 m	School Buffer 1 Mile	1 Standard Deviational Ellipses	2 Standard Deviational Ellipses	Parish
Address 800 m buffer	-	24.7%	9.7%	52.0%	13.6%	15.5%	5.5%	3.7%	35.8%	17.8%	23.8%
Address 1600 m buffer	100.0%	-	23.9%	57.5%	55.1%	29.4%	18.2%	14.3%	58.3%	39.0%	52.5%
Convex hull	60.4%	45.4%	-	63.8%	48.2%	95.9%	67.9%	53.4%	78.2%	55.1%	38.9%
Combined 800 m buffer	100.0%	27.1%	18.1%	-	26.1%	31.1%	100.0%	27.1%	40.0%	23.8%	25.3%
Combined 1600 m buffer	100.0%	100.0%	40.4%	100.0%	-	49.7%	100.0%	100.0%	64.5%	49.1%	54.7%
Path area	24.3%	11.9%	21.2%	25.3%	12.1%	-	27.2%	13.4%	25.0%	15.0%	12.6%
School 800 m buffer	5.5%	4.5%	10.5%	52.0%	13.6%	16.5%	-	24.7%	8.8%	8.4%	5.7%
School 1600 m buffer	15.2%	14.3%	26.5%	57.5%	55.1%	31.1%	100.0%	-	23.2%	21.9%	14.4%
One SD ellipses	65.1%	40.3%	22.2%	40.6%	26.0%	27.3%	16.9%	13.6%	-	26.2%	30.7%
Two SD ellipses	87.7%	69.6%	52.3%	71.1%	55.9%	57.0%	54.9%	44.2%	97.8%	-	55.2%
Parish	79.0%	60.5%	18.4%	44.1%	33.5%	23.3%	9.6%	8.2%	48.5%	31.7%	-

Table 4. Mean count of GPS loggings located within each neighbourhood (n = 187). Bottom of table holds results of ANOVA for logging count in neighbourhoods.

Neighbourhood		GPS Logging Count in Neighbourhoods	
		Mean	σ
Place-based neighbourhoods			
Parish		250,302.4 (73.98%)	142,687.9
Address 800 m buffer		248,100.1 (72.93%)	143,025.2
Address 1600 m buffer		256,252.2 (76.30%)	142,999.4
School 800 m buffer		46,299.8 (17.09%)	60,025.4
School 1600 m buffer		70,197.3 (25.50%)	91,348.3
Combined 800 m buffer		281,020.8 (84.71%)	149,402.6
Combined 1600 m buffer		290,858.0 (88.82%)	148,838.4
Person-based neighbourhoods			
Convex hull		321,796.1 (100%)	152,318.5
One SD ellipses		264,880.8 (81.64%)	132,870.1
Two SD ellipses		302,087.2 (94.35%)	145,004.9
Path area		321,796.1 (100%)	152,318.5
ANOVA	*F*-test values	509.8	
	Significance level	<0.001	
Tukey's HSD test		47 of 55 pairs have significant different means	

3.3. Comparison of Exposure to Supermarkets in Neighbourhoods

The number of supermarkets located within each neighbourhood served as a measure of the exposure to supermarkets, and the results are presented in Table 5. The mean amount of supermarkets located in the neighbourhoods varies from 2.18 for the address 800 m buffer to 26.44 for convex hull. The mean amount of supermarkets in each neighbourhood type has a strong positive linear relationship with the size of the neighbourhood areas (cor. coef. = 0.80 and p = 0.003). When taking the neighbourhood area sizes into account, the neighbourhoods' school 800 m buffer and path area distinguish themselves by having significantly more supermarkets per square kilometre.

Table 5 presents the results of the one-way ANOVA. The high F-values and significance levels below 0.001 for all denote that almost no chance exists that the exposure to supermarkets are equal for all 11 neighbourhoods. The null hypothesis of no difference between supermarket exposures in neighbourhoods is rejected. Tukey's HSD tests were calculated to compare the individual pairs and the proportions of significant pairs are presented in the last row of Table 5. A distinction is made between the amount of significant pairs for the urban and rural samples for both the raw data count and supermarkets per square kilometre. The complete results of Tukey's HSD test are available in the Supplementary Material.

The one-way ANOVA and Tukey's HSD test highlighted the differences between urban and rural neighbourhoods. The results of the Welch t-test presented in Table 6 accentuate the significant difference for supermarket exposure in the urban and rural samples. Non-significant differences exist between one SD ellipses and both school buffers that are most likely the result of the schools being identical for urban and rural participants. All the place-based neighbourhoods have negative t-values, which indicate higher supermarket exposure in the urban sample. However, the *t*-values are positive for the individual-based neighbourhood types, which indicate a higher supermarket exposure in the rural sample.

Table 5. Mean exposure to supermarkets in each neighbourhood for total (n = 187), urban (n = 94), rural (n = 93) per km^2 (n = 187), per km^2 urban (n = 94) and per km^2 rural (n = 93) samples. The lower portion of the table presents results of ANOVA and Tukey's HSD for neighbourhoods.

		Supermarkets		Supermarkets Urban Areas		Supermarkets Rural Areas		Supermarkets pr. km^2		Supermarkets pr. km^2 (Urban)		Supermarkets pr. km^2 (Rural)	
Neighbourhood		Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ
Place-based neighbourhoods													
Parish		3.43	2.31	4.64	2.19	2.20	1.70	0.79	1.13	1.47	1.26	0.09	0.08
Address 800 m buffer		2.18	2.51	3.50	2.83	0.85	1.05	1.09	1.25	1.74	1.40	0.42	0.52
Address 1600 m buffer		6.01	5.87	10.14	5.55	1.83	1.87	0.74	0.72	1.24	0.68	0.22	0.23
School 800 m buffer		4.93	2.87	4.76	3.19	5.11	2.51	2.45	1.43	2.36	1.58	2.54	1..25
School 1600 m buffer		12.65	6.28	12.29	6.50	13.01	6.07	1.55	0.77	1.51	0.79	1.59	0.74
Combined 800 m buffer		6.79	3.75	7.61	4.38	5.96	2.75	1.75	1.00	2.02	1.17	1.48	0.68
Combined 1600 m buffer		16.70	7.61	18.61	8.36	14.75	6.23	1.12	0.55	1.33	0.59	0.90	0.38
Person-based neighbourhoods													
Convex hull		26.44	15.97	22.22	15.69	30.70	15.17	1.32	1.16	1.90	1.27	0.73	0.62
One SD ellipses		6.36	8.32	5.88	7.00	6.85	9.49	1.18	1.84	1.90	2.29	0.44	0.67
Two SD ellipses		20.04	18.75	16.15	14.07	23.97	21.89	0.97	1.05	1.47	1.22	0.44	0.41
Path area		11.44	6.25	10.41	6.30	12.47	6.06	2.82	1.45	3.39	1.53	2.23	1.09
ANOVA	*F*-test values	137.8		55.35		100.6		60.49		19.51		134.3	
	Sig. level	<0.001		<0.001		<0.001		<0.001		<0.001		<0.001	
Tukey's HSD test		44 of 55 pairs have significant different means		39 of 55 pairs have significant different means		42 of 55 pairs have significant different means		33 of 55 pairs have significant different means		19 of 55 pairs have significant different means		41 of 55 pairs have significant different means	

Comparing supermarket exposure per square kilometre in urban and rural neighbourhoods resulted in significant differences for all neighbourhood types except for the two school buffers as indicated in the previous *t*-test in Table 6. The remaining *t*-values are all negative indicating a higher supermarket exposure per square kilometre in the urban sample.

Table 6. The results of the Welch *t*-test comparing urban and rural neighbourhood exposure to supermarkets.

Neighbourhood	*t*	df	Sig. (Two-Sided)	95% conf. Interval of the Differences	
				Lower	Upper
Place-based neighbourhoods					
Parish	−8.489	175.316	** <0.001	−3.000	−1.868
Address 800 m buffer	−8.512	118.519	** <0.001	−3.267	−2.034
Address 1600 m buffer	−13.749	114.016	** <0.001	−9.508	−7.113
School 800 m buffer	0.839	176.145	0.403	−0.476	1.181
School 1600 m buffer	0.787	184.374	0.432	−1.091	2.538
Combined 800 m buffer	−3.086	156.853	** 0.002	−2.705	−0.594
Combined 1600 m buffer	−3.567	171.904	** <0.001	−5.970	−1.716

Table 6. *Cont.*

Neighbourhood	*t*	df	Sig. (Two-Sided)	95% conf. Interval of the Differences Lower	Upper
Person-based neighbourhoods					
Convex hull	3.756	184.902	** <0.001	4.023	12.928
One SD ellipses	0.792	169.149	0.430	−1.443	3.376
Two SD ellipses	2.902	156.711	** 0.004	2.497	13.141
Path area	2.277	184.838	* 0.024	0.275	3.842

Notes: * statistically significant below the 0.05 level, ** Statistically significant below the 0.01 level.

3.4. Comparison of Exposure to Fast Food Outlets in Neighbourhoods

Table 7 presents the results of fast food exposure in neighbourhoods. The mean amount of fast food outlets that are located within each neighbourhood vary from 3.81 for the place-based neighbourhoods to 46.92 fast food outlets for the person-based neighbourhoods. More fast food outlets per square kilometre are located near the schools than in other locations.

Table 7. Mean exposure to fast food outlets in each neighbourhood for total ($n = 187$), urban ($n = 94$), rural ($n = 93$) per km^2 ($n = 187$), per km^2 urban ($n = 94$) and per km^2 rural ($n = 93$) samples. The lower portion of the table presents results of ANOVA and Tukey's HSD for neighbourhoods.

		Fast Food Outlets		Fast Food Outlets Urban Areas		Fast Food Outlets Rural Areas		Fast Food Outlets pr. km^2		Fast Food Outlets pr. km^2 (Urban)		Fast Food Outlets pr. km^2 (Rural)	
Neighbourhood		Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ	Mean	σ
Place-based neighbourhoods													
Parish		4.06	4.16	6.44	4.56	1.67	1.54	1.86	4.76	3.603	6.260	0.091	0.105
Address 800 m buffer		3.81	6.51	6.85	8.06	0.74	1.05	1.90	3.24	3.407	4.007	0.369	0.523
Address 1600 m buffer		9.79	13.71	17.81	15.45	1.68	2.54	1.20	1.68	2.188	1.898	0.206	0.311
School 800 m buffer		13.71	11.99	13.27	12.36	14.15	11.65	6.82	5.96	6.597	6.146	7.037	5.799
School 1600 m buffer		26.99	21.41	26.06	21.94	27.92	20.93	3.32	2.63	3.203	2.696	3.431	2.572
Combined 800 m buffer		16.47	12.97	18.03	14.06	14.89	11.64	4.28	3.47	4.842	3.901	3.703	2.893
Combined 1600 m buffer		33.07	22.27	36.54	23.02	29.57	21.02	2.24	1.59	2.649	1.753	1.820	1.291
Person-based neighbourhoods													
Convex hull		46.92	25.11	42.44	25.08	51.45	24.44	2.87	3.60	4.399	4.416	1.329	1.323
One SD ellipses		11.30	16.97	11.74	15.75	10.86	18.20	2.23	3.84	3.687	4.770	0.751	1.573
Two SD ellipses		34.98	31.61	31.10	26.11	38.91	36.05	2.12	4.10	3.499	5.375	0.729	0.883
Path area		24.29	12.65	23.60	13.18	24.99	12.12	6.45	4.53	8.213	5.198	4.658	2.797
ANOVA	*F*-test values	110.7		42.69		78.45		46.06		14.78		80.82	
	Sig. level	<0.001		<0.001		<0.001		<0.001		<0.001		<0.001	
Tukey's HSD test		45 of 55 pairs have significant different means		35 of 55 pairs have significant different means		47 of 55 pairs have significant different means		29 of 55 pairs have significant different means		20 of 55 pairs have significant different means		36 of 55 pairs have significant different means	

The results of the one-way ANOVA for fast food exposure are presented in Table 7. The *F*-values varies from 14.78 to 110.7 and is a hint of how many pairs of neighbourhoods have significantly different mean fast food exposure. Significance levels are below 0.001 for all analysis of variance, indicating that almost no chance exists that the exposure to fast food outlets are equal for all 11 neighbourhoods in any of the six analyses of variance. Tukey's HSD tests were calculated to compare the individual pairs and the proportions of significant pairs are presented in the last row of Table 7. Fewer significantly different pairs of neighbourhoods are found to experience fast food exposure in urban areas than in the rural sample. The complete results of Tukey's HSD test are available in the Supplementary Material.

The results of Tukey's HSD test were significantly different between the fast food outlet exposure in rural neighbourhoods (47/36 of 55) and some of the urban neighbourhoods (35/20 of 55). The Welch *t*-test compares the fast food exposure in the urban and rural neighbourhoods. The results of the *t*-tests are presented in Table 8. Significant differences exist between the mean exposures to fast food outlets for the home-based neighbourhood's parish, address 800 and 1600 m buffer. For all three, the *t*-values are negative denoting a higher exposure in the urban sample.

Table 8. The results of the Welch *t*-test for comparing fast food outlet exposure in urban and rural neighbourhoods.

Neighbourhood	*t*	df	Sig. (Two-Sided)	95% Conf. Interval of the Differences	
				Lower	Upper
Place-based neighbourhoods					
Parish	−9.598	114.176	** <0.001	−5.754	−3.785
Address 800 m buffer	−7.288	96.203	** <0.001	−7.772	4.445
Address 1600 m buffer	−9.988	98.067	** <0.001	−19.336	−12.926
School 800 m buffer	0.503	184.587	0.615	−2.581	4.350
School 1600 m buffer	0.593	184.756	0.554	−4.326	8.048
Combined 800 m buffer	−1.664	179.407	0.097	−6.862	0.583
Combined 1600 m buffer	−2.163	183.83	* 0.032	−13.332	−0.613
Person-based neighbourhoods					
Convex hull	2.489	184.958	* 0.014	1.871	16.160
One SD ellipses	−0.355	180.686	0.723	−5.798	4.029
Two SD ellipses	1.697	167.599	0.092	−1.277	16.913
Path area	0.753	184.021	0.452	−2.258	5.045

Notes: * Statistically significant below the 0.05 level, ** statistically significant below the 0.01 level.

The results of the Welch *t*-test for comparing fast food exposure per square kilometre in urban and rural neighbourhoods resulted in significant differences for all neighbourhoods except both school buffers. The *t*-values are all negative, which indicate a higher fast food exposure per square kilometre in the urban sample.

4. Discussion

4.1. Place Based vs. *People Based Neighbourhood Definitions*

The understanding of place as a concept stretches from the individual adhering to their own unique place determined by their everyday life and behaviour to the claim that the individual unconsciously relates their behaviour and choices to more structured patterns based on social and physical environment characteristics [2]. However, often the discussion about place is ignored due to pragmatic considerations, such as data only being accessible in administrative units. Administrative divisions as the concept for place are therefore often the natural choice for many researchers without considering the administrative divisions' ability to encapsulate the relevant behaviour. The consequence is a wrong assumption or generalisation that all individuals have equal behaviour patterns, limiting the exposure to a confined area and limiting diversity in food supply choices.

This study reveals that the administrative divisions are not a suitable neighbourhood type to capture the measured behaviour. This finding is supported by the fact that only 12.8% (24 of 187) of the participants attend school in their residential parish, and the exposures to supermarkets and fast food outlets around the schools are more than three and six times higher, respectively, than in the parishes. This fact coincides with previous studies that found similar relationships between exposures near home and school [11,15,17]. However, the differences between home and school neighbourhoods are significantly more distinctive for participants living in a rural area and attending schools in urban areas.

The place-based neighbourhood definitions do not take into account the diversity in individual behaviour. This problem is most likely the result of assuming people carry out most of their activities in their residential location, which is contradicted by the high mobility in the participant sample. The participants in this study are young adults, and most have a high mobility level even without the ability to drive a car. The participant's mobility must be taken into account because it weakens the influence of residential neighbourhoods. However, other studies with low mobility group samples, such as the elderly and the disadvantaged people, are probably more sensitive to the residential neighbourhood exposure [20].

The use of the term neighbourhood in food environment research adheres to spaces defined by fixed boundaries, such as administrative units, or a fixed distance, such as buffers, that define a school or residential neighbourhood [4]. When referring to individual-measured areas, a more appropriate term instead of neighbourhood is "activity spaces" as suggested by Zenk and colleagues [39]. This division between terms can potentially improve researchers' understanding of the differences between the place-based and person-based exposure measures.

Defining individual activity spaces is advantageous for providing increased specificity in a multiple space exposure measurement. However, as Ball and colleagues note, the collection of activity space attribute data can be time and labour intensive because the individual activity spaces do not align spatially with existing administrative divisions [1]. The activity spaces defined by the individual's behaviour most likely vary in area size, which increases the complexity of analysis when comparing different individuals' exposure. Moreover, comparisons across different studies are very difficult if the activity spaces vary in area size. The equal size of neighbourhoods based on buffers makes them easier to compare between studies in different countries. However, the buffers are limited to a few locations, and as this study reveals, the buffers and the administrative divisions have similar problems in capturing exposure during commuting or leisure time activities. The researcher's perception is that the use of multiple-location buffers provides a much better basis for measuring exposure than single-area buffers and administrative divisions. Applying buffers on either home or school only provides one piece in the complex puzzle of measuring the complete exposure. Many studies have limited the research area to a residential/school neighbourhood (for example, a 1 km buffer) [5,9,13,16,24,27–33] or administratively defined boundaries [22–26]. The studies thereby only consider data inside the sample area of interest. Data in adjacent areas are not implemented, which could be problematic because the effect of exposure across study boundaries is not considered. Another problem with the buffer areas created is how to define a relevant distance since found associations may vary depending on this definition [4]. To bypass these problems, researchers should consider measuring actual activity spaces, which is possible using GPS.

4.2. Implications for Research

The neighbourhoods' ability to capture the activity measured by GPS varies, particularly for those neighbourhood types that are confined to one or two locations and enclose a smaller percentage of the measured activity. The parishes are typically more than eight times larger in area than the address 800 m buffer and two times the 1600 m buffer, but they enclose only 1% more and 2.5% less, respectively, of the measured activity. This finding indicates that most activity around the residential locations is tied very closely (within 800 m) to the home, whereas an enlargement of the residential neighbourhood to a 1600 m buffer or a parish has little effect on capturing more of the measured activity. Approximately 85% of the measured activities are near the home or school, but the final 15% poses a challenge for researchers to measure because it constitutes the behaviours that are most affected by individual preferences.

Individual characteristics as confounders are crucial to take into account personal preferences when analysing relationships between the food environment and health outcomes [1,2]. However, not all preferences can be adjusted through common confounders such as income, ethnicity and education level. Consequently, methods used for defining neighbourhoods must accommodate the individual behavioural characteristics [20]. However, to achieve this effect, researchers must carefully scrutinise the behaviour to be measured to fully understand the phenomenon. The way a space is defined should reflect the context in which it is applied [14]. Therefore, to measure the exposure to food environment, researchers must make qualified assumptions about where people shop, the distance they are willing to travel to shop and other individual preferences [2]. Thus, paying attention to the individual is important when developing studies of the interaction between the population and the environment. As Larson and Story concluded, most food environment studies have methodological problems that reduce the credibility of their findings [40]. Problems occur with assessing the physical access to food sources in the environment [4] and linking access to a food source with food purchases and intake. Further analysis of individual behaviour could potentially be used to link the food source exposure to individual food purchasing through analysing movement and stop flows in space-time data.

The results of this study are consistent with several other studies [1,2,15,20,21] advocating for more individual-based neighbourhood definitions taking into account multiple environments for exposure beyond home, school or work communities. Exposure during commuting time and leisure activities are particularly difficult to incorporate when the neighbourhoods are place based. Kwan further questions the use of arbitrary definitions of neighbourhoods instead of considering the actual spaces in which individuals' exposure occur [41]. The main objections to the static and administrative bounded spatial definitions in ecological exposure measures found in this study and accentuated by Kwan are: (1) the assumption that the residential neighbourhoods are the most relevant in affecting food exposure; and (2) individuals who live in the same spatial areas experience the same level of exposure, regardless of time spent in the area and residential locations within the area [41]. The results from this study contradict the assumptions since individuals also spend a substantial time outside their residential neighbourhood, and the variance of individual activity space sizes illustrates the variety in individuals' exposure.

Comparisons between urban and rural samples (*t*-tests) clearly reveal differences in exposure to supermarkets and fast food outlets in some neighbourhoods. Tukey's HSD test similarly reveals that more neighbourhood types are significantly different in the rural sample than in the urban sample. Hence, a separation between urban and rural samples would create more homogenous samples. Individual activity spaces will vary depending on factors such as income, personal mobility (ability to drive, access to a vehicle, walking disabilities, *etc.*), age and other individual preferences. People living in rural areas are more likely to travel to a more populated area because these areas often provide greater access to work opportunities, food or cultural events, for example. On the other hand, urban residents are less likely to commute to rural areas, as their needs are mostly satisfied in the cities. The daily activity spaces of rural residents are presumably larger if they have no restrictions on their movement or travel abilities.

4.3. Limitations

The activity data measured by GPS clearly indicates that participants are using multiple locations and are thereby not restricted to their immediate residential environments. The survey period of one week is a short time frame for analysis of the participants' behaviour. Short tracking periods could include locations, which might be visited infrequently and vice versa [20]. This phenomenon is the shortcoming of GPS technologies because recording consecutive involvement at such a level for longer periods is difficult. The development of tracking technologies is a fast growing field, and technologies such as Bluetooth, Wi-Fi and cellular phone networks could potentially be used to track participants in a way that requires less involvement from the individuals [42], mostly because all these technologies are included in most mobile phones today and therefore do not require participants to carry and maintain additional devices. The development of these technologies provides a promising improvement for empirical place research [21].

This study used GPS devices set to measure at seven-second intervals, which was the minimum interval available between loggings. A seven-second interval between registrations is a short time and discharges the battery faster than at a higher interval. A low interval between registrations is preferable for some uses, but the logging interval could probably be 15 s or more to measure the extent of the activity spaces. However, some problems occur with a high registration frequency. Activity measured by GPS can experience periods with loss of data that interferes with the registration interval. Activity space measures as standard deviational ellipses are calculated from the centre of gravity of the measured point locations and uneven intervals between registrations therefore affect the extent of the calculated spaces. Several methods have been proposed for resolving this issue by estimating missing data [39] or interpolation between registrations. Further, studies' ability to measure individuals' use of food retailers is dependent on a low interval between registrations. To detect stops at food retailers, several consecutive registrations at the same location are needed. Determining a maximum interval between registrations is difficult without further research, but a large interval between registrations results in a smaller dataset that is easier to analyse. Studies that apply GPS to measure activity must consider the accuracy required (interval between registrations) and the expected travel types and speed of participants.

The individual based neighbourhoods are better at capturing multiple space activity, but the measured exposure could be an exaggeration, which could be the case for the convex hull and two SD ellipses when compared to path area. The neighbourhood type convex hull has a large mean area size, particularly for the rural samples. Comparing convex hull with path area, which are both based on GPS tracking, reveals a 25% higher supermarket exposure for the convex hull neighbourhood type. However, if the area sizes for both neighbourhood types are used to adjust the exposure, then the exposure in path area is twice that of the convex hull. Path area is more focused on where the actual activity has occurred, but it does not capture deviant activities that would happen at other times than the single week when the activity was tracked. Therefore, whether the path area may underestimate the exposure remains unclear. To answer this question, researchers must delve into the understanding of people's behaviour. Second, studying the relationship between measured exposure and the actual choices of food buying is relevant because this research could broaden the insight to defining a proper neighbourhood for measuring exposure to food outlets.

Any study of this type must use the appropriate spatial area to measure the exposure. However, many studies have applied place-based neighbourhoods with little focus on identifying these areas [41]. Among the most discussed methodological issues in research applying spatial data is the Modifiable Area Unit Problem (MAUP). MAUP refers to the issue that the areal units to which data are assigned might influence results. Neighbourhoods based on administrative divisions or buffers are highly susceptible to the MAUP. The place-based neighbourhoods allow little variation between individuals compared to the person-based neighbourhoods (Table 1). Large differences exist between individual activity spaces such as the convex hull and standard deviational ellipses where the standard deviation for each type of activity space is larger than the mean area size. This finding clearly indicates a large spread between individual activity spaces. Considering the actual spatial and temporal exposure would allow for a more accurate measure of exposure and address the MAUP [41]. This result would allow individuals to have individual exposure measures although they live in the same neighbourhood.

5. Conclusions

This study presents significant differences between the exposure to supermarkets and fast food outlets for different neighbourhood types. Second, significant differences were found for exposure to food outlets between urban and rural neighbourhoods.

Neighbourhoods are a fuzzy concept that varies in meaning depending on the conducted study and on each person's individual perception of their neighbourhood. Complexity and heterogeneity of human mobility no longer appear to correspond to the use of residential neighbourhoods but stress the need for methods and measures of individual activity and exposure. Exposure to the food environment occurs in multiple environments, but measuring individuals' activity spaces in multiple environments is challenging. The lack of focus on neighbourhood or activity space definitions in studies of the food environment is unfortunate, mainly due to the large amount of research analysing relationships between the food environment and health outcomes in which no evidence demonstrates that the neighbourhood exposures used coincide with the actual exposure. Tracking technologies can provide space-time data on the behaviour of individuals, and these data can be used to define neighbourhoods for measuring exposure to the food environment.

Author Contributions

Anders K. Lyseen and Henning S. Hansen wrote the paper and Bent E. Mikkelsen revised the paper. Anders K. Lyseen, Henrik Harder and Anders S. Jensen conceived and designed the experiments. Anders K. Lyseen analysed and interpreted the data.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Ball, K.; Timperio, A.F.; Crawford, D.A. Understanding Environmental Influences on Nutrition and Physical Activity Behaviors: Where Should We Look and What Should We Count? *Int. J. Behav. Nutr. Phys. Act.* **2006**, *3*, 1–8.
2. Lytle, L.A. Measuring the Food Environment: State of the Science. *Am. J. Prev. Med.* **2009**, *36*, 134–144.
3. Thornton, L.E.; Pearce, J.R.; Macdonald, L.; Lamb, K.E.; 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*, doi:10.1186/1476-072X-11-29.
4. Lucan, S.C. 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**, *2*, 205–212.
5. Leung, C.W.; Laraia, B.A.; Kelly, M.; Nickleach, D.; Adler, N.E.; Kushi, L.H.; Yen, I.H. The Influence of Neighborhood Food Stores on Change in Young Girls' Body Mass Index. *Am. J. Prev. Med.* **2011**, *41*, 43–51.
6. Rossen, L.M.; Curriero, F.C.; Cooley-Strickland, M.; Pollack, K.M. Food Availability en Route to School and Anthropometric Change in Urban Children. *J. Urban Health* **2013**, *90*, 653–666.
7. Oreskovic, N.M.A.; Winickoff, J.P.B.; Kuhlthau, K.A.C.; Romm, D.C.; Perrin, J.M.B. Obesity and the Built Environment among Massachusetts Children. *Clin. Pediatr.* **2009**, *48*, 904–912.
8. Wang, M.C.; Kim, S.; Gonzalez, A.A.; MacLeod, K.E.: Winkleby, M.A. Socioeconomic and Food-Related Physical Characteristics of the Neighbourhood Environment are Associated with Body Mass Index. *J. Epidemiol. Community Health* **2007**, *61*, 491–498.
9. Burgoine, T.; Forouhi, N.G.; Griffin, S.J.; Wareham, N.J.; Monsivais, P. Associations between Exposure to Takeaway Food Outlets, Takeaway Food Consumption, and Body Weight in Cambridgeshire, UK: Population Based, Cross Sectional Study. *BMJ* **2014**, *348*, 1464.

10. Pearce, J.A.; Hiscock, R.A.; Blakely, T.B.; Witten, K.C. A National Study of the Association between Neighbourhood Access to Fast-Food Outlets and the Diet and Weight of Local Residents. *Health Place* **2009**, *15*, 193–197.
11. Jeffery, R.W.; Baxter, J.; McGuire, M.; Linde, J. Are Fast Food Restaurants an Environmental Risk Factor for Obesity? *Int. J. Behav. Nutr. Phys. Act.* **2006**, *3*, doi:10.1186/1479-5868-3-2.
12. Liu, G.C.; Wilson, J.S.B.; Qi, R.C.; Ying, J.D. Green Neighborhoods, Food Retail and Childhood Overweight: Differences by Population Density. *Am. J. Health Promot.* **2007**, *21*, 317–325.
13. Laxer, R.E.; Janssen, I. The Proportion of Excessive Fast-Food Consumption Attributable to the Neighbourhood Food Environment among Youth Living within 1 km of Their School. *Appl. Physiol. Nutr. Metabol.* **2013**, *39*, 480–486.
14. Kwan, M.P. The Uncertain Geographic Context Problem. *Ann. Assoc. Am. Geogr.* **2012**, *5*, 958–968.
15. 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*, 1–11.
16. Giles-Corti, B.; Timperio, A.; Bull, F.; Pikora, T. Understanding Physical Activity Environmental Correlates: Increased Specificity for Ecological Models. *Exerc. Sport Sci. Rev.* **2005**, *33*, 175–181.
17. Hurvitz, P.M.; Moudon, A.V. Home *versus* Nonhome Neighborhood: Quantifying Differences in Exposure to the Built Environment. *Am. J. Prev. Med.* **2012**, *42*, 411–417.
18. Odoms-Young, A.M.; Zenk, S.; Mason, M. Measuring Food Availability and Access in African-American Communities: Implications for Intervention and Policy. *Am. J. Prev. Med.* **2009**, *36*, 145–150.
19. Kerr, J.; Duncan, S.; Schipperjin, J. Using Global Positioning Systems in Health Research: A Practical Approach to Data Collection and Processing. *Am. J. Prev. Med.* **2011**, *41*, 532–540.
20. 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.
21. Rainham, D.; McDowell, I.; Krewski, D.; Sawada, M. Conceptualizing the Healthscape: Contributions of Time Geography, Location Technologies and Spatial Ecology to Place and Health Research. *Soc. Sci. Med.* **2010**, *70*, 668–676.
22. Powell, L.M.; Han, E.; Zenk, S.N.; Khan, T.; Quinn, C.M.; Gibbs, K.P.; Pugach, O. Field Validation of Secondary Commercial Data Sources on the Retail Food Outlet Environment in the U.S. *Health Place* **2011**, *17*, 1122–1131.
23. Apparicio, P.; Cloutier, M.; Shearmur, R. The Case of Montréal's Missing Food Deserts: Evaluation of Accessibility to Food Supermarkets. *Int. J. Health Geogr.* **2007**, *6*, 1–13.
24. Block, J.P.; Scribner, R.A.; DeSalvo, K.B. Fast Food, Race/Ethnicity, and Income: A Geographic Analysis. *Am. J. Prev. Med.* **2004**, *27*, 211–217.
25. Duran, A.C.; Diez Roux, A.V.; Latorre, M.D.R.; Jaime, P.C. Neighborhood Socioeconomic Characteristics and Differences in the Availability of Healthy Food Stores and Restaurants in Sao Paulo, Brazil. *Health Place* **2013**, *23*, 39–47.
26. Moore, L.V.; Diez Roux, A.V. Associations of Neighborhood Characteristics with the Location and Type of Food Stores. *Am. J. Public Health* **2006**, *96*, 325–331.
27. Chen, X.; Clark, J. Interactive Three-Dimensional Geovisualization of Space-Time Access to Food. *Appl. Geogr.* **2013**, *43*, 81–86.

28. Austin, S.B.; Melly S.J.; Sanchez, B.N.; Patel, A.; Buka, S.; Gortmaker, S.L. 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*, 1575–1581.
29. Kestens, Y.; Daniel, M. Social Inequalities in Food Exposure around Schools in an Urban Area. *Am. J. Prev. Med.* **2010**, *39*, 33–40.
30. Nash, D.M.; Gilliland, J.A.; Evers, S.E.; Wilk, P.; Campbell, M.K. Determinants of Diet Quality in Pregnancy: Sociodemographic, Pregnancy-Specific, and Food Environment Influences. *J. Nutr. Educ. Behav.* **2013**, *45*, 627–634.
31. Neckerman, K.M.; Bader, M.D.M.; Richards, C.A.; Purciel, M.; Quinn, J.W.; Thomas, J.S.; Warbelow, C.; Weiss, C.C.; Lovasi, G.S.; Rundle, A. Disparities in the Food Environments of New York City Public Schools. *Am. J. Prev. Med.* **2010**, *39*, 195–202.
32. Seliske, L.M.; Pickett, W.; Boyce, W.F.; Janssen, I. Density and Type of Food Retailers Surrounding Canadian Schools: Variations across Socioeconomic Status. *Health Place* **2009**, *15*, 903–907.
33. Sturm, R. Disparities in the Food Environment Surrounding US Middle and High Schools. *Public Health* **2008**, *122*, 681–690.
34. Newsome, T.H.; Walcott, W.A.; Smith, P.D. Urban Activity Spaces: Illustrations and Application of a Conceptual Model for Integrating the Time and Space Dimensions. *Transportation* **1998**, *25*, 357–377.
35. Qi, F.; Du, F. Trajectory Data Analyses for Pedestrian Space-Time Activity Study. *J. Vis. Exp.* **2013**, *72*, doi:10.3791/50130.
36. Smith, G.; Gidlow, C.; Davey, R.; Foster, C. What Is My Walking Neighbourhood? A Pilot Study of English Adults' Definitions of Their Local Walking Neighbourhoods. *Int. J. Behav. Nutr. Act.* **2010**, *7*, 1–8.
37. Lyseen, A.K.; Hansen, H.S. Spatial and Semantic Validation of Secondary Food Source Data. *ISPRS Int. J. Geo-Inf.* **2014**, *3*, 236–253.
38. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. 2014. Available online: http://www.R-project.org/ (accessed on 19 February 2015).
39. Zenk, S.N.; Schulz, A.J.; Matthews, S.A.; Odoms-Young, A.; Wilbur, J.; Wegrzyn, L.; Gibbs, K.; Braunschweig, C.; Stokes, C. Activity Space Environment and Dietary and Physical Activity Behaviors: A Pilot Study. *Health Place* **2011**, *17*, 1150–1161.
40. Larson, N.; Story, M. A Review of Environmental Influences on Food Choices. *Ann. Behav. Med.* **2009**, *38*, 56–73.
41. Kwan, M.P. From Place-Based to People-Based Exposure Measures. *Soc. Sci. Med.* **2009**, *9*, 1311–1313.
42. Mikkelsen, B.E.; Lyseen, A.K.; Dobroczynski, M.; Hansen, H.S. Behavioural Nutrition & Big Data: How Geodata, Register Data & GPS, Mobile Positioning, Wi-Fi, Bluetooth & Thermal Cameras Can Contribute to the Study of Human Food Behaviour. In Proceedings of the Measuring Behavior 2014, Wageningen, The Netherlands, 27–29 August 2014.

APPENDIX V – SUPPLEMENTARY MATERIAL – PAPER IV: DEFINING NEIGHBOURHOODS AS A MEASURE OF EXPOSURE TO THE FOOD ENVIRONMENT

Supplementary Information

Defining Neighbourhoods as a Measure of Exposure to the Food Environment

Table S1. Results Tukey's HSD test for comparison of area size.

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Parish	**	Address buffer 800 m	15.797	0.005	2.713	28.861
		Address buffer 1 mile	9.661	0.379	−3.413	22.735
	**	School buffer 800 m	15.787	0.005	2.713	28.861
		School buffer 1 mile	9.661	0.379	−3.413	22.735
	*	Address & school buffer 800 m	13.886	0.026	0.812	26.961
		Address & school buffer 1 mile	2.688	1.000	−10.386	15.762
	**	Convex hull	−33.337	<0.000	−46.411	−20.262
		1 standard deviational ellipses	0.020	1.000	−13.054	13.094
	**	2 standard deviational ellipses	−33.592	<0.000	−46.667	−20.518
		Path area	13.042	0.051	−0.032	26.116
Address buffer 800 m		Address buffer 1 mile	−6.126	0.916	−19.200	6.948
		School buffer 800 m			Exactly same area size	
		School buffer 1 mile	−6.126	0.917	−19.200	6.948
		Address & school buffer 800 m	−1.900	1.000	−14.974	11.174
	*	Address & school buffer 1 mile	−13.098	0.049	−26.173	−0.024
	**	Convex hull	−49.123	<0.000	−62.197	−36.049
	**	1 standard deviational ellipses	−15.766	0.005	−28.840	−2.692
	**	2 standard deviational ellipses	49.379	<0.000	−62.453	−36.305
		Path area	−2.745	1.000	−15.819	10.329
Address buffer 1 mile		School buffer 800 m	6.126	0.917	−6.948	19.200
		School buffer 1 mile			Exactly same area size	
		Address & school buffer 800 m	4.226	0.994	−8.848	17.300
		Address & school buffer 1 mile	−6.972	0.826	−20.045	6.101
	**	Convex hull	−42.997	<0.000	−56.071	−29.923
		1 standard deviational ellipses	−9.640	0.383	−22.714	3.434
	**	2 standard deviational ellipses	−42.253	<0.000	−56.327	−30.179
		Path area	3.381	0.999	−9.693	16.455
School buffer 800 m		School buffer 1 mile	−6.126	0.917	−19.200	6.948
		Address & school buffer 800 m	−1.900	1.000	−14.974	11.174
	*	Address & school buffer 1 mile	−13.098	0.049	−26.173	−0.024
	**	Convex hull	−49.123	<0.000	−62.197	−36.049
	**	1 standard deviational ellipses	−15.766	0.005	−28.840	−2.692
	**	2 standard deviational ellipses	−49.379	<0.000	−62.453	−36.305
		Path area	−2.745	1.000	−15.819	10.329

Table S1. *Cont.*

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
School buffer 1 mile		Address & school buffer 800 m	4.226	0.994	−8.848	17.300
		Address & school buffer 1 mile	−6.972	0.826	−20.046	6.102
	**	Convex hull	−42.997	<0.000	−56.071	−29.923
		1 standard deviational ellipses	−9.640	0.383	−22.714	3.434
	**	2 standard deviational ellipses	−43.253	<0.000	−56.327	−30.179
		Path area	3.381	0.999	−9.693	16.455
Address & school buffer 800 m		Address & school buffer 1 mile	−11.198	0.174	−24.272	1.876
	**	Convex hull	−47.223	<0.000	−60.297	−34.149
	*	1 standard deviational ellipses	−13.866	0.027	−26.940	−0.792
	**	2 standard deviational ellipses	−47.479	<0.000	−60.553	−34.405
		Path area	−0.845	1.000	−13.919	12.230
Address & school buffer 1 mile	**	Convex hull	−36.025	<0.000	−49.099	−22.951
		1 standard deviational ellipses	−2.668	1.000	−15.742	10.406
	**	2 standard deviational ellipses	−36.281	<0.000	−49.355	−23.207
		Path area	10.354	0.275	−2.720	23.428
Convex hull	**	1 standard deviational ellipses	33.357	<0.000	20.283	46.431
		2 standard deviational ellipses	−0.256	1.000	−13.330	12.818
	**	Path area	46.379	<0.000	33.304	59.453
1 standard deviational ellipses	**	2 standard deviational ellipses	−33.613	<0.000	−46.687	−20.539
		Path area	13.022	0.052	−0.053	26.096
2 standard deviational ellipses	**	Path area	46.634	<0.000	33.560	59.709

Notes: * Statistically significant below the 0.05 level, ** Statistically significant below the 0.01 level.

Table S2. Results Tukey's HSD test for comparison of area size in urban sample.

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Parish		Address buffer 800 m	3.691	0.644	−2.240	9.621
		Address buffer 1 mile	−2.435	0.965	−8.366	3.495
		School buffer 800 m	3.691	0.644	−2.240	9.621
		School buffer 1 mile	−2.435	0.965	−8.366	3.495
		Address & school buffer 800 m	1.900	0.994	−4.031	7.830
	**	Address & school buffer 1 mile	−8.291	<0.000	−14.222	−2.360
	**	Convex hull	−15.439	<0.000	−21.370	−9.509
		1 standard deviational ellipses	1.176	1.000	−4.754	7.107
	**	2 standard deviational ellipses	−1.099	<0.000	−16.920	−5.059
		Path area	2.252	0.980	−3.679	8.182
Address buffer 800 m	*	Address buffer 1 mile	−6.126	0.036	−12.057	−0.195
		School buffer 800 m	Exactly same area size			
	*	School buffer 1 mile	−6.126	0.036	−12.057	−0.195
		Address & school buffer 800 m	−1.791	0.997	−7.722	4.140
	**	Address & school buffer 1 mile	−11.982	<0.000	−17.912	−6.051
	**	Convex hull	−19.130	<0.000	−25.061	−13.199
		1 standard deviational ellipses	−2.515	0.956	−8.445	3.416
	**	2 standard deviational ellipses	−14.680	<0.000	−20.611	−8.749
		Path area	−1.439	0.999	−7.370	4.491

Table S2. *Cont.*

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Address buffer 1 mile	*	School buffer 800 m	6.126	0.036	0.195	12.057
		School buffer 1 mile			Exactly same area size	
		Address & school buffer 800 m	4.335	0.395	−1.596	10.266
		Address & school buffer 1 mile	−5.856	0.056	−11.786	0.075
	**	Convex hull	−13.004	<0.000	−18.934	−7.073
		1 standard deviational ellipses	3.612	0.674	−2.319	9.542
	**	2 standard deviational ellipses	−8.554	<0.000	−14.485	−2.623
		Path area	4.687	0.277	−1.244	10.618
School buffer 800 m	*	School buffer 1 mile	−6.126	0.036	−12.057	−0.195
		Address & school buffer 800 m	−1.791	0.997	−7.722	4.140
	**	Address & school buffer 1 mile	−11.982	<0.000	−17.912	−6.051
	**	Convex hull	−19.130	<0.000	−25.061	−13.199
		1 standard deviational ellipses	−2.515	0.956	−8.445	3.416
	**	2 standard deviational ellipses	−14.680	<0.000	−20.611	−8.749
		Path area	−1.439	0.999	−7.370	4.491
School buffer 1 mile		Address & school buffer 800 m	4.335	0.395	−1.596	10.266
		Address & school buffer 1 mile	−5.856	0.056	−11.786	0.075
	**	Convex hull	−13.004	<0.000	−18.934	−7.073
		1 standard deviational ellipses	3.612	0.674	−2.319	9.542
	**	2 standard deviational ellipses	−8.554	<0.000	−14.485	−2.623
		Path area	4.687	0.277	−1.244	10.618
Address & school buffer 800 m	**	Address & school buffer 1 mile	−10.191	<0.000	−16.121	−4.260
	**	Convex hull	−17.339	<0.000	−23.269	−11.408
		1 standard deviational ellipses	−0.723	1.000	−6.654	5.207
	**	2 standard deviational ellipses	−12.889	<0.000	−18.820	−6.958
		Path area	0.352	1.000	−5.579	6.283
Address & school buffer 1 mile	**	Convex hull	−7.148	0.005	−13.079	−1.218
	**	1 standard deviational ellipses	9.467	<0.000	3.537	15.398
		2 standard deviational ellipses	−2.698	0.930	−8.629	3.232
	**	Path area	10.543	<0.000	4.612	16.473
Convex hull	**	1 standard deviational ellipses	16.615	<0.000	10.685	22.546
		2 standard deviational ellipses	4.450	0.354	−1.481	10.380
	**	Path area	17.691	<0.000	11.760	23.621
1 standard deviational ellipses	**	2 standard deviational ellipses	−12.166	<0.000	−18.096	−6.235
		Path area	1.075	1.000	−4.855	7.006
2 standard deviational ellipses	**	Path area	13.241	<0.000	7.310	19.172

Notes: * Statistically significant below the 0.05 level, ** Statistically significant below the 0.01 level.

Table S3. Results Tukey's HSD test for comparison of area size in rural sample.

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Parish	**	Address buffer 800 m	28.013	0.007	4.369	51.657
		Address buffer 1 mile	21.887	0.099	−1.757	45.531
	**	School buffer 800 m	28.013	0.007	4.369	51.657
		School buffer 1 mile	21.887	0.099	−1.757	45.531
	*	Address & school buffer 800 m	26.002	0.018	2.358	49.646
		Address & school buffer 1 mile	13.786	0.730	−9.858	37.430
	**	Convex hull	−51.426	<0.000	−75.071	−27.782
		1 standard deviational ellipses	−1.148	1.000	−24.792	22.496
	**	2 standard deviational ellipses	−5.644	<0.000	−80.083	−32.794
	*	Path area	23.948	0.044	0.304	47.592
Address buffer 800 m		Address buffer 1 mile	−6.126	0.999	−29.770	17.518
		School buffer 800 m	Exactly same area size			
		School buffer 1 mile	−6.126	0.999	−29.770	17.518
		Address & school buffer 800 m	2.011	1.000	−21.633	25.655
		Address & school buffer 1 mile	−14.227	0.690	−37.871	9.417
	**	Convex hull	−79.439	<0.000	−103.083	−55.795
	**	1 standard deviational ellipses	−29.161	0.003	−52.805	−5.517
	**	2 standard deviational ellipses	−84.451	<0.000	−108.095	−60.807
		Path area	−4.064	1.000	−27.708	19.580
Address buffer 1 mile		School buffer 800 m	6.126	0.999	−17.518	29.770
		School buffer 1 mile	Exactly same area size			
		Address & school buffer 800 m	4.115	1.000	−19.529	−27.760
		Address & school buffer 1 mile	−8.101	0.991	−31.745	15.543
	**	Convex hull	−73.331	<0.000	−96.957	−49.669
		1 standard deviational ellipses	−23.035	0.064	−46.679	0.610
	**	2 standard deviational ellipses	−78.325	<0.000	−101.969	−54.681
		Path area	2.062	1.000	−21.582	25.706
School buffer 800 m		School buffer 1 mile	−6.126	0.999	−29.770	17.518
		Address & school buffer 800 m	−2.011	1.000	−25.655	21.633
		Address & school buffer 1 mile	−14.227	0.690	−37.871	9.417
	**	Convex hull	−79.439	<0.000	−103.083	−55.795
	**	1 standard deviational ellipses	−29.161	0.004	−52.805	−5.517
	**	2 standard deviational ellipses	−84.451	<0.000	−108.095	−60.807
		Path area	−4.064	1.000	−27.708	19.580
School buffer 1 mile		Address & school buffer 800 m	4.115	1.000	−19.529	27.760
		Address & school buffer 1 mile	−8.101	0.991	−31.745	15.543
	**	Convex hull	−73.313	<0.000	−96.957	−49.669
		1 standard deviational ellipses	−23.035	0.064	−46.679	0.610
	**	2 standard deviational ellipses	−78.325	<0.000	−101.969	−54.681
		Path area	2.062	1.000	−21.582	25.706
Address & school buffer 800 m		Address & school buffer 1 mile	−12.217	0.852	−35.861	11.427
	**	Convex hull	−77.429	<0.000	−101.073	−53.785
	**	1 standard deviational ellipses	−27.150	0.010	−50.794	−3.506
	**	2 standard deviational ellipses	−82.441	<0.000	−106.085	−58.797
		Path area	−2.054	1.000	−25.698	21.590

Table S3. *Cont.*

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Address & school buffer 1 mile	**	Convex hull	−65.212	<0.000	−88.856	−41.568
		1 standard deviational ellipses	−14.933	0.622	−38.577	8.711
	**	2 standard deviational ellipses	−70.224	<0.000	−93.868	−46.580
		Path area	10.163	0.952	−13.481	33.807
Convex hull	**	1 standard deviational ellipses	50.279	<0.000	26.635	73.923
		2 standard deviational ellipses	−5.012	1.000	−28.656	18.632
	**	Path area	75.375	<0.000	51.731	99.019
1 standard deviational ellipses	**	2 standard deviational ellipses	−55.291	<0.000	−78.935	−31.647
	*	Path area	25.096	0.027	1.452	48.740
2 standard deviational ellipses	**	Path area	80.387	<0.000	56.743	104.031

Notes: * Statistically significant below the 0.05 level, ** Statistically significant below the 0.01 level.

Table S4. Results Tukey's HSD test for comparison of percent of GPS activity within neighbourhoods.

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Parish		Address buffer 800 m	1.051	0.999	−4.559	6.663
		Address buffer 1 mile	−2.322	0.963	−7.933	3.289
	**	School buffer 800 m	56.891	<0.000	51.280	62.502
	**	School buffer 1 mile	48.484	<0.000	42.872	54.095
	**	Address & school buffer 800 m	−10.733	<0.000	−16.345	−5.122
	**	Address & school buffer 1 mile	−14.844	<0.000	−20.455	−9.232
	**	Convex hull	−26.020	<0.000	−31.631	−20.409
	**	1 standard deviational ellipses	−7.58	0.001	−13.270	−2.047
	**	2 standard deviational ellipses	−20.374	<0.000	−25.985	−14.762
	**	Path area	−25.474	<0.000	−31.086	−19.863
Address buffer 800 m		Address buffer 1 mile	−3.374	0.692	−8.985	2.237
	**	School buffer 800 m	55.839	<0.000	50.228	61.451
	**	School buffer 1 mile	47.432	<0.000	41.820	53.043
	**	Address & school buffer 800 m	−11.785	<0.000	−17.397	−6.174
	**	Address & school buffer 1 mile	−15.895	<0.000	−21.507	−10.284
	**	Convex hull	−27.072	<0.000	−32,683	−21.461
	**	1 standard deviational ellipses	−8.710	<0.000	−14.322	−3.099
	**	2 standard deviational ellipses	−21.425	<0.000	−27.037	−15.814
	**	Path area	−26.526	<0.000	−32.138	−20.915
Address buffer 1 mile	**	School buffer 800 m	59.213	<0.000	53.602	64.825
	**	School buffer 1 mile	50.806	<0.000	45.195	56.417
	**	Address & school buffer 800 m	−8.411	<0.000	−14.023	−2.800
	**	Address & school buffer 1 mile	−12.521	<0.000	−18.133	−6.910
	**	Convex hull	−23.698	<0.000	−29.309	−18.087
		1 standard deviational ellipses	−5.336	0.079	−10.948	0.275
	**	2 standard deviational ellipses	−18.051	<0.000	−23.663	−12.440
	**	Path area	−23.152	<0.000	−28.763	−17.541

Table S4. *Cont.*

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
School buffer 800 m	**	School buffer 1 mile	−8.407	<0.000	−14.018	−2.796
	**	Address & school buffer 800 m	−67.625	<0.000	−73.236	−62.014
	**	Address & school buffer 1 mile	−71.735	<0.000	−77.346	−66.124
	**	Convex hull	−82.912	<0.000	−88,523	−77.300
	**	1 standard deviational ellipses	−64.550	<0.000	−70.161	−58.938
	**	2 standard deviational ellipses	−77.265	<0.000	−82.876	−71.654
	**	Path area	−82.366	<0.000	−87.977	−76.754
School buffer 1 mile	**	Address & school buffer 800 m	−59.217	<0.000	−64.829	−53.606
	**	Address & school buffer 1 mile	−63.328	<0.000	−68.939	−57.716
	**	Convex hull	−74.504	<0.000	−80.116	−68.893
	**	1 standard deviational ellipses	−56.142	<0.000	−61.754	−50.531
	**	2 standard deviational ellipses	−68.858	<0.000	−74.469	−63.246
	**	Path area	−73.958	<0.000	−79.57	−68.347
Address & school buffer 800 m		Address & school buffer 1 mile	−4.110	0.393	−9.721	1.501
	**	Convex hull	−15.286	<0.000	−20.898	−9.675
		1 standard deviational ellipses	3.075	0.800	−2.536	8.686
	**	2 standard deviational ellipses	−9.640	<0.000	−15.251	−4.028
	**	Path area	−14.741	<0.000	−20.352	9.129
Address & school buffer 1 mile	**	Convex hull	−11.177	<0.000	−16.788	−5.565
	**	1 standard deviational ellipses	7.185	0.002	1.573	12.796
		2 standard deviational ellipses	−5.530	0.058	−11.141	0.081
	**	Path area	−10.630	<0.000	−16.242	−5.019
Convex hull	**	1 standard deviational ellipses	18.361	<0.000	12.750	23.973
	*	2 standard deviational ellipses	5.646	0.046	0.035	11.258
		Path area	−	1.000	−	−
1 standard deviational ellipses	**	2 standard deviational ellipses	−12.715	<0.000	−18.326	−7.103
	**	Path area	−17.815	<0.000	−23.427	−12.204
2 standard deviational ellipses	*	Path area	−5.646	0.046	−11.258	−0.035

Notes: * Statistically significant below the 0.05 level, ** Statistically significant below the 0.01 level.

Table S5. Results Tukey's HSD test for comparison of supermarket exposure.

Neighbourhood		Neighbourhood	Mean diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Parish		Address buffer 800 m	1.246	0.960	−1.735	4.227
		Address buffer 1 mile	−2.578	0.164	−5.559	0.404
		School buffer 800 m	−1.503	0.872	−4.484	1.479
	**	School buffer 1 mile	−9.219	<0.000	−12.201	−6.238
	*	Address & school buffer 800 m	−3.358	0.013	−6.340	−0.377
	**	Address & school buffer 1 mile	−13.267	<0.000	−16.249	−10.286
	**	Convex hull	−23.011	<0.000	−25.992	−20.029
		1 standard deviational ellipses	−2.936	0.058	−5.917	0.046
	**	2 standard deviational ellipses	−16.610	<0.000	−19.591	−13.628
	**	Path area	−8.011	<0.000	−10.992	−5.029

Table S5. *Cont.*

Neighbourhood		Neighbourhood	Mean diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Address buffer 800 m	**	Address buffer 1 mile	−3.824	0.002	−6.805	−0.842
		School buffer 800 m	−2.749	0.103	−5.730	0.233
	**	School buffer 1 mile	−10.465	<0.000	−13.447	−7.484
	**	Address & school buffer 800 m	−4.604	<0.000	−7.586	−1.623
	**	Address & school buffer 1 mile	−14.513	<0.000	−17.495	−11.532
	**	Convex hull	−24.257	<0.000	−27.238	−21.275
	**	1 standard deviational ellipses	−4.182	<0.000	−7.163	−1.200
	**	2 standard deviational ellipses	−17.856	<0.000	−20.837	−14.874
	**	Path area	−9.257	<0.000	−12.238	−6.275
Address buffer 1 mile		School buffer 800 m	1.075	0.986	−1.907	4.056
	**	School buffer 1 mile	−6.642	<0.000	−9.623	−3.660
		Address & school buffer 800 m	−0.781	0.999	−3.762	2.201
	**	Address & school buffer 1 mile	−10.690	<0.000	−13.671	−7.708
	**	Convex hull	−20.433	<0.000	−23.415	−17.452
		1 standard deviational ellipses	−0.358	1.000	−3.340	2.623
	**	2 standard deviational ellipses	−14.032	<0.000	−17.013	−11.051
	**	Path area	−5.433	<0.000	−8.415	−2.452
School buffer 800 m	**	School buffer 1 mile	−7.717	<0.000	−10.698	−4.735
		Address & school buffer 800 m	−1.856	0.645	−4.837	1.126
	**	Address & school buffer 1 mile	−11.765	<0.000	−14.746	−8.783
	**	Convex hull	−21.508	<0.000	−24.489	−18.527
		1 standard deviational ellipses	−1.433	0.903	−4.415	1.548
	**	2 standard deviational ellipses	−15.107	<0.000	−18.088	−12.126
	**	Path area	−6.508	<0.000	−9.489	−3.527
School buffer 1 mile	**	Address & school buffer 800 m	5.861	<0.000	2.880	8.842
	**	Address & school buffer 1 mile	−4.048	0.001	−7.030	−1.067
	**	Convex hull	−13.791	<0.000	−16.773	−10.810
	**	1 standard deviational ellipses	6.283	<0.000	3.302	9.265
	**	2 standard deviational ellipses	−7.390	<0.000	−10.372	−4.409
		Path area	1.209	0.968	−1.773	4.190
Address & school buffer 800 m	**	Address & school buffer 1 mile	−9.909	<0.000	−12.890	−6.928
	**	Convex hull	−19.652	<0.000	−22.634	−163.671
		1 standard deviational ellipses	0.422	1.000	−2.559	3.404
	**	2 standard deviational ellipses	−13.251	<0.000	−16.233	−10.270
	**	Path area	−4.652	<0.000	−7.634	−1.671
Address & school buffer 1 mile	**	Convex hull	−9.743	<0.000	−12.725	−6.762
	**	1 standard deviational ellipses	10.332	<0.000	7.350	13.313
	*	2 standard deviational ellipses	−3.342	0.014	−6.324	−0.361
	**	Path area	5.257	<0.000	2.275	8.238
Convex hull	**	1 standard deviational ellipses	20.075	<0.000	17.093	23.056
	**	2 standard deviational ellipses	6.401	<0.000	3.420	9.382
	**	Path area	15.000	<0.000	12.019	17.981
1 standard deviational ellipses	**	2 standard deviational ellipses	−13.674	<0.000	−16.655	−10.692
	**	Path area	−5.075	<0.000	−8.056	−2.093
2 standard deviational ellipses	**	Path area	8.599	<0.000	5.618	11.580

Notes: * Statistically significant below the 0.05 level, ** Statistically significant below the 0.01 level.

Table S6. Results Tukey's HSD test for comparison of supermarket exposure in urban sample.

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Parish**		Address buffer 800 m	1.138	0.996	−2.64	4.940
	**	Address buffer 1 mile	−5.500	<0.000	−9.302	−1.697
		School buffer 800 m	−0.117	1.000	−3.919	3.685
	**	School buffer 1 mile	−7.648	<0.000	−11.450	−3.846
		Address & school buffer 800 m	−2.968	0.294	−6.770	0.833
	**	Address & school buffer 1 mile	−13.968	<0.000	−17.770	−10.166
	**	Convex hull	−17.585	<0.000	−21.387	−13.783
		1 standard deviational ellipses	−1.244	0.993	−5.046	2.557
	**	2 standard deviational ellipses	−11.510	<0.000	−15.312	−7.708
	**	Path area	−5.776	<0.000	−9.578	−1.974
Address buffer 800 m	**	Address buffer 1 mile	−6.638	<0.000	−10.440	−2.836
		School buffer 800 m	−1.255	0.993	−5.057	2.546
	**	School buffer 1 mile	−8.787	<0.000	−12.589	−4.985
	*	Address & school buffer 800 m	−4.106	0.022	−7.908	−0.304
	**	Address & school buffer 1 mile	−15.106	<0.000	−18.908	−11.304
	**	Convex hull	−18.723	<0.000	−22.525	−14.921
		1 standard deviational ellipses	−2.382	0.634	−6.185	1.419
	**	2 standard deviational ellipses	−12.648	<0.000	−16.450	−8.846
	**	Path area	−6.914	<0.000	−10.716	−3.112
Address buffer 1 mile	**	School buffer 800 m	5.382	<0.000	1.580	9.185
		School buffer 1 mile	−2.148	0.766	−5.950	1.653
		Address & school buffer 800 m	2.531	0.542	−1.270	6.333
	**	Address & school buffer 1 mile	−8.468	<0.000	−12−270	−4.666
	**	Convex hull	−12.085	<0.000	−15.887	−8.283
	*	1 standard deviational ellipses	4.255	0.015	0.453	8.057
	**	2 standard deviational ellipses	−.010	<0.000	−9.812	−2.208
		Path area	−0.276	1.000	−4.078	3.525
School buffer 800 m	**	School buffer 1 mile	−7.531	<0.000	−11.531	−3.729
		Address & school buffer 800 m	−2.851	0.355	−6.653	0.950
	**	Address & school buffer 1 mile	−13.851	<0.000	−17.653	−10.049
	**	Convex hull	−17.468	<0.000	−21.270	−13.666
		1 standard deviational ellipses	−1.127	0.997	−4.929	2.674
	**	2 standard deviational ellipses	−11.393	<0.000	−15.195	−7.591
	**	Path area	−5.659	<0.000	−9.461	−1.857
School buffer 1 mile	**	Address & school buffer 800 m	4.680	0.004	0.878	8.482
	**	Address & school buffer 1 mile	−6.319	<0.000	−10.121	−2.517
	**	Convex hull	−9.936	<0.000	−13.738	−6.134
	**	1 standard deviational ellipses	6.404	<0.000	2.602	10.206
	*	2 standard deviational ellipses	−3.861	0.043	−7.663	−0.059
		Path area	1.872	0.887	−1.926	5.674
Address & school buffer 800 m	**	Address & school buffer 1 mile	−11.000	<0.000	−14.802	−7.197
	**	Convex hull	−14.617	<0.000	−18.419	−10.815
		1 standard deviational ellipses	1.723	0.931	−2.078	5.525
	**	2 standard deviational ellipses	−8.542	<0.000	−12.344	−4.740
		Path area	−2.808	0.378	−6.610	0.993

Table S6. *Cont.*

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Address & school buffer 1 mile		Convex hull	−3.617	0.079	−7.419	0.185
	**	1 standard deviational ellipses	12.723	<0.000	8.921	16.524
		2 standard deviational ellipses	2.457	0.588	−1.344	6.259
	**	Path area	8.191	<0.000	4.389	11.993
Convex hull	**	1 standard deviational ellipses	16.340	<0.000	12.538	20.340
	**	2 standard deviational ellipses	6.074	<0.000	2.272	9.876
	**	Path area	11.808	<0.000	8.006	15.611
1 standard deviational ellipses	**	2 standard deviational ellipses	−10.265	<0.000	−14.067	−6.463
	**	Path area	−4.531	0.006	−8.333	−0.729
2 standard deviational ellipses	**	Path area	5.734	<0.000	1.932	9.536

Notes: * Statistically significant below the 0.05 level, ** Statistically significant below the 0.01 level.

Table S7. Results Tukey's HSD test for comparison of supermarket exposure in rural sample.

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Parish		Address buffer 800 m	1.354	0.995	−3.002	5.712
		Address buffer 1 mile	0.376	1.000	−3.980	4.733
		School buffer 800 m	−2.903	0.542	−7.260	1.453
	**	School buffer 1 mile	−10.806	<0.000	−15.163	−6.449
		Address & school buffer 800 m	−3.752	0.168	−8.109	0.604
	**	Address & school buffer 1 mile	−12.559	<0.000	−16.916	−8.201
	**	Convex hull	−28.494	<0.000	−32.815	−24.137
	*	1 standard deviational ellipses	−4.645	0.025	−9.002	−0.288
	**	2 standard deviational ellipses	−21.763	<0.000	−26.120	−17.406
	**	Path area	−10.268	<0.000	−14.625	−5.911
Address buffer 800 m		Address buffer 1 mile	−0.978	0.999	−5.335	3.378
		School buffer 800 m	−4.258	0.062	−8.615	0.099
	**	School buffer 1 mile	−12.161	<0.000	−16.518	−7.804
	**	Address & school buffer 800 m	−5.107	0.008	−9.464	−0.750
	**	Address & school buffer 1 mile	−13.913	<0.000	−18.271	−9.556
	**	Convex hull	−29.849	<0.000	−34.206	−25.492
	**	1 standard deviational ellipses	−6.000	0.001	−10.357	−1.642
	**	2 standard deviational ellipses	−23.118	<0.000	−27.475	−18.761
	**	Path area	−11.623	<0.000	−15.980	−7.266
Address buffer 1 mile		School buffer 800 m	−3.279	0.349	−7.636	1.077
	**	School buffer 1 mile	−11.182	<0.000	−15.539	−6.825
		Address & school buffer 800 m	−4.129	0.082	−8.486	0.228
	**	Address & school buffer 1 mile	−12.934	<0.000	−17.292	−8.578
	**	Convex hull	−28.871	<0.000	−33.228	−24.513
	**	1 standard deviational ellipses	−5.021	0.009	−9.378	−0.664
	**	2 standard deviational ellipses	−22.139	<0.000	−26.496	−17.782
	**	Path area	−10.645	<0.000	−15.002	−6.287

Table S7. *Cont.*

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
School buffer 800 m	**	School buffer 1 mile	−7.903	<0.000	−12.260	−3.546
		Address & school buffer 800 m	−0.849	0.999	−5.206	3.507
	**	Address & school buffer 1 mile	−9.655	<0.000	−14.013	−5.298
	**	Convex hull	−25.591	<0.000	−29.948	−21.234
		1 standard deviational ellipses	−1.741	0.971	−6.099	2.615
	**	2 standard deviational ellipses	−18.860	<0.000	−23.217	−14.503
	**	Path area	−7.365	<0.000	−11.722	−3.008
School buffer 1 mile	**	Address & school buffer 800 m	7.053	<0.000	2.696	11.410
		Address & school buffer 1 mile	−1.752	0.969	−6.109	2.604
	**	Convex hull	−17.688	<0.000	−22.045	−13.330
	**	1 standard deviational ellipses	6.161	<0.000	1.804	10.518
	**	2 standard deviational ellipses	−10.956	<0.000	−15.314	−6.599
		Path area	0.537	0.999	−3.819	4.895
Address & school buffer 800 m	**	Address & school buffer 1 mile	−8.806	<0.000	−13.163	−4.449
	**	Convex hull	−24.741	<0.000	−29.099	−20.384
		1 standard deviational ellipses	−0.892	0.999	−5.249	3.464
	**	2 standard deviational ellipses	−18.010	<0.000	−22.367	−13.653
	**	Path area	−6.516	<0.000	−10.873	−2.158
Address & school buffer 1 mile	**	Convex hull	−15.935	<0.000	−20.292	−11.578
	**	1 standard deviational ellipses	7.913	<0.000	3.556	12.271
	**	2 standard deviational ellipses	−9.204	<0.000	−13.561	−4.847
		Path area	2.290	0.837	−2.066	6.647
Convex hull	**	1 standard deviational ellipses	23.849	<0.000	19.492	28.206
	**	2 standard deviational ellipses	6.731	<0.000	2.374	11.088
	**	Path area	18.225	<0.000	13.868	22.582
1 standard deviational ellipses	**	2 standard deviational ellipses	−17.118	<0.000	−21.475	−12.761
	**	Path area	−5.623	0.002	−9.980	−1.266
2 standard deviational ellipses	**	Path area	11.494	<0.000	7.137	15.851

Notes: * Statistically significant below the 0.05 level, ** Statistically significant below the 0.01 level.

Table S8. Results Tukey's HSD test for comparison of supermarket per square kilometre exposure in urban sample.

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Parish		Address buffer 800 m	−0.264	0.958	−0.893	0.365
		Address buffer 1 mile	0.230	0.984	−0.398	0.860
	**	School buffer 800 m	−0.888	<0.000	−1.517	−0.259
		School buffer 1 mile	−0.033	1.000	−0.662	0.595
		Address & school buffer 800 m	−0.543	0.165	−1.172	0.086
		Address & school buffer 1 mile	0.140	0.999	−0.488	0.770
		Convex hull	−0.429	0.503	−1.059	0.199
		1 standard deviational ellipses	−0.431	0.497	−1.061	0.197
		2 standard deviational ellipses	−0.002	1.000	−0631	0.627
	**	Path area	−1.915	<0.000	−2.545	−1.286

Table S8. *Cont.*

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Address buffer 800 m		Address buffer 1 mile	0.494	0.284	−0.134	1.124
		School buffer 800 m	−0.624	0.054	−1.253	0.005
		School buffer 1 mile	0.230	0.984	−0.398	0.860
		Address & school buffer 800 m	−0.278	0.940	−0.908	0.350
		Address & school buffer 1 mile	0.405	0.594	−0.224	1.034
		Convex hull	−0.165	0.998	−0.795	0.463
		1 standard deviational ellipses	−0.167	0.998	−0.796	0.462
		2 standard deviational ellipses	0.262	0.960	−0.367	0.891
	**	Path area	−1.651	<0.000	−2.280	−1.022
Address buffer 1 mile	**	School buffer 800 m	−1.119	<0.000	−1.748	−0.489
		School buffer 1 mile	−0.264	0.958	−0.893	0.365
	**	Address & school buffer 800 m	−0.773	0.004	−1.403	−0.144
		Address & school buffer 1 mile	−0.089	0.999	−0.719	0.539
	*	Convex hull	−0.660	0.030	−1.289	−0.031
	*	1 standard deviational ellipses	−0.662	0.029	−1.291	−0.032
		2 standard deviational ellipses	−0.232	0.983	−0.861	0.396
	**	Path area	−2.145	<0.000	−2.775	−1.516
School buffer 800 m	**	School buffer 1 mile	0.855	0.001	0.225	1.484
		Address & school buffer 800 m	0.345	0.797	−0.283	0.974
	**	Address & school buffer 1 mile	1.029	<0.000	0.399	1.658
		Convex hull	0.458	0.399	−0.171	1.087
		1 standard deviational ellipses	0.456	0.405	−0.172	1.086
	**	2 standard deviational ellipses	0.886	<0.000	0.257	1.515
	**	Path area	−1.027	<0.000	−1.656	−0.397
School buffer 1 mile		Address & school buffer 800 m	−0.509	0.243	−1.138	0.119
		Address & school buffer 1 mile	0.174	0.998	−0.455	0.803
		Convex hull	−0.39	0.626	−1.025	0.232
		1 standard deviational ellipses	−0.398	0.621	−1.027	0.231
		2 standard deviational ellipses	0.031	1.000	−0.597	0.660
	**	Path area	−1.882	<0.000	−2.511	−1.252
Address & school buffer 800 m	*	Address & school buffer 1 mile	0.684	0.020	0.054	1.313
		Convex hull	0.113	0.999	−0.516	0.742
		1 standard deviational ellipses	0.111	0.999	−0.517	0.740
		2 standard deviational ellipses	0.541	0.169	−0.088	1.170
	**	Path area	−1.372	<0.000	−2.002	−0.743
Address & school buffer 1 mile		Convex hull	−0.571	0.116	−1.200	0.058
		1 standard deviational ellipses	−0.572	0.113	−1.201	0.057
		2 standard deviational ellipses	−0.142	0.999	−0.772	0.486
	**	Path area	−2.056	<0.000	−2.685	−1.427
Convex hull		1 standard deviational ellipses	−0.002	1.000	−0.631	0.627
		2 standard deviational ellipses	0.427	0.510	−0.201	1.057
	**	Path area	−1.485	<0.000	−2.115	−0.856
1 standard deviational ellipses		2 standard deviational ellipses	0.429	0.504	−0.199	1.058
	**	Path area	−1.484	<0.000	−2.113	−0.854
2 standard deviational ellipses	**	Path area	−1.913	<0.000	−2.543	−1.284

Notes: * Statistically significant below the 0.05 level, ** Statistically significant below the 0.01 level.

Table S9. Results Tukey's HSD test for comparison of supermarket per square kilometre exposure in rural sample.

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Parish		Address buffer 800 m	−0.325	0.054	−0.653	0.003
		Address buffer 1 mile	−0.127	0.975	−0.455	0.200
	**	School buffer 800 m	.2.443	<0.000	−2.771	−2.115
	**	School buffer 1 mile	−1.501	<0.000	−1.829	−1.173
	**	Address & school buffer 800 m	−1.384	<0.000	−1.712	−1.056
	**	Address & school buffer 1 mile	−0.811	<0.000	−.1.139	−0.483
	**	Convex hull	−0.638	<0.000	−0.966	−0.310
	*	1 standard deviational ellipses	−0.343	0.030	−0.671	0.016
	*	2 standard deviational ellipses	−0.351	0.025	−0.679	−0.023
	**	Path area	−2.142	<0.000	−2.470	−1.814
Address buffer 800 m		Address buffer 1 mile	0.197	0.686	−0.130	0.525
	**	School buffer 800 m	−2.117	<0.000	−2.445	−1.789
	**	School buffer 1 mile	−1.176	<0.000	−1.504	−0.848
	**	Address & school buffer 800 m	1.058	<0.000	−1.386	−0.730
	**	Address & school buffer 1 mile	−0.486	<0.000	−0.814	−0.158
		Convex hull	−0.312	0.077	−0.640	0.015
		1 standard deviational ellipses	−0.018	1.000	−0.346	0.309
		2 standard deviational ellipses	−0.025	1.000	−0.355	0.302
	**	Path area	−1.817	<0.000	−2.145	−1.489
Address buffer 1 mile	**	School buffer 800 m	−2.315	<0.000	−2.643	−1.987
	**	School buffer 1 mile	−1.374	<0.000	−1.702	−1.046
	**	Address & school buffer 800 m	−1.256	<0.000	−1.584	−0.928
	**	Address & school buffer 1 mile	−0.684	<0.000	−1.013	−0.356
	**	Convex hull	−0.510	<0.000	−0.838	−0.182
		1 standard deviational ellipses	−0.216	0.558	−0.544	0.111
		2 standard deviational ellipses	−0.223	0.509	−0.551	0.104
	**	Path area	−2.014	<0.000	−2.342	−1.686
School buffer 800 m	**	School buffer 1 mile	0.941	<0.000	0.613	1.269
	**	Address & school buffer 800 m	1.058	<0.000	0,731	1.386
	**	Address & school buffer 1 mile	1.631	<0.000	1.303	1.959
	**	Convex hull	1.804	<0.000	1.476	2.133
	**	1 standard deviational ellipses	2.099	<0.000	1.771	2.427
	**	2 standard deviational ellipses	2.092	<0.000	1.764	2.420
		Path area	0.300	0.107	−0027	0.628
School buffer 1 mile		Address & school buffer 800 m	0.117	0.986	−0.210	0.445
	**	Address & school buffer 1 mile	0.690	<0.000	0.361	1.018
	**	Convex hull	0.863	<0.000	0.535	1.191
	**	1 standard deviational ellipses	1.158	<0.000	0.830	1.486
	**	2 standard deviational ellipses	1.151	<0.000	0.823	1.479
	**	Path area	−0.640	<0.000	−0.968	−0.312

Table S9. *Cont.*

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Address & school buffer 800 m	**	Address & school buffer 1 mile	0.572	<0.000	0.244	0.900
	**	Convex hull	0.746	<0.000	0.418	1.074
	**	1 standard deviational ellipses	1.040	<0.000	0.712	1.368
	**	2 standard deviational ellipses	1.033	<0.000	0.705	1.361
	**	Path area	−0.758	<0.000	−1.086	−0.430
Address & school buffer 1 mile		Convex hull	0.173	0.831	−0.154	0.501
	**	1 standard deviational ellipses	0.468	<0.000	0.140	0.795
	**	2 standard deviational ellipses	0.461	<0.000	0.133	0.789
	**	Path area	−1.330	<0.000	−1.658	−1.002
Convex hull		1 standard deviational ellipses	0.294	0.125	−0.033	0.622
		2 standard deviational ellipses	0.287	0.148	−0.040	0.615
	**	Path area	−1.504	<0.000	−1.832	−1.176
1 standard deviational ellipses		2 standard deviational ellipses	−0.007	1.000	−0.334	0.321
	**	Path area	−1.798	<0.000	−2.126	−1.470
2 standard deviational ellipses	**	Path area	−1.791	<0.000	−2.119	−1.463

Notes: * Statistically significant below the 0.05 level, ** Statistically significant below the 0.01 level.

Table S10. Results Tukey's HSD test for comparison of fast food outlet exposure.

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Parish		Address buffer 800 m	0.251	1.000	−5.770	6.273
		Address buffer 1 mile	−5.722	0.080	−11.743	0.299
	**	School buffer 800 m	−9.642	<0.000	−15.663	−3.621
	**	School buffer 1 mile	−22.925	<0.000	−28.946	−16.904
	**	Address & school buffer 800 m	−12.406	<0.000	−18.428	−6.385
	**	Address & school buffer 1 mile	−29.011	<0.000	−35.032	−22.990
	**	Convex hull	−42.856	<0.000	−48.877	−36.834
	**	1 standard deviational ellipses	−7.241	0.005	−13.262	−1.219
	**	2 standard deviational ellipses	−30.920	<0.000	−36.941	−24.899
	**	Path area	−20.225	<0.000	−26.246	−14.203
Address buffer 800 m		Address buffer 1 mile	−5.973	0.054	−11.994	0.048
	**	School buffer 800 m	−9.893	<0.000	−15.914	−3.872
	**	School buffer 1 mile	−23.176	<0.000	−29.198	−17.155
	**	Address & school buffer 800 m	−12.658	<0.000	−18.679	−6.637
	**	Address & school buffer 1 mile	−29.262	<0.000	−35.283	−23.241
	**	Convex hull	−43.107	<0.000	−49.128	−37.086
	**	1 standard deviational ellipses	−7.492	0.003	−13.513	−1.471
	**	2 standard deviational ellipses	−31.171	<0.000	−37.192	−25.150
	**	Path area	−20.476	<0.000	−26.497	−14.455

Table S10. *Cont.*

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Address buffer 1 mile		School buffer 800 m	−3.920	0.579	−9.941	2.101
	**	School buffer 1 mile	−17.203	<0.000	−23.224	−11.182
	*	Address & school buffer 800 m	−6.684	0.016	−12.706	−0.663
	**	Address & school buffer 1 mile	−23.289	<0.000	−29.310	−17.268
	**	Convex hull	−37.134	<0.000	−43.155	−31.112
		1 standard deviational ellipses	−1.519	0.999	−7.540	4.502
	**	2 standard deviational ellipses	−25.198	<0.000	−31.219	−19.177
	**	Path area	−14.503	<0.000	−20.524	−8.481
School buffer 800 m	**	School buffer 1 mile	−13.283	<0.000	−19.305	−7.262
		Address & school buffer 800 m	−2.765	0.927	−8.786	3.256
	**	Address & school buffer 1 mile	−19.369	<0.000	−25.390	−13.348
	**	Convex hull	−33.214	<0.000	−39.235	−27.193
		1 standard deviational ellipses	2.401	0.971	−3.620	8.422
	**	2 standard deviational ellipses	−21.278	<0.000	−27.299	−15.257
	**	Path area	−10.583	<0.000	−16.604	−4.562
School buffer 1 mile	**	Address & school buffer 800 m	10.519	<0.000	4.498	16.540
	*	Address & school buffer 1 mile	−6.086	0.045	−12.107	−0.064
	**	Convex hull	−19.930	<0.000	−25.952	−13.909
	**	1 standard deviational ellipses	15.684	<0.000	9.663	21.706
	**	2 standard deviational ellipses	−7.995	0.001	−14.016	−1.973
		Path area	2.701	0.937	−3.321	8.722
Address & school buffer 800 m	**	Address & school buffer 1 mile	−16.604	<0.000	−22.625	−10.583
	**	Convex hull	−30.449	<0.000	−36.470	−24.428
		1 standard deviational ellipses	5.166	0.173	−0.855	11.187
	**	2 standard deviational ellipses	−18.513	<0.000	−24.535	−12.492
	**	Path area	−7.818	0.001	−13.839	−1.797
Address & school buffer 1 mile	**	Convex hull	−13.845	<0.000	−19.866	−7.824
	**	1 standard deviational ellipses	21.770	<0.000	15.749	27.791
		2 standard deviational ellipses	−1.909	0.995	−7.930	4.112
	**	Path area	8.786	<0.000	2.765	14.807
Convex hull	**	1 standard deviational ellipses	35.615	<0.000	29.594	41.636
	**	2 standard deviational ellipses	11.936	<0.000	5.915	17.957
	**	Path area	22.631	<0.000	16.610	28.652
1 standard deviational ellipses	**	2 standard deviational ellipses	−23.679	<0.000	−29.700	−17.658
	**	Path area	−12.984	<0.000	−19.005	−6.963
2 standard deviational ellipses	**	Path area	10.695	<0.000	4.674	16.716

Notes: * Statistically significant below the 0.05 level, ** Statistically significant below the 0.01 level.**Table S11.** Results Tukey's HSD test for comparison of fast food outlet exposure in urban sample.

Neighbourhood	Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
				Lower	Upper
Parish	Address buffer 800 m	−0.414	1.000	−8.710	7.880

	**	Address buffer 1 mile	−11.372	0.001	−19.667	−3.077
		School buffer 800 m	−6.829	0.222	−15.124	1.465
	**	School buffer 1 mile	−19.627	<0.000	−27.922	−11.332
	**	Address & school buffer 800 m	−11.595	<0.000	−19.890	−3.300
	**	Address & school buffer 1 mile	−30.106	<0.000	−38.401	−21.811
	**	Convex hull	−36.000	<0.000	−44.295	−27.704
		1 standard deviational ellipses	−5.308	0.603	−13.603	2.986
	**	2 standard deviational ellipses	−24.659	<0.000	−32.954	−16.364
	**	Path area	−17.159	<0.000	−25.454	−8.864
Address buffer 800 m	**	Address buffer 1 mile	−10.957	0.001	−19.252	−2.662
		School buffer 800 m	−6.414	0.308	−14.710	1.880
	**	School buffer 1 mile	−19.212	<0.000	−27.507	−10.917
	**	Address & school buffer 800 m	−11.180	0.001	−19.476	−2.885
	**	Address & school buffer 1 mile	−29.691	<0.000	−37.986	−21.396
	**	Convex hull	−35.585	<0.000	−43.880	−27.289
		1 standard deviational ellipses	−4.893	0.715	−13.188	3.401
	**	2 standard deviational ellipses	−24.244	<0.000	−32.539	−15.949
	**	Path area	−16.744	<0.000	−25.039	−8.449
Address buffer 1 mile		School buffer 800 m	4.542	0.799	−3.752	12.837
		School buffer 1 mile	−8.255	0.052	−16.550	0.039
		Address & school buffer 800 m	−0.223	1.000	−8.518	8.071
	**	Address & school buffer 1 mile	−18.734	<0.000	−27.029	−10.438
	**	Convex hull	−24.627	<0.000	−32.992	−16.332
		1 standard deviational ellipses	6.063	0.394	−2.231	14.359
	**	2 standard deviational ellipses	−13.287	<0.000	−21.582	−4.992
		Path area	−5.787	0.469	−14.082	−2.507
School buffer 800 m	**	School buffer 1 mile	−12.797	<0.000	−21.093	−4.502
		Address & school buffer 800 m	−4.765	0.747	−13.061	3.529
	**	Address & school buffer 1 mile	−23.276	<0.000	−31.571	−14.981
	**	Convex hull	−29.170	<0.000	−37.465	−20.875
		1 standard deviational ellipses	1.521	0.999	−6.773	9.816
	**	2 standard deviational ellipses	−17.829	<0.000	−26.124	−9.534
	**	Path area	−10.329	0.003	−18.624	−2.034
School buffer 1 mile		Address & school buffer 800 m	8.031	0.068	−0.263	16.327
	**	Address & school buffer 1 mile	−10.478	0.002	−18.773	−2.183
	**	Convex hull	−16.372	<0.000	−24.667	−8.077
	**	1 standard deviational ellipses	14.319	<0.000	6.023	22.614
		2 standard deviational ellipses	−5.031	0.790	−13.327	3.263
		Path area	2.468	0.996	−5.827	10.763
Address & school buffer 800 m	**	Address & school buffer 1 mile	−18.510	<0.000	−26.805	−10.215
	**	Convex hull	−24.404	<0.000	−32.699	−16.109
		1 standard deviational ellipses	6.287	0.338	−2.007	14.582
	**	2 standard deviational ellipses	−13.063	<0.000	−21.359	−4.768
		Path area	−5.563	0.531	−13.859	2.731

Table S11. *Cont.*

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Address & school buffer 1 mile		Convex hull	−5.893	0.440	−14.188	2.401
	**	1 standard deviational ellipses	24.797	<0.000	16.502	33.093
		2 standard deviational ellipses	5.446	0.564	−2.848	13.742
	**	Path area	12.946	<0.000	4.651	21.242
Convex hull	**	1 standard deviational ellipses	30.691	<0.000	22.396	38.986
	**	2 standard deviational ellipses	11.340	<0.000	3.045	19.635
	**	Path area	18.840	<0.000	10.545	27.135
1 standard deviational ellipses	**	2 standard deviational ellipses	−19.351	<0.000	−27.646	−11.055
	**	Path area	−11.851	<0.000	−20.146	−3.555
2 standard deviational ellipses		Path area	7.500	0.119	−0.795	15.795

Note: * Statistically significant below the 0.05 level, ** Statistically significant below the 0.01 level.

Table S12. Results Tukey's HSD test for comparison of fast food outlet exposure in rural sample.

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Parish		Address buffer 800 m	0.924	0.999	−7.558	9.407
		Address buffer 1 mile	−0.010	1.000	−8.493	8.471
	**	School buffer 800 m	−12.483	<0.000	−20.966	−4.001
	**	School buffer 1 mile	−26.258	<0.000	−34.740	−17.775
	**	Address & school buffer 800 m	−13.225	<0.000	−21.708	−4.743
	**	Address & school buffer 1 mile	−27.903	<0.000	−36.385	−19.420
	**	Convex hull	−49.784	<0.000	−58.267	−41.302
	*	1 standard deviational ellipses	−9.193	0.021	−17.676	−0.710
	**	2 standard deviational ellipses	−37.247	<0.000	−45.730	−28.764
	**	Path area	−23.322	<0.000	−31.805	−14.839
Address buffer 800 m		Address buffer 1 mile	−0.935	0.999	−9.418	7.547
	**	School buffer 800 m	−13.408	<0.000	−21.891	−4.925
	**	School buffer 1 mile	−27.182	<0.000	−35.65	−18.700
	**	Address & school buffer 800 m	−14.150	<0.000	−22.633	−5.667
	**	Address & school buffer 1 mile	−28.827	<0.000	−37.310	−20.345
	**	Convex hull	−50.709	<0.000	−59.192	−42.226
	**	1 standard deviational ellipses	−10.118	0.006	−18.601	−1.635
	**	2 standard deviational ellipses	−38.172	<0.000	−46.654	−29.689
	**	Path area	−24.247	<0.000	−32.730	−15.764
Address buffer 1 mile	**	School buffer 800 m	−12.473	<0.000	−20.955	−3.990
	**	School buffer 1 mile	−26.247	<0.000	−34.730	−17.764
	**	Address & school buffer 800 m	−13.215	<0.000	−21.697	−4.732
	**	Address & school buffer 1 mile	−27.892	<0.000	−36.275	−19.409
	**	Convex hull	−49.774	<0.000	−58.256	−41.291
	*	1 standard deviational ellipses	−9.182	0.021	−17.665	−0.700
	**	2 standard deviational ellipses	−37.236	<0.000	−45.719	−28.753
	**	Path area	−23.311	<0.000	−31.794	−14.829

Table S12. *Cont.*

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
School buffer 800 m	**	School buffer 1 mile	−13.774	<0.000	−22.256	−5.291
		Address & school buffer 800 m	−0.741	1.000	−9.224	7.740
	**	Address & school buffer 1 mile	−15.419	<0.000	−23.902	−6.936
	**	Convex hull	−37.301	<0.000	−45.783	−28.818
		1 standard deviational ellipses	3.290	0.976	−5.192	11.773
	**	2 standard deviational ellipses	−24.763	<0.000	−33.246	−16.280
	**	Path area	−10.838	0.002	−19.321	−2.355
School buffer 1 mile	**	Address & school buffer 800 m	13.032	<0.000	4.549	21.515
		Address & school buffer 1 mile	−1.645	0.999	−10.127	6.837
	**	Convex hull	−23.526	<0.000	−32.009	−15.044
	**	1 standard deviational ellipses	17.064	<0.000	8.581	25.547
	**	2 standard deviational ellipses	−10.989	0.002	−19.471	−2.506
		Path area	2.935	0.989	−5.547	11.418
Address & school buffer 800 m	**	Address & school buffer 1 mile	−14.677	<0.000	−32.160	−6.194
	**	Convex hull	−36.559	<0.000	−45.041	−28.076
		1 standard deviational ellipses	4.032	0.908	−4.450	12.515
	**	2 standard deviational ellipses	−24.021	<0.000	−32.504	−15.538
	**	Path area	−10.096	0.006	−18.579	−1.614
Address & school buffer 1 mile	**	Convex hull	−21.881	<0.000	−30.364	−13.398
	**	1 standard deviational ellipses	18.709	<0.000	10.226	27.192
	*	2 standard deviational ellipses	−9.344	0.017	−17.826	−0.861
		Path area	4.580	0.813	−3.902	13.063
Convex hull	**	1 standard deviational ellipses	40.591	<0.000	32.108	49.074
	**	2 standard deviational ellipses	12.537	<0.000	4.054	21.020
	**	Path area	26.462	<0.000	17.979	34.945
1 standard deviational ellipses	**	2 standard deviational ellipses	−28.053	<0.000	−36.536	−19.571
	**	Path area	−14.129	<0.000	−22.611	−5.646
2 standard deviational ellipses	**	Path area	13.924	<0.000	5.441	22.407

Notes: * Statistically significant below the 0.05 level, ** Statistically significant below the 0.01 level.

Table S13. Results Tukey's HSD test for comparison of fast food outlet per square kilometre exposure in urban sample.

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Parish		Address buffer 800 m	0.195	0.999	−1.910	2.301
		Address buffer 1 mile	1.414	0.529	−0.691	3.520
	**	School buffer 800 m	−2.994	<0.000	−5.101	−0.888
		School buffer 1 mile	0.399	0.999	−1.706	2.506
		Address & school buffer 800 m	−1.239	0.718	−3.345	0.867
		Address & school buffer 1 mile	0.954	0.931	−1.151	3.060
		Convex hull	−0.796	0.980	−2.902	1.310
		1 standard deviational ellipses	−0.084	1.000	−2.190	2.021
		2 standard deviational ellipses	0.103	1.000	−2.002	2.210
	**	Path area	−4.610	<0.000	−6.716	−2.504

Table S13. *Cont.*

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Address buffer 800 m		Address buffer 1 mile	1.218	0.738	−0.887	3−324
	**	School buffer 800 m	−3.190	<0.000	−5.296	−1.084
		School buffer 1 mile	0.204	0.999	−1.901	2.310
		Address & school buffer 800 m	−1.434	0.507	−3.541	0.671
		Address & school buffer 1 mile	0.758	0.986	−1.347	2.864
		Convex hull	−0.991	0.913	−3.097	1.114
		1 standard deviational ellipses	−0.279	0.999	−2.386	1.826
		2 standard deviational ellipses	−0.091	1.000	−2.197	2.014
	**	Path area	−4.806	<0.000	−6.912	−2.700
Address buffer 1 mile	**	School buffer 800 m	−4.409	<0.000	−6.515	−2.303
		School buffer 1 mile	−1.014	0.901	−3.120	1.091
	**	Address & school buffer 800 m	−2.653	0.003	−4.759	−0.547
		Address & school buffer 1 mile	−0.460	0.999	−2.566	1.646
	*	Convex hull	−2.210	0.030	−4.316	−0.104
		1 standard deviational ellipses	−1.498	0.437	−3.604	0.607
		2 standard deviational ellipses	−1.310	0.643	−3.416	0.795
	**	Path area	−6.024	<0.000	−8.131	−3.918
School buffer 800 m	**	School buffer 1 mile	3.394	<0.000	1.288	5.501
		Address & school buffer 800 m	1.755	0.206	−0.350	3.861
	**	Address & school buffer 1 mile	3.949	<0.000	1.843	6.055
	*	Convex hull	2.198	0.032	0.093	4.304
	**	1 standard deviational ellipses	2.910	<0.000	0.804	5.016
	**	2 standard deviational ellipses	3.098	<0.000	0.992	5.204
		Path area	−1.615	0.321	−3.721	0.490
School buffer 1 mile		Address & school buffer 800 m	−1.639	0.299	−3.745	0.467
		Address & school buffer 1 mile	0.554	0.998	−1.551	2.660
		Convex hull	−1.195	0.761	−3.302	0.910
		1 standard deviational ellipses	−0.484	0.999	−2.590	1.621
		2 standard deviational ellipses	−0.295	0.999	−2.402	1.810
	**	Path area	−5.010	<0.000	−7.116	2.904
Address & school buffer 800 m	*	Address & school buffer 1 mile	2.193	0.033	0.087	4.299
		Convex hull	0.443	0.999	−1.663	2.549
		1 standard deviational ellipses	1.154	0.798	−0.951	3.261
		2 standard deviational ellipses	1.343	0.608	−0.763	3.449
	**	Path area	−3.371	<0.000	−5.477	−1.265
Address & school buffer 1 mile		Convex hull	−1.750	0.209	−3.856	0.355
		1 standard deviational ellipses	−1.038	0.886	−3.144	1.067
		2 standard deviational ellipses	−0.850	0.968	−2.956	1.255
	**	Path area	−5.564	<0.000	−7.671	−3.458
Convex hull		1 standard deviational ellipses	0.711	0.991	−1,394	2.817
		2 standard deviational ellipses	0.899	0.953	−1.206	3.006
	**	Path area	−3.814	<0.000	−5.920	−1.708
1 standard deviational ellipses		2 standard deviational ellipses	0.188	1.000	−1.917	2.294
	**	Path area	−4.526	<0.000	−6.632	−2.420
2 standard deviational ellipses	**	Path area	−4.714	<0.000	−6.820	−2.608

Notes: * Statistically significant below the 0.05 level, ** Statistically significant below the 0.01 level.

Table S14. Results Tukey's HSD test for comparison of fast food outlet per square kilometre exposure in rural sample.

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Parish		Address buffer 800 m	−0.278	0.999	−1.415	0.858
		Address buffer 1 mile	−0.115	0.999	−1.252	1.021
	**	School buffer 800 m	−6.947	<0.000	−8.083	−5.810
	**	School buffer 1 mile	−3.341	<0.000	−4.477	−2.204
	**	Address & school buffer 800 m	−3.612	<0.000	−4.749	−2.476
	**	Address & school buffer 1 mile	−1.729	<0.000	−2.866	−0.593
	*	Convex hull	−1.239	0.019	−2.375	−0.102
		1 standard deviational ellipses	−0.660	0.734	−1.797	0.476
		2 standard deviational ellipses	−0.639	0.771	−1.775	0.497
	**	Path area	−4.567	0	−5.703	−3.430
Address buffer 800 m		Address buffer 1 mile	0.163	0.999	−0.973	1.299
	**	School buffer 800 m	−6.668	<0.000	−7.805	−5.532
	**	School buffer 1 mile	−3.062	<0.000	−4.199	−1.926
	**	Address & school buffer 800 m	−3.334	<0.000	−4.471	−2.197
	**	Address & school buffer 1 mile	−1.451	0.002	−2.588	−0.314
		Convex hull	−0.960	0.189	−2.097	0.175
		1 standard deviational ellipses	−0.382	0.991	−1.518	0.754
		2 standard deviational ellipses	−0.361	0.994	−1.497	0.775
	**	Path area	−4.288	0	−5.425	−3.152
Address buffer 1 mile	**	School buffer 800 m	−6.831	<0.000	−7.968	−5.695
	**	School buffer 1 mile	−3.225	<0.000	−4.362	−2.089
	**	Address & school buffer 800 m	−33.497	<0.000	−4.633	2.360
	**	Address & school buffer 1 mile	−1.614	<0.000	−2.750	−0.477
		Convex hull	−1.123	0.055	−2.260	0.013
		1 standard deviational ellipses	−0.544	0.903	−1.681	0.597
		2 standard deviational ellipses	−0.523	0.924	−1.660	0.612
	**	Path area	−4.451	<0.000	−5.588	−3.315
School buffer 800 m	**	School buffer 1 mile	3.605	<0.000	2.469	4.742
	**	Address & school buffer 800 m	3.334	<0.000	2.197	4.471
	**	Address & school buffer 1 mile	5.217	<0.000	4.080	6.354
	**	Convex hull	5.708	<0.000	4.571	6.844
	**	1 standard deviational ellipses	6.286	<0.000	5.150	7.423
	**	2 standard deviational ellipses	6.307	<0.000	5.171	7.444
	**	Path area	2.380	<0.000	1.243	3.516
School buffer 1 mile		Address & school buffer 800 m	−0.271	0.999	−1.408	0.865
	**	Address & school buffer 1 mile	1.611	<0.000	0.474	2.748
	**	Convex hull	2.102	<0.000	0.965	3.238
	**	1 standard deviational ellipses	2.680	<0.000	1.544	3.817
	**	2 standard deviational ellipses	2.701	<0.000	1.565	3.838
	*	Path area	−1.225	0.022	−2.362	−0.089

Table S14. *Cont.*

Neighbourhood		Neighbourhood	Mean Diff.	Sig.	95% Conf. Interval of the Differences	
					Lower	Upper
Address & school buffer 800 m	**	Address & school buffer 1 mile	1.882	<0.000	0.746	3.019
	**	Convex hull	2.373	<0.000	1.236	3.510
	**	1 standard deviational ellipses	2.952	<0.000	1.815	4.089
	**	2 standard deviational ellipses	2.973	<0.000	1.836	4.110
		Path area	−0.954	0.197	−2.091	0.182
Address & school buffer 1 mile		Convex hull	0.490	0.950	−0.645	1.627
		1 standard deviational ellipses	1.069	0.087	−0.067	2.206
		2 standard deviational ellipses	1.090	0.073	−0.046	2.227
	**	Path area	−2.837	<0.000	−3.974	−1.701
Convex hull		1 standard deviational ellipses	0.578	0.863	−0.557	1.715
		2 standard deviational ellipses	0.599	0.834	−0.536	1.736
	**	Path area	−3.328	<0.000	−4.464	−2.191
1 standard deviational ellipses		2 standard deviational ellipses	0.021	1.000	−1.115	1.157
	**	Path area	−3.906	<0.000	−5.043	−2.770
2 standard deviational ellipses	**	Path area	−3.927	<0.000	−5.064	−2.791

Notes: * Statistically significant below the 0.05 level, ** Statistically significant below the 0.01 level.

ISSN (online): 2246-1248
ISBN (online): 978-87-7112-492-7

AALBORG **UNIVERSITY PRESS**