Aus dem
Helmholtz Zentrum München
Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH)
Institut für Gesundheitsökonomie und Management im Gesundheitswesen

Institutsleiter: Prof. Dr. Reiner Leidl

Integrating information concerning the obesogenic environment from geocoding services into the surveillance of diabetes risk factors

Einbeziehung von Informationen zur adipogenen Umwelt aus Geokodierungsdiensten in die Surveillance von Diabetesrisikofaktoren

Dissertation
zum Erwerb des Doktorgrades der Humanbiologie
an der Medizinischen Fakultät der
Ludwig-Maximilians-Universität zu München

vorgelegt von
Maximilian Alexander Präger

aus
München

Jahr
2024

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Berichterstatter:	Prof. Dr. Rolf Holle
Mitberichterstatter:	Prof. Dr. Christine Meisinger
	Prof. Dr. Eberhard Standl
Mitbetreuung durch den promovierten Mitarbeiter:	Dr. Werner Maier
Dekan:	Prof. Dr. med. Thomas Gudermann
Tag der mündlichen Prüfung:	07.05.2024

Affidavit

Präger, Maximilian Alexander

Surname, first name

I hereby declare, that the submitted thesis entitled

Einbeziehung von Informationen zur adipogenen Umwelt aus Geokodierungsdiensten in die Surveillance von Diabetesrisikofaktoren

Integrating information concerning the obesogenic environment from geocoding services into the surveillance of diabetes risk factors

is my own work. I have only used the sources indicated and have not made unauthorized use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the dissertation presented here has not been submitted in the same or similar form to any other institution for the purpose of obtaining an academic degree.

Munich, June 11th 2024

Maximilian Präger

Table of contents

Abbreviations

API	Application programming interface
BIRCH	Balanced iterative reducing and clustering using hierarchies
BMI	Body mass index
CDC	Centers for Disease Control and Prevention
DALY	Disability adjusted life year
DBSCAN	Density-based spatial clustering of applications with noise
DIAB-CORE	Diabetes Collaborative Research of Epidemiologic Studies
FKZ	Förderkennzeichen (in English: grant number)
GIS	Geographic information system
IDF	International Diabetes Federation
KDE	Kernel density estimation
NCD	Non-communicable disease
NDSS	National Diabetes Surveillance System
OSM	OpenStreetMap
POI	Point of interest
QGIS	Quantum-GIS
RKI	Robert Koch Institute
SORS	Spatial obesity risk score
STEPS	Stepwise approach to surveillance
T2DM	Type 2 diabetes mellitus
UK	United Kingdom
US	United States
WHO	World Health Organization

1. Introduction

"The increasingly obesogenic environment we live in makes it harder for individuals to avoid unhealthy lifestyle choices. The obesogenic environment can be considered to be at the root of the prevention challenge in Type 2 diabetes."

Public Health England. Health matters: preventing Type 2 Diabetes (1)

Sedentary lifestyles, seen in recent decades in western countries, as well as increasingly observed in other regions worldwide, have led to a fundamental rise in obesity and diabetes as well as other non-communicable disease rates, thus leading to a shift in the importance from infectious toward non-infectious disease, i.e., the so-called "epidemiological transition" (2). Understanding the promoting mechanisms, as well as the interconnection of obesity and diabetes, is crucial to develop methods to mitigate these severe public health issues. To enrich the available methodology, this thesis aims to develop a geographic tool to visualize distributions of obesity-related environmental factors on maps. In addition, composite risks are targeted, which are estimated based on these factors. It is examined and discussed whether this tool can be applied within diabetes surveillance.

1.1 Overweight and obesity

According to the World Health Organization (WHO), overweight and obesity are defined based on a body mass index (BMI) greater than 25 and 30, respectively, with prevalence rates growing over recent decades (3). Thus, reducing these rates as well as treating obese patients is a major concern of public health activities. Typical interventions to mitigate obesity are dietary therapies targeting calorie restrictions or meal replacement (4). In addition, medication is used, for example to control the absorption of ingested fat. Severe overweight cases, for which other treatment options have not been successful, can undergo bariatric surgery (4).

Obesity can be classified into the phenotypes “acquired obesity”, which occurs as a result of unhealthy lifestyles, as well as ”inherited obesity”, which in most cases already appears very early in life before puberty and can be classified into “genetic” or “non-genetic” (5). Understanding the promoters of these conditions is a fundamental but complex task. Mechanisms behind the promotion of or reduction in obesity and weight management are often described based on energy expenditure or energy imbalance, which means that calories ingested by the individual exceed the number of calories burned by exercise. A deeper understanding of these drivers of body fat composition is helpful for obesity management (6).

Newer approaches to classify obesity have been proposed that take further variables and profiles into account, such as cardiometabolic aspects (7). Thus, these definitions can help to describe further subtypes of obesity for which the classical definitions fall short. In general, a focus on obesogenic factors is of major importance in obesity-related research. Obesity risks can be determined in many dimensions. For health care and epidemiology, factors that can be influenced through interventions and policies are important. Environmental obesity risk factors fall into this category, and it should be analyzed where they originate from and to what extent they can be adjusted to improve people’s health. A first step in this process is the discovery and visualization of the obesity-related factors of the so-called “built environment”.

1.2 Obesity risks and the environment

A specific epidemiological approach in obesity-related research is represented by the term “obesogenic environment”, which is also the central concept of this dissertation. It is used to connect obesity risks with a geographic scale and thus enables spatial analyses in the field of overweight/obesity. “Obesogenic environments” were initially described by Swinburn and colleagues (8) in their 1999 publication. The main idea behind this concept is that certain environmental factors promote or prevent obesity. These factors are often related to the built environment, i.e., to conditions that were generated by humans and thus should be targeted by health sciences. The environmental factors considered in this dissertation showed consistent effects within the literature, as outlined in the first study. In most cases, they are related to either food or physical activity, and are either positively or negatively correlated with obesity. The obesity correlates can be combined and set off against each other within the developed methodology of this dissertation. The available data are located in a geographic space that allows the derivation of maps that are important for relevant stakeholders, e.g., for health policy makers (9). These maps are helpful to visualize distributions of factor patterns and can thus be used to identify potential accumulated obesity risks (10). They are useful in cases of scarce budgets in order to decide which regions would profit the most from structural improvements, i.e., to detect gaps between regions with a high percentage of obese people and government actions to counteract unhealthy structures (11). In addition, they can be applied on a small scale, such as within city districts, or on a broader scale, such as whole metropolitan areas or even countries (12). Additional details regarding useful methodological approaches in this field are covered in the second article in this dissertation, as well as within the following sections.

Taking into account the health consequences of obesity, it is important to intervene at an early stage before subsequent diseases can emerge or progress. The most important disease that is tightly coupled with obesity caused by unhealthy lifestyles is Type 2 diabetes mellitus (T2DM). Thus, understanding this disease and monitoring of influential factors can help to improve population health.

1.3 Diabetes and disease surveillance

Diabetes is a metabolic disorder characterized by a lack of insulin secretion and/or disturbed insulin effect (13). The prevalence of diabetes has increased substantially worldwide over recent years, and its complications have evolved over a broader range, thus leading to a greater proportion of multimorbid cases (14). Among the diabetes risk factors, obesity is a major factor responsible for insulin resistance, and the vast majority of T2DM patients are obese (15). This underlines the importance of obesity and weight control among diabetes patients.

Important therapies against diabetes are lifestyle programs that are supplemented with medications for weight management, such as glucose-lowering drugs (16). Evidence concerning diabetes prevalence and incidence can be derived from studies of the DIAB-CORE Consortium (17) for the German context, as well as the IDF Diabetes Atlas (18), which adopts a global view on diabetes. As the health consequences of diabetes are severe, onset and progression have to be detected at an early stage and mitigated. In addition, disease monitoring is a common approach to counteract incident and prevalent cases for diabetes prevention. Monitoring risk factors on an aggregated population level is often covered by disease surveillance systems (19). In the 1960s, surveillance was defined at the Centers for Disease Control and Prevention (CDC) as an extensive collection of disease-related data that is assessed and reported, as well as disseminated to policy makers (20). Surveillance was initially defined for communicable diseases; however, since the mid-1970s, a shift in the focus from communicable toward non-communicable diseases (NCDs) could be observed (21), which made it necessary to implement surveillance methods also in NCD domains such as the field of diabetes. The WHO thus established a stepwise approach to surveillance (STEPS) for NCDs (22).

The aim of diabetes surveillance programs is to establish a reporting approach based on a predefined set of indicators, and the related data can be classified into primary and secondary data (23). Diabetes surveillance programs play an important role in national health systems worldwide. In Canada, the National Diabetes Surveillance System (NDSS) was established by the government in 1999 and became an overall chronic disease surveillance system afterward (24). Data could be used for prevalence estimates, as well as for all-cause mortality of patients with diabetes. Other examples of successful diabetes surveillance programs include programs in the US as well as in Australia (25).

In Germany, diabetes caused 268,000 years of life lost regarding mortality in 2017 and showed a substantial regional inequality concerning the distribution of the disease (26). A clear trend has been observed during recent years toward increasing prevalence rates and disability adjusted life years (DALYs) (27). In addition, patient groups with diabetes showed higher indirect costs compared with non-diabetic control groups (28). Thus, diabetes surveillance plays an important role in the German context to understand the composition and structure related to diabetes and to enable effective prevention and control mechanisms (29). The German Diabetes Surveillance system was established by the Robert Koch Institute (RKI) in 2015 (30). In the field of diabetes mellitus, the RKI follows some well-established approaches, such as monitoring the dynamics of diseases, determinants, and secondary diseases related to diabetes (31). For example, spatial monitoring of diabetes indicators such as overweight/obesity and physical activity was a primary topic as part of these approaches (32). In cooperation with the

RKI, the Diabetes Surveillance Project (grant from the Federal Ministry of Health, Germany (FKZ: GE20160324)) was initiated to develop and extend surveillance methods related to T2DM, on which this dissertation is partly based.

Results of research activities derived from diabetes surveillance activities in Germany are also used within German national medical guidelines concerning T2DM (33). Findings derived from surveillance include descriptions of the course of the disease and prevalence/incidence rates. These guidelines were developed to improve daily medical practice and are often important sources of information for health practitioners (34).

This work has a focus on obesity and not directly on diabetes, as obesity can be a valuable approximation for effects and metrics related to diabetes on account of their close connectivity. This can be seen from the fact that a high percentage of people with diabetes are obese and that diabetes has the strongest association with obesity compared with other diseases (35). Focusing on obesity instead of diabetes made it possible to take advantage of a vast amount of literature concerning obesity and the environment, on which the methodology is based.

1.4 Health geography—origins and common methods

Obesity and the environment are common dimensions in health geography (36). Useful geographic methods have emerged over the past centuries and can be applied in this field. Health geography describes and handles the effect of geography/environment/space on the health of individuals (37). It is a discipline in which humans and the environment act together and which encompasses social, political, and cultural dimensions related to health (36). Important milestones in health geographic history were the findings by geographer Charles Picquet in 1832, who designed specific map coloring for cholera in the city of Paris (38). A further important milestone was disease maps again related to cholera, which were developed by John Snow in 1854 who identified contaminated water as the root cause (39,40). These initial phases were directly related to diseases rather than to health-related aspects; however, there was a remarkable shift afterward from medical geography toward health geography, as well as toward a broader population-based focus (37).

Common health-related metrics have been developed that can be interpreted in a spatial dimension. Disease prevalence and incidence, for example, can be calculated region specifically. These disease-related metrics can be depicted as, for example, choropleth maps (41). A further geographic construct in the field of health sciences and geography is area-level deprivation, which can be described as the socioeconomic characteristics of a region and can be decomposed into variables such as education, income, and housing (42). The German Index of Multiple Deprivation was established based on the example from the UK (43), and has been used in epidemiologic and health care research in Germany (44).

Geographic information systems (GIS) can be used to generate and handle spatial health-related data. A GIS is an information system that is predominantly used for the generation, handling, and analysis of geographic data, which can be shown on maps (45). There are some free software tools and programming languages that are commonly used for geographic analysis, such as QGIS, R, and Python, in contrast to some proprietary tools such as ArcGIS (46).

1.5 Geocoding and online map services

This dissertation is based on the concept of geocoding, which means matching address descriptions of locations with geographic coordinates (47). This is a common approach in health-related geography (48), which allows integration of location-related data into the analyses. One common approach in the area of geocoding is translating street addresses into spatial coordinates. A more recent example of this approach was a study examining the correlation of traffic and ozone, for which geocoding of homes was needed (49).

Computerized geocoding approaches became popular in health-related disciplines in the 1980s, especially in cases of larger data (50,51). However, geocoding required a significant amount of computer resources, costs, and skilled staff and was not always available to health care professionals. Technical innovations helped to overcome these issues. Thus, one of the first online geocoding services targeting a broader audience was offered by Google (52), which also showed high-level structuring and data quality. Further examples of online geocoding services are described by Monir and colleagues (53).

Apart from pure geocoding functionality, a common feature of online geocoding services is the possibility to download spatial data points, which were geocoded in the past, via a variable- and location-based query (54). The geocoding service thus constitutes a database containing, for example, spatial point of interest (POI) data, i.e., the geographic location of an object together with additional information related to that object (54). This approach also significantly reduces costs related to persistent data storage and thus avoids data generation in the field. Data used in this dissertation were extracted from the two online geocoding services, Google Maps and OpenStreetMap (OSM). Other similar popular services are Bing Maps and Here Maps (55).

1.6 Geographic data quality and sensitivity

The increasing public availability of geographic data via online databases (56) poses a question regarding data quality standards and sensitivity related to these data. Central dimensions of spatial data quality described within the ArcGIS software are completeness, logical consistency, spatial accuracy, thematic accuracy, temporal quality, as well as data usability (57). Especially concerning user-generated content, as in the case of OSM, measures of data quality are important. Thus, data quality was examined in the first study. Regarding the quality dimensions, completeness especially played an important role in the data validation approach. In general, data usability and logical consistency were also important within the methodology of this dissertation, as well-suited spatial techniques could be implemented and the data completely matched with the requirements of the two studies. Temporal aspects were of minor importance owing to the data snapshots that were used. This facilitated the initial development of a spatial risk score; however, future versions can be extended in a time dimension. Finally, accuracy measures have been applied within the literature and were also included in consideration.

The public availability of the data underlines that the data is not highly critical such as, for example, medical claims data. It is structured on a more aggregated area level, i.e., factors are more generally described as promoting or preventing weight gain in a geographic space. Thus, the freely available data can either be used efficiently in a stand-alone fashion without having to take care of data protection regulation, as shown in both articles in this dissertation, or even as a source of database enrichment (58). The latter approach needs a variable on which different types of data sources can be merged.

1.7 Analyzing geographic data related to obesity and diabetes

In this dissertation, POI data related to obesity were downloaded from online geocoding services and analyzed via GIS methods. The relevant variables and their correlation with obesity were based on evidence from the literature. These obesity-related data points can be depicted on a map using geographic coordinates in latitude–longitude pairs (59). In general, influential factors on obesity/diabetes have complex underlying mechanisms and thus require consideration as well as understanding in many dimensions (60). Therefore, it is essential to have a spatial database such as the online geocoding services previously described that covers a variety of different factors. Each health-related spatial object, e.g., a green space or a restaurant, can be represented by a single spatial data point located at the given coordinates (61). As these obesity-related data points represent various factor types with different sizes and scopes, it is important to take the differences between these factors into account. One possible solution is a sensitivity analysis that examines weighting schemes, as was done within the risk score estimation framework. Despite the heterogeneous shape of the POI types, they can be summarized into a composite overall picture, as the main outcome of these data points is their effect on the increase in or mitigation of obesity risks.

Further spatial objects that can be handled by a GIS are lines and spatial polygons, which are often used to represent geographic areas (62,63). Two methods of area representation are used in this dissertation. The first approach is based on geographic borders in shape file format (64), which can be downloaded from official registries, and are provided, for example, by the Bavarian government (65). The second type of areal representation is the so-called bounding box, which is a rectangular geographic shape (66) and is also used for data download from the online geocoding service OpenStreetMap (67). In contrast, the download area in Google Maps has a circular format.

Spatial point data can be analyzed with clustering methods to describe the accumulation of these points within the study area of interest. Popular techniques include hierarchical clustering, as well as partitioning methods such as k-means clustering (68). In this dissertation, we use a density-based clustering method (Density-based spatial clustering of applications with noise, DBSCAN) for data pre-processing (69). In addition, a risk score for diabetes was developed based on kernel density estimation (KDE), a technique that is also able to describe accumulated point patterns in a pre-defined estimation window (70).

Further common geographic metrics are local Moran’s I for spatial autocorrelation, kriging for interpolation, as well as so-called “gravity models” for spatial interaction (71).

Spatial obesity and diabetes-related studies found in the literature often take risk factors as their inputs and predict disease prevalence as output. Among those risk factors, food environment and physical activity can often be found (72). In addition, common versions of diabetes risk scores were derived from patient-related characteristics such as age- and weight-related input factors (73,74). In this dissertation, we provide disease risks as outputs instead of estimated prevalence rates. Further details regarding this approach can be found in the following sections.

The starting point of the main analysis was a “points in polygon” approach, i.e., the downloaded data points are bounded by a pre-defined study area. Owing to complex geographic boundaries, study areas often deviate from the download area of online geocoding services, which have a rather rectangular or circular format. Thus, the downloaded data points need to be checked to

see whether they lie within the study area. The respective study area is used to visualize the spatial points and serves as the estimation window for the subsequent density modeling approach, which is described within the following sections.
For the implementation of the spatial methodology, the R environment for statistical computing (75) was used, as several useful packages exist which efficiently enable GIS analyses, including overlay plots, spatial interpolation techniques and, most important, point pattern analysis (76). A good overview on spatial methods implemented in R can also be found in Bivand et al 2013 (77). R also enables efficient sensitivity and uncertainty analyses via some native data structures, as well as some syntactical constructs that handle multi-dimensional data. To download geographic data, we used the Google places API (78), the osmar R package (79), and the Overpass Turbo API (80). In addition, a repository established by the Bavarian government (65) served as the source for the borders of the study area.

1.8 Aim and content of this dissertation

The main purpose of this dissertation is to develop a geographic tool to assess the obesogenic environment that can be applied within diabetes surveillance. Two steps are important in this process. First, the general feasibility of using possible data sources is examined. The second step can be described as deriving statistical measures and risk score maps for the target areas using GIS approaches. Data are captured from online geocoding services Google Maps and OpenStreetMap. Several geographic areas on a smaller scale in Bavaria are chosen as pilot areas for testing, data generation, data processing, as well as data validation. This risk score can be applied to larger areas than the pilot areas.
The first article, “Using data from online geocoding services for the assessment of environmental obesogenic factors: a feasibility study”, describes the feasibility of data extraction via online geocoding services, as well as data visualization and evaluation of data quality in the field. Literature searches are performed to derive an extensive list of factors that are correlated with obesity. The geocoding services performed adequately during field validation. OSM had a higher positive predictive value, whereas Google Maps showed higher sensitivity. The work presented in this article is the pre-requisite for the subsequent analysis.
The approach in the second article, “A spatial obesity risk score for describing the obesogenic environment using kernel density estimation: development and parameter variation”, aims to summarize the different risk structures and show a compound picture on maps. Important steps

include data preprocessing, ex-ante visualizations on maps, testing for spatial heterogeneity, as well as developing the spatial obesity risk score (SORS) based on KDE. Uncertainty analysis is used to evaluate the score and to choose the optimal parameters. Bandwidth and size of the study area have the greatest impact on the result. In addition, risk score maps show parts of the study areas with higher or lower obesity risks.

1.9 Stakeholder implications

The geographic tool presented in this dissertation has an effect on each type of prevention, i.e., primary, secondary, as well as tertiary prevention. The maps show disadvantaged and privileged areas and may motivate policy makers to counteract obesogenic and promote protective environmental structures. This would defer or prevent the onset of overweight and obesity, as well as progression in the population of interest. Concerning the factors described previously, two approaches become obvious: increasing the availability or strengthening the effect of physical activity facilities, as well as making the supply of unhealthy food less attractive. In general, interventions that are targeted at the environment level were found to be cost effective, especially as their positive impact took effect directly at a population level (81). Several success stories are described within the literature in which policies had a positive effect on overweight and diabetes. For example, the health effects of changes in food composition being animal-based vs vegetarian-based could be compared across several countries and showed a larger impact in high-income countries (82). Furthermore, interventions could be designed at a community level considering the regional risk level, such as the supply of healthy food (83). Common and effective physical activity policies included physical activity interventions (84).
Some modeling approaches for generating evidence for the complex underlying mechanisms, such as mathematical models explaining the dynamics of obesity or cost-effectiveness studies, were helpful in the past to implement the policies described above. Thus, there was an increasing demand to intensify these efforts (81). This underlines the importance of overweight/diabetes modeling tools for policy makers. Data generation, data processing, and risk score estimation within the studies in this dissertation are automated, and corresponding code is published on GitHub (85) or attached to the published manuscript. The provided code templates can thus be adapted to a different region by downloading the respective data and by loading it into the data processing software afterward. For this purpose, the coordinates of the geographic area of interest need to be entered into the templates for subsequent data download

and data analysis. This also enables the efficient application of the developed geographic tool on a different scale and may even offer additional code extension possibilities for stakeholder groups.

1.10 Author contribution

The author of this dissertation (MP) developed the design of the two studies together with the team. MP generated all database queries, visualizations, processed the data, and implemented the statistical methodology for both studies based on code snippets that were made publicly available. Furthermore, the author coordinated and performed an extensive literature search for the first study. In the context of the field validations, the tasks were divided within the team, and MP evaluated the data in one of the four study areas in the field. The risk score algorithm of the second study was developed and optimized by MP. The author of the dissertation contributed to the internal methodological discussions regarding POI validation, literature search approach, risk score evaluation, and parameter variation. Finally, all manuscripts were written by MP, and all team members commented and agreed on the manuscripts.

References

1. Public Health England. Health matters: preventing Type 2 Diabetes [Internet]. [cited 2023 Aug 22]. Available from: https://www.gov.uk/government/publications/health-matters-preventing-type-2-diabetes/health-matters-preventing-type-2-diabetes
2. Egger G, Dixon J. Beyond obesity and lifestyle: a review of 21st century chronic disease determinants. Biomed Res Int. 2014;2014:731685.
3. World Health Organization. Obesity [Internet]. [cited 2023 Aug 22]. Available from: https://www.who.int/health-topics/obesity#tab=tab_1
4. Ruban A, Stoenchev K, Ashrafian H, Teare J. Current treatments for obesity. Clin Med (Lond). 2019 May;19(3):205–12.
5. Archer E, Lavie CJ. Obesity subtyping: the etiology, prevention, and management of acquired versus inherited obese phenotypes. Nutrients. 2022 May 30;14(11).
6. Westerterp KR. Exercise, energy balance and body composition. Eur J Clin Nutr. 2018 Sep;72(9):1246–50.
7. Yárnoz-Esquiroz P, Olazarán L, Aguas-Ayesa M, Perdomo CM, García-Goñi M, Silva C, et al. ‘Obesities’: Position statement on a complex disease entity with multifaceted drivers. Eur J Clin Invest. 2022 Jul;52(7):e13811.
8. Swinburn B, Egger G, Raza F. Dissecting obesogenic environments: the development and application of a framework for identifying and prioritizing environmental interventions for obesity. Prev Med. 1999 Dec;29(6 Pt 1):563–70.
9. Burns AC, Whitacre PT eds. Perspectives from United Kingdom and United States policy makers on obesity prevention: Workshop summary. National Academies Press, 2010.
10. Parker L, Burns AC, Sanchez E, editors. Local Government Actions to Prevent Childhood Obesity. The National Academies Press. Washington (DC); 2009.
11. Stamatakis KA, Leatherdale ST, Marx CM, Yan Y, Colditz GA, Brownson RC. Where is obesity prevention on the map?: Distribution and predictors of local health department prevention activities in relation to county-level obesity prevalence in the United States. J Public Health Manag Pract. 2012 Oct;18(5):402–11.
12. Fan JX, Hanson HA, Zick CD, Brown BB, Kowaleski-Jones L, Smith KR. Geographic scale matters in detecting the relationship between neighbourhood food environments and obesity risk: an analysis of driver license records in Salt Lake County, Utah. BMJ Open. 2014 Aug 19;4(8):e005458.
13. Petersmann A, Müller-Wieland D, Müller UA, Landgraf R, Nauck M, Freckmann G, et al. Definition, classification and diagnosis of diabetes mellitus. Exp Clin Endocrinol Diabetes. 2019 Dec;127(S 01):S1–7.
14. Pearson-Stuttard J, Buckley J, Cicek M, Gregg EW. The changing nature of mortality and morbidity in patients with diabetes. Endocrinol Metab Clin North Am. 2021 Sep;50(3):357–68.
15. Chobot A, Górowska-Kowolik K, Sokołowska M, Jarosz-Chobot P. Obesity and diabetes-Not only a simple link between two epidemics. Diabetes Metab Res Rev. 2018 Oct;34(7):e3042.
16. Wilding JPH. Medication use for the treatment of diabetes in obese individuals. Diabetologia. 2018 Feb;61(2):265–72.

17. Schipf S, Ittermann T, Tamayo T, Holle R, Schunk M, Maier W, et al. Regional differences in the incidence of self-reported type 2 diabetes in Germany: results from five population-based studies in Germany (DIAB-CORE Consortium). J Epidemiol Community Health. 2014 Nov;68(11):1088–95.
18. Unwin N, Gan D, Whiting D. The IDF Diabetes Atlas: providing evidence, raising awareness and promoting action. Diabetes Res Clin Pract. 2010 Jan;87(1):2–3.
19. Nsubuga P, White ME, Thacker SB, Anderson MA, Blount SB, Broome CV, et al. Public health surveillance: a tool for targeting and monitoring interventions. Disease Control Priorities in Developing Countries 2nd edition. 2006;
20. Arita I, Nakane M, Nakano T. Surveillance of disease: Overview. International Encyclopedia Of Public Health. 2008;275.
21. Jaacks LM, Vandevijvere S, Pan A, McGowan CJ, Wallace C, Imamura F, et al. The obesity transition: stages of the global epidemic. Lancet Diabetes Endocrinol. 2019 Mar;7(3):231–40.
22. WHO. WHO STEPwise approach to surveillance [Internet]. [cited 2022 Apr 17]. Available from: https://www.euro.who.int/en/health-topics/noncommunicable-diseases/pages/monitoring-and-surveillance/tools-and-initiatives/who-stepwise-approach-to-surveillance
23. Robert Koch Institut. Data sources [Internet]. [cited 2022 May 26]. Available from: https://diabsurv.rki.de/Webs/Diabsurv/EN/project/methodology/data_sources/data_sources-node.html
24. LeBlanc AG, Jun Gao Y, McRae L, Pelletier C. At-a-glance – Twenty years of diabetes surveillance using the Canadian Chronic Disease Surveillance System. Health Promot Chronic Dis Prev Can. 2019 Nov;39(11):306–9.
25. Joshy G, Simmons D. Diabetes information systems: a rapidly emerging support for diabetes surveillance and care. Diabetes Technol Ther. 2006 Oct;8(5):587–97.
26. Robert Koch-Institut. Diabetes in Germany – adults. Years of life lost (YLL) [Internet]. [cited 2022 Aug 31]. Available from: https://diabsurv.rki.de/Webs/Diabsurv/EN/diabetes-in-germany/4-37_years_of_life_lost_YLL.html
27. Goodall R, Alazawi A, Hughes W, Bravis V, Salciccioli JD, Marshall DC, et al. Trends in type 2 diabetes mellitus disease burden in European Union countries between 1990 and 2019. Sci Rep. 2021;11(1):1–14.
28. König H, Rommel A, Baumert J, Schmidt C, König HH, Brettschneider C, et al. Excess costs of type 2 diabetes and their sociodemographic and clinical determinants: a cross-sectional study using data from the German Health Interview and Examination Survey for Adults (DEGS1). BMJ Open. 2021 Apr 21;11(4):e043944.
29. Khokhar B, Quan H, Kaplan GG, Butalia S, Rabi D. Exploring novel diabetes surveillance methods: a comparison of administrative, laboratory and pharmacy data case definitions using THIN. J Public Health (Oxf). 2018 Sep 1;40(3):652–8.
30. Reitzle L, Paprott R, Färber F, Heidemann C, Schmidt C, Thamm R, et al. Gesundheitsberichterstattung im Rahmen von Public Health Surveillance: Das Beispiel Diabetes. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz. 2020 Sep 1;63(9):1099–107.

31. Robert Koch Institut. Diabetes Surveillance [Internet]. [cited 2021 Dec 12]. Available from: https://diabsurv.rki.de/Webs/Diabsurv/EN/project/background/background-node.html
32. Gabrys L, Heidemann C, Baumert J, Teti A, Du Y, Paprott R, et al. Selecting and defining indicators for diabetes surveillance in Germany. J Health Monit. 2018 Jun;3(Suppl 3):3–21.
33. Bundesärztekammer (BÄK), Kassenärztliche Bundesvereinigung (KBV), Arbeitsgemeinschaft der Wissenschaftlichen, Medizinischen Fachgesellschaften (AWMF). Nationale VersorgungsLeitlinie Typ-2-Diabetes – Langfassung. [Internet]. [cited 2023 Sep 22]. Available from: www.leitlinien.de/diabetes
34. Brenner S, Oberaigner W, Stummer H. In guidelines physicians trust? Physician perspective on adherence to medical guidelines for type 2 diabetes mellitus. Heliyon. 2020 Aug;6(8):e04803.
35. Leitner DR, Fruhbeck G, Yumuk V, Schindler K, Micic D, Woodward E, et al. Obesity and type 2 diabetes: two diseases with a need for combined treatment strategies – EASO can lead the way. Obes Facts. 2017;10(5):483–92.
36. Dummer TJB. Health geography: supporting public health policy and planning. CMAJ. 2008 Apr 22;178(9):1177.
37. Augustin J, Andrees V, Walsh D, Reintjes R, Koller D. Spatial aspects of health – developing a conceptual framework. Int J Environ Res Public Health. 2023 Jan 19;20(3).
38. Bipu NA. Geographic information system & spatial analysis. Int J Empirical Educ Res. 2019; 3(18): 70-90.
39. Whitty CJM, Watt FM. Map clusters of diseases to tackle multimorbidity. Nature. 2020 Mar;579(7800):494–6.
40. Information Resources Management Association. Healthcare Policy and Reform: Concepts, Methodologies, Tools, and Applications. IGI Global; 2019.
41. Dekker LH, Rijnks RH, Navis GJ. Regional variation in type 2 diabetes: evidence from 137 820 adults on the role of neighbourhood body mass index. Eur J Public Health. 2020 Feb 1;30(1):189–94.
42. Maroko AR, Doan TM, Arno PS, Hubel M, Yi S, Viola D. Integrating social determinants of health with treatment and prevention: a new tool to assess local area deprivation. Prev Chronic Dis. 2016 Sep 15;13:E128.
43. Noble M, Wright G, Smith G, Dibben C. Measuring multiple deprivation at the small-area level. Environ Plan A. 2006 Jan 1;38(1):169–85.
44. Maier W, Holle R, Hunger M, Peters A, Meisinger C, Greiser KH, et al. The impact of regional deprivation and individual socio-economic status on the prevalence of Type 2 diabetes in Germany. A pooled analysis of five population-based studies. Diabet Med. 2013 Mar;30(3):e78-86.
45. esri. What is GIS? [Internet]. [cited 2022 Jan 8]. Available from: https://www.esri.com/en-us/what-is-gis/overview
46. Graser A, Olaya V. Processing: A Python framework for the seamless integration of geoprocessing tools in QGIS. ISPRS International Journal of Geo-Information. 2015;4(4):2219–45.
47. Davis CA, Fonseca FT. Assessing the certainty of locations produced by an address geocoding system. GeoInformatica. 2007 Mar 1;11(1):103–29.

48. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Geocoding and monitoring of US socioeconomic inequalities in mortality and cancer incidence: does the choice of area-based measure and geographic level matter?: the Public Health Disparities Geocoding Project. Am J Epidemiol. 2002 Sep 1;156(5):471–82.
49. McConnell R, Berhane K, Yao L, Lurmann FW, Avol E, Peters JM. Predicting residential ozone deficits from nearby traffic. Sci Total Environ. 2006;363(1–3):166–74.
50. Pittman J, Andrews H, Struening E. The use of zip coded population data in social area studies of service utilization. Eval Program Plann. 1986;9(4):309–17.
51. Howe HL. Geocoding NY State Cancer Registry. Am J Public Health. 1986 Dec;76(12):1459–60.
52. Cura R, Dumenieu B, Abadie N, Costes B, Perret J, Gribaudi M. Historical collaborative geocoding. ISPRS International Journal of Geo-Information. 2018;7(7).
53. Monir N, Abdul Rasam A, Ghazali R, Suhandri H, Cahyono A. Address geocoding services in geospatial-based epidemiological analysis: a comparative reliability for domestic disease mapping. International Journal of Geoinformatics. 2021;17(5).
54. Yeow, L.W.; Low, R.; Tan, Y.X.; Cheah, L. Point-of-Interest (POI) Data Validation Methods: An Urban Case Study. ISPRS Int. J. Geo-Inf. 2021, 10(11): 735. https://doi.org/10.3390/ijgi10110735
55. Skopeliti A, Stamou L. Online map services: contemporary cartography or a new cartographic culture? ISPRS Int J. Geo-Inf. 2019;8(5).
56. Peters M, Zeeb H. Availability of open data for spatial public health research. Ger Med Sci. 2022;20:Doc01.
57. esri. Identify data quality requirements. [cited 2022 Jul 23]. Identify data quality requirements. Available from: https://pro.arcgis.com/de/pro-app/2.8/help/data/validating-data/identify-data-quality-requirements.htm
58. Mobasheri A, Huang H, Degrossi LC, Zipf A. Enrichment of OpenStreetMap data completeness with sidewalk geometries using data mining techniques. Sensors. 2018;18(2).
59. He X, Zhang Z, Yang Z. Extraction of urban built-up area based on the fusion of night-time light data and point of interest data. R Soc Open Sci. 2021 Aug;8(8):210838.
60. Dwivedi P, Huang D, Yu W, Nguyen Q. Predicting geographical variation in health-related quality of life. Prev Med. 2019 Sep;126:105742.
61. Yang X, Ye T, Zhao N, Chen Q, Yue W, Qi J, et al. Population mapping with multisensor remote sensing images and point-of-interest data. Remote Sensing. 2019;11(5):574.
62. Blaschke, T.; Merschdorf, H.; Cabrera-Barona, P.; Gao, S.; Papadakis, E.; Kovacs-Györi, A. Place versus Space: From Points, Lines and Polygons in GIS to Place-Based Representations Reflecting Language and Culture. ISPRS Int. J. Geo-Inf. 2018, 7, 452. https://doi.org/10.3390/ijgi7110452.
63. Documentation QGIS. Vector Data [Internet]. [cited 2022 Aug 22]. Available from: https://docs.qgis.org/2.14/en/docs/gentle_gis_introduction/vector_data.html
64. Danielson JJ, Gesch DB. Global multi-resolution terrain elevation data 2010 (GMTED2010). US Department of the Interior, US Geological Survey; 2011.
65. Bayerische Vermessungsverwaltung. OpenData [Internet]. 2021 [cited 2021 Aug 11]. Available from: https://www.ldbv.bayern.de/produkte/weitere/opendata.html

66. Jilani, M, Padraig, C, Michela, B. "Automated quality improvement of road network in OpenStreetMap." Agile Workshop (action and interaction in volunteered geographic information). 2013. (p. 19).
67. Almendros-Jiménez JM, Becerra-Terón A, Torres M. Integrating and querying OpenStreetMap and linked Geo Open Data. The Computer Journal. 2019 Mar 1;62(3):321–45.
68. Andreopoulos B, An A, Wang X, Schroeder M. A roadmap of clustering algorithms: finding a match for a biomedical application. Brief Bioinform. 2009 May 1;10(3):297–314.
69. Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In kdd (Vol. 96, No. 34, pp. 226-231).
70. Vestal BE, Carlson NE, Ghosh D. Filtering spatial point patterns using Kernel densities. Spat Stat. 2021 Mar;41:100487.
71. Fortin, MJ. Spatial structure in population data. *Population ecology in practice*. Oxford: Wiley Blackwell, 2020. 299-314.
72. Quiñones S, Goyal A, Ahmed ZU. Geographically weighted machine learning model for untangling spatial heterogeneity of type 2 diabetes mellitus (T2D) prevalence in the USA. Sci Rep. 2021 Mar 26;11(1):6955.
72. Griffin SJ, Little PS, Hales CN, Kinmonth AL, Wareham NJ. Diabetes risk score: towards earlier detection of type 2 diabetes in general practice. Diabetes Metab Res Rev. 2000 Jun;16(3):164–71.
74. Janghorbani M, Adineh H, Amini M. Evaluation of the Finnish Diabetes Risk Score (FINDRISC) as a screening tool for the metabolic syndrome. Rev Diabet Stud. 2013 Winter;10(4):283–92.
75. R Core Team (2016). R: A Language and Environment for Statistical Computing, Vienna, Austria.
76. Bivand R. CRAN Task View: Analysis of Spatial Data [Internet]. [cited 2021 Dec 5]. Available from: https://cran.r-project.org/web/views/Spatial.html
77. Bivand RS, Pebesma EJ, Gómez-Rubio V, Pebesma EJ. Applied Spatial Data Analysis with R. Vol. 747248717. Springer; 2008.
78. Place Search. Places API for Web. Google Developers [Internet]. 2019 [cited 2019 Feb 21]. Available from: https://developers.google.com/places/web-service/search
79. Manuel J. A. Eugster and Thomas Schlesinger. osmar: OpenStreetMap and R. R Journal, 2012. Accepted for publication on 2012-08-14.
Eugster, MJA, Schlesinger, T. osmar: OpenStreetMap and R. The R Journal (2013) 5:1, pages 53-63.
80. Overpass turbo [Internet]. 2021 [cited 2021 Aug 11]. Available from: https://overpass-turbo.eu/
81. Gortmaker SL, Swinburn BA, Levy D, Carter R, Mabry PL, Finegood DT, et al. Changing the future of obesity: science, policy, and action. Lancet. 2011 Aug 27;378(9793):838–47.
82. Springmann M, Wiebe K, Mason-D'Croz D, Sulser TB, Rayner M, Scarborough P. Health and nutritional aspects of sustainable diet strategies and their association with

environmental impacts: a global modelling analysis with country-level detail. Lancet Planet Health. 2018 Oct;2(10):e451–61.
83. Cuadros DF, Li J, Musuka G, Awad SF. Spatial epidemiology of diabetes: Methods and insights. World J Diabetes. 2021 Jul 15;12(7):1042–56.
84. Tseng E, Zhang A, Shogbesan O, Gudzune KA, Wilson RF, Kharrazi H, et al. Effectiveness of policies and programs to combat adult obesity: a systematic review. J Gen Intern Med. 2018 Nov;33(11):1990–2001.
85. GitHub - MAPraeger/GOcode [Internet]. 2021 [cited 2021 Dec 23]. Available from: https://github.com/MAPraeger/GOcode

2 Summary

2.1 English summary

Background

Obesity is a severe global health-related issue. Many premature deaths are caused by the health condition. The term obesogenic environment describes the influence of environmental factors on the development of obesity. Diabetes is a severe disease and closely related to obesity. Online geocoding services can be used to generate data for visualizing and calculating obesity risks. The aim of the work is to develop an obesity risk tool based on these data which can be applied for diabetes surveillance.

Methods

In our first study, the feasibility of the use of online geocoding systems for diabetes surveillance was tested. As a first step, an extensive literature review was executed to discover environmental factors that are related to obesity. Subsequently, the transferability of these factors into pre-defined variables of online geocoding services such as OpenStreetMap (OSM) was tested. The corresponding spatial data points were downloaded for the pilot areas and visualized on maps. We determined the quality of our data sources within field validations. In a second study, we developed a spatial obesity risk score (SORS) based on kernel density estimations (KDE), which uses the same data-collecting mechanism. For this purpose, we first tested the spatial heterogeneity of the raw data using Ripley's bivariate K function together with uncertainty. Our risk score was based on subtracting negative risks from positive risks. Thus, we were able to generate maps showing a composite picture of high- and low-risk areas based on the score values. Baseline parameters were chosen and varied in a deterministic sensitivity analysis afterward. We used a probabilistic approach to determine suitable parameter choices and to identify their sensitivity. Finally, we compared the KDE-based map with an alternative intensity-based map.

Results

We identified several studies in our literature search that describe factors that were positively or negatively correlated with obesity. Respective data could be operationalized, downloaded from online geocoding services, and further processed. Additionally, the data points could be shown on maps. During the validations in the field, Google Maps showed higher sensitivities, whereas OSM had a higher positive predictive value on average across the pilot areas. In the second study, the spatial heterogeneity was significant and underlined the importance of the subsequent risk score approach. In addition, we were able to show ex-ante densities separately for the positive and negative environmental factors. The risk score showed some dense regions in our study area indicating an increased and lowered risk for the development of obesity, whereas the alternative intensity-based methods were rather able to identify some protective structures The bandwidth and the amount of edge correction were the model parameters with the greatest impact on the results.

Conclusion

Results showed reasonable data quality for the online geocoding services under consideration. The risk score made it possible to gain a composite picture of risk patterns within the study areas. Furthermore, the automation process made it possible to extend the analyses to other and even larger geographic areas, given that the respective data were downloaded and the code adapted to the desired location. Model parameters should be chosen carefully, especially those with a high impact on risk score values. The tool made it possible for policy makers to generate a simpler and more intuitive interpretation of the available data. Further studies should examine the SORS with real world data. In all, we were able to show that the tool can possibly be integrated into diabetes surveillance.

2.2 Deutsche Zusammenfassung

Hintergrund

Übergewicht ist ein schwerwiegendes weltweites Gesundheitsproblem. Viele vorzeitige Todesfälle gehen auf diesen Gesundheitszustand zurück. Der Ausdruck „Obesogenic environment“ beschreibt den Einfluss von Umweltfaktoren auf die Entwicklung von Übergewicht. Diabetes ist eine schwerwiegende Erkrankung und steht eng mit Übergewicht in Beziehung. Online-Geokodierungsdienste können verwendet werden, um Daten für die Visualisierung und zur Berechnung von Risiken für Übergewicht zu erzeugen. Das Ziel der Arbeit ist es auf Basis dieser Daten ein Risiko-Tool für Übergewicht zu entwickeln, das in der Diabetes Surveillance eingesetzt werden kann.

Methoden

In unserer ersten Studie wurde die Machbarkeit der Verwendung von Online-Geokodierungsdiensten für die Diabetes-Surveillance getestet. Zunächst wurde eine ausführliche Literaturrecherche ausgeführt, um Faktoren zu ermitteln, die mit Übergewicht assoziiert sind. Im Anschluss wurde getestet, inwieweit sich diese Faktoren auf vordefinierte Variablen der Online-Geokodierungsdienste wie beispielsweise OpenStreetMap (OSM) abbilden lassen. Die entsprechenden raumbezogenen Datenpunkte wurden für das Pilotgebiet heruntergeladen und auf Karten visualisiert. Mittels Validierungen vor Ort wurde die Datenqualität bestimmt. In einer zweiten Studie wurde ein auf Übergewicht abzielender, räumlicher Risikoscore (Spatial Obesity Risk Score, SORS) entwickelt, der auf Kerndichteschätzungen und auf der gleichen Methode der Datengenerierung basiert. Zu diesem Zweck testeten wir zunächst auf räumliche Heterogenität der Rohdaten mittels Ripleys bivariater K-Funktion, die um Unsicherheit erweitert wurde. Unser Risiko-Score wurde durch die Subtraktion des positiven Risikos vom negativen Risiko ermittelt. Folglich war es uns möglich ein zusammengesetztes Bild durch Karten mit hohen und niedrigen Risikos mithilfe der Scores aufzuzeigen. Basisparameter wurden gewählt und in deterministischen Sensitivitätsanalysen verändert. Wir nutzten einen probabilistischen Ansatz, um geeignete

Parameter zu bestimmen und deren Sensitivität zu identifizieren. Schließlich verglichen wir die Visualisierung des Risiko-Scores mit einer alternativen Methode auf Basis von Intensitäten.

Ergebnisse

Im Rahmen unserer Literaturrecherche wurden mehrere Studien identifiziert, in denen die behandelten Faktoren positiv oder negativ mit Übergewicht korreliert waren. Die entsprechenden Daten konnten operationalisiert, von Online-Gekodierungsdiensten heruntergeladen und weiterverarbeitet werden. Zusätzlich konnten die Datenpunkte auf Karten dargestellt werden. Im Rahmen der Validierungen vor Ort zeigte Google Maps höhere Sensitivitäten, wohingegen OSM durchschnittlich einen höheren positiven prädiktiven Wert über die Studiengebiete hinweg hatte. In der zweiten Studie war die räumliche Heterogenität signifikant und unterstrich die Bedeutung des nachfolgenden Risiko-Score-Ansatzes. Darüber hinaus konnten wir ex-ante Dichten separat für positive und negative Umweltfaktoren aufzeigen. Der Risiko-Score zeigte dichte Regionen, die auf erhöhtes und niedrigeres Risiko für Übergewicht hinwiesen, während die alternative intensitätsbasierte Methode eher die protektiven Strukturen nachweisen konnte. Die Bandweite und die Randkorrektur waren die beiden Parameter mit dem höchsten Einfluss auf die Ergebnisse.

Schlussfolgerung

Die Ergebnisse zeigten eine angemessene Datenqualität der betrachteten Online-Geokodierungsdienste. Der Risiko-Score ermöglichte ein zusammengesetztes Bild von Risikopatterns in den Studiengebieten. Außerdem war es möglich mittels Automatisierung die Analyse auf andere und sogar größere Gebiete auszuweiten. Dafür war es nötig entsprechende Daten herunterzuladen und den Code auf die Ziellokation anzupassen. Modellparameter sollten mit Bedacht gewählt werden, besonders die Parameter mit großem Einfluss auf die Risiko-Scores. Das Tool ermöglichte für politische Entscheidungsträger eine einfachere und intuitivere Interpretation der vorliegenden Daten. Weitere Studien sollten den SORS mit Real-World-Daten untersuchen. Insgesamt konnte gezeigt werden, dass das Tool im Rahmen der Diabetes Surveillance eingesetzt werden kann.

3 Articles

3.1 Article 1: Using data from online geocoding services for the assessment of environmental obesogenic factors: a feasibility study

Published in:

Int J Health Geogr (2019) 18:13, https://doi.org/10.1186/s12942-019-0177-9

Maximilian Präger, Christoph Kurz, Julian Böhm, Michael Laxy, Werner Maier.

Präger *et al. Int J Health Geogr* (2019) 18:13
https://doi.org/10.1186/s12942-019-0177-9

International Journal of
Health Geographics

RESEARCH

Open Access

Using data from online geocoding services for the assessment of environmental obesogenic factors: a feasibility study

Maximilian Präger[1,2], Christoph Kurz[1,2], Julian Böhm[1,2], Michael Laxy[1,2] and Werner Maier[1,2*]

Abstract

Background: The increasing prevalence of obesity is a major public health problem in many countries. Built environment factors are known to be associated with obesity, which is an important risk factor for type 2 diabetes. Online geocoding services could be used to identify regions with a high concentration of obesogenic factors. The aim of our study was to examine the feasibility of integrating information from online geocoding services for the assessment of obesogenic environments.

Methods: We identified environmental factors associated with obesity from the literature and translated these factors into variables from the online geocoding services Google Maps and OpenStreetMap (OSM). We tested whether spatial data points can be downloaded from these services and processed and visualized on maps. True- and false-positive values, false-negative values, sensitivities and positive predictive values of the processed data were determined using search engines and in-field inspections within four pilot areas in Bavaria, Germany.

Results: Several environmental factors could be identified from the literature that were either positively or negatively correlated with weight outcomes in previous studies. The diversity of query variables was higher in OSM compared with Google Maps. In each pilot area, query results from Google showed a higher absolute number of true-positive hits and of false-positive hits, but a lower number of false-negative hits during the validation process. The positive predictive value of database hits was higher in OSM and ranged between 81 and 100% compared with a range of 63–89% for Google Maps. In contrast, sensitivities were higher in Google Maps (between 59 and 98%) than in OSM (between 20 and 64%).

Conclusions: It was possible to operationalize obesogenic factors identified from the literature with data and variables available from geocoding services. The validity of Google Maps and OSM was reasonable. The assessment of environmental obesogenic factors via geocoding services could potentially be applied in diabetes surveillance.

Keywords: Obesogenic environment, Geocoding services, Validation, Diabetes

*Correspondence: werner.maier@helmholtz-muenchen.de
[1] Institute of Health Economics and Health Care Management, Helmholtz Zentrum München – German Research Center for Environmental Health (GmbH), Ingolstädter Landstraße 1, 85758 Neuherberg, Germany
Full list of author information is available at the end of the article

Background

Obesity, commonly defined as a body mass index (BMI) of $\geq 30\ kg/m^2$ in adults [1], is the result of a complex multifactorial relationship (e.g. genetic, socioeconomic, and cultural factors) [2]. The prevalence of obesity is affected by lifestyle habits, consumption patterns as well as the urban development [2]. Since the 1980s, the prevalence of obesity has risen considerably and doubled in many countries [3]. Furthermore, a high BMI seems to be associated with a significant proportion of mortality and disability cases [4, 5]. Obesity is therefore recognized as a serious worldwide epidemic.

A number of severe health conditions are correlated with being very overweight, e.g. cardiovascular disease and hypertension, but in particular type 2 diabetes mellitus (T2DM) [6], which is the second leading cause of BMI-related deaths in 2015 [4]. Furthermore, obesity and overweight are the single most relevant predictors for T2DM [7]. Because some studies revealed the simultaneous spread of obesity and diabetes, the term 'diabesity' has been used in the literature in order to illustrate the close connectedness [8].

The built environment, comprising buildings, spaces and products generated or influenced by humans, has a strong influence on promoting or preventing diseases [9, 10]. The built environment can act on three different scales: the macro level describes the sprawl or the compactness of a region on a higher aggregated level, e.g. at the nationwide level, whereas the meso level is concerned with the community or neighbourhood environment, in which the access to certain facilities is of major interest. The micro level constitutes a person-related perspective, for example regarding qualities of urban design, and is often connected with the concept of walkability [11]. Factors of the built environment may contribute to obesity, for example via the availability of unhealthy food or the absence of green spaces [12], and consequently create obesogenic environments. Following Swinburn and colleagues [13], obesogenic environments can be described as 'the sum of influences that the surroundings, opportunities, or conditions of life have on promoting obesity in individuals or populations'.

In order to evaluate features of the built environment, tools based on the use of geographic information systems (GIS) have been developed using remote sensing techniques applicable as desk-based approaches [14]. In the past, researchers have shown great interest in commercial data within GIS-based analyses [15, 16]. Recently, freely available data from online geocoding services such as Google Maps and OpenStreetMap (OSM) have become increasingly popular [17, 18]. These services are often accessed via embedded application programming interfaces (APIs) to search data within the geographical databases, e.g. for food-related data [19]. These freely available data can be further applied to assess the environmental risk of the development of obesity by describing high- and low-risk geographical areas originating from the accumulation of obesogenic and protective environmental factors [20]. Further applications of such data could refer to environmental pollution or geographical access to primary health care [21, 22].

The aim of our study was to examine the feasibility of integrating information from online geocoding services into the assessment of environmental obesogenic factors which could potentially be used for diabetes surveillance. Diabetes risk has often been estimated e.g. using data from national surveys, but also from administrative data [23]. Thus secondary data from online geocoding services could be a potential complementary data source for diabetes surveillance. Considering this, two steps were required: First, we checked whether obesogenic and protective factors can be derived from the literature and translated into variables from online geocoding services. Second, we compared Google Maps and OSM regarding their validity and reliability of queried data.

Methods

Design of the validation process

To prepare subsequent validations, we initially identified environmental factors correlated with obesity from the literature. Based on these results and on expert discussions, we have chosen variables from Google Maps and OSM and downloaded these for four regions in Bavaria, Germany. Subsequently, these downloaded data points were validated in the field and by using search engines. An overview of the methods applied during the two phases of preparation and validation is shown in Fig. 1, and further details are provided below.

Literature search and extraction of variables

We applied a search strategy within PubMed using the search terms 'obesogenic', 'environmental factors', 'systematic' and 'review'. After screening the results, two reviews were determined to be relevant for our analysis. The first review by Mackenbach and colleagues [24] provided a systematic search strategy and identified correlates of environmental factors with obesity. The second publication was a review of GIS methods by Jia and colleagues [25], in which correlations of variables with weight status and obesity were described. Following Mackenbach et al. [24], we created a table in order to summarize the factors from our literature search. In a first step, we extracted environmental factors from the studies covered by the two reviews. In a second step, we grouped the publications describing these environmental factors and extracted and summarized information from

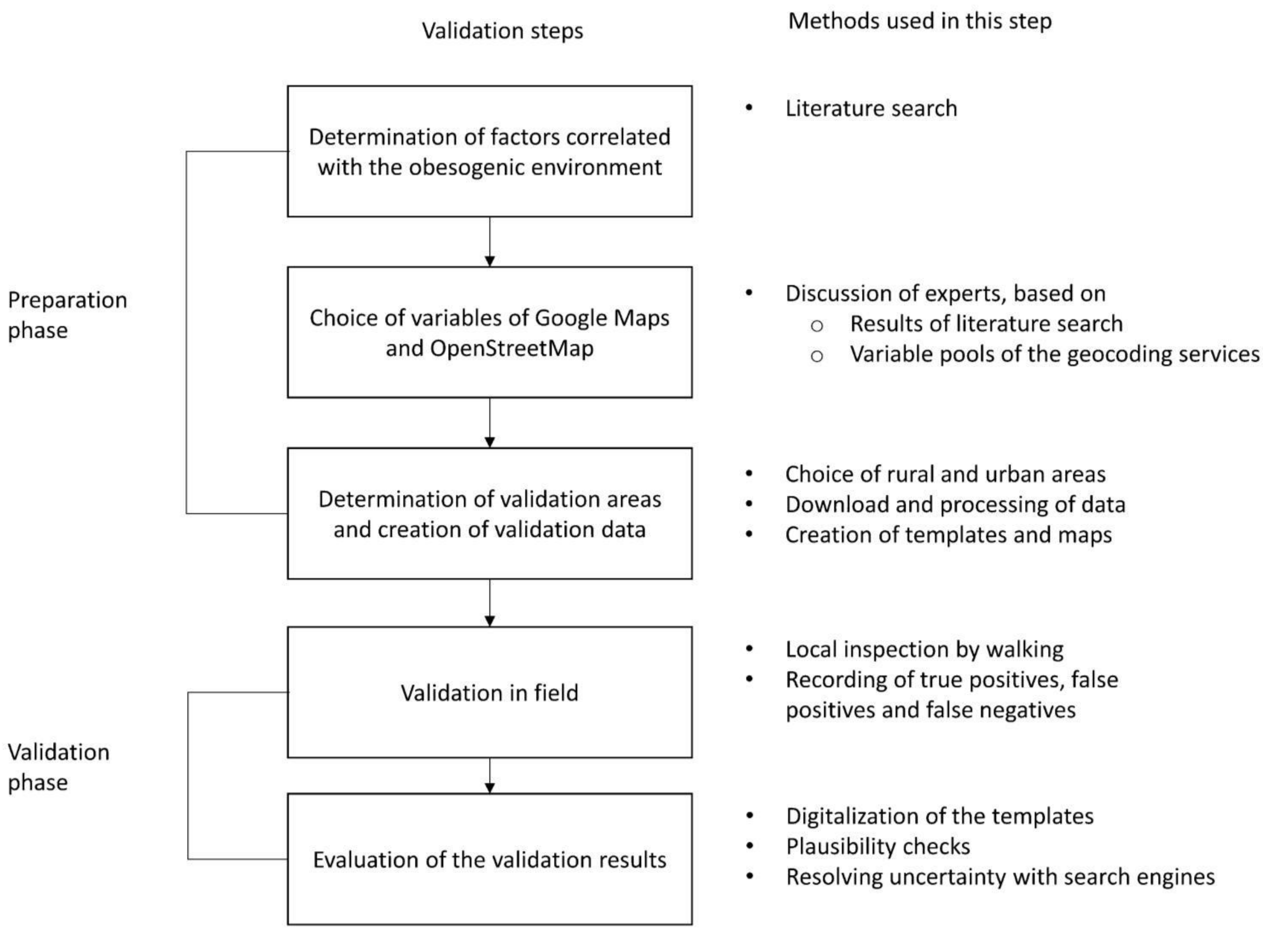

Fig. 1 Flow chart of the validation process

these publications in order to determine their correlation with obesity.

Subsequently, we extended the systematic search strategy provided by Mackenbach et al. in order to identify recent additional studies within PubMed, EMBASE, Web of Science, Cochrane Library, PsychInfo and Google Scholar. We completed the variable table with the additionally identified publications, and information from these studies was used to update the correlations of the environmental factors.

Definition of correlation

For each given environmental factor, we summed up the numbers of studies describing a positive and significant correlation with obesity. Analogously, we counted the numbers of studies describing negative and significant correlations with obesity for the same given factor. Subsequently, we defined this factor as overall positively correlated if at least three publications could be found and if the ratio of the number of positive correlations for the factor divided by the number of negative correlations for the same factor was 2 or higher. Dividing by 0 in this sense can be interpreted as causing infinity. Studies showing no significant correlation were not taken into account. Analogously, if the number of negative correlations divided by the number of positive correlations equals 2 or more, we assumed the factor to be overall negatively correlated. Otherwise, we supposed that no association existed. For example, if a factor was described with a positive correlation in five publications and with a negative correlation in 12 publications, an overall negative correlation was assumed as $12/5 \geq 2$. This calculation procedure was performed for each extracted environmental factor.

Determining the variables from the geocoding services

We checked environmental factors identified within the literature search regarding mapping possibilities with variables from Google Maps and OSM. Google Maps data, among other sources, are derived from official registries, e.g. from the Agency for Digitisation, High-Speed Internet and Surveying in Bavaria [26, 27]. OpenStreetMap, in contrast, is based on volunteered geographical information (VGI), i.e. it is based on user-generated content [28]. We have chosen both geocoding services because their data were freely available at low cost. Furthermore, their accuracy has been investigated for Germany in the past. Apparently, Google Maps showed higher completeness and higher precision of coordinates than OSM [17]. Besides environmental obesogenic factors, additional variables concerning the regional healthcare structure were taken into account. Four researchers in our team independently rated the relevance of the variables from the geocoding services with respect to the results of the literature search. After discussion, the variables best operationalizing the identified factors from the literature were determined and downloaded from Google Maps and OSM. We focused our analysis on single points of interest (POIs). Therefore, complex variable constructs, such as 'neighbourhood walkability' and 'land use mix', were not considered, as these compound measures are based e.g. on residential density or numbers of developed hectares which cannot be directly derived from online geocoding services. For an overview on the composites of these variables see Feng et al. [29]. Furthermore, six broader categories, 'food', 'doctor', 'sport', 'education', 'transport' and 'other', were determined via expert discussions within our team, and each operationalized variable, for which POIs were returned by at least one geocoding service, was assigned to one of those categories. Based on this approach, it was possible to visualize the distribution of environmental factors on a higher aggregated level and improve interpretability of the field validation results.

Choosing locations for the validation process

We have chosen four pilot areas in the German federal state of Bavaria for the validation process. Our aim was to investigate the data quality of the geocoding services within regions of different population density and urbanization level. Size and population count of each area were derived from the German Federal Statistical Office and the statistical offices of the German Länder (federal states) [30]. The first region was a sparsely populated municipality in the south-west of Bavaria with fewer than 2000 inhabitants encompassing an area of about 36.31 km^2. This area constituted a rural region containing few amenities (Area A). The second area was a street in a medium-sized major district town near Munich, the capital of Bavaria, with fewer than 45,000 inhabitants and a size of around 34.96 km^2 (Area B). Finally, the densely populated city of Munich (about 1.5 million inhabitants, total area 310.71 km^2) was selected for the validation. From the whole city of Munich, both a denser area close to the city centre and an area with a relatively lower density of amenities was chosen (Areas C and D).

Database extraction and processing of data

Google Maps data were downloaded using queries in uniform resource locator (URL) format targeting the Google Places API. Furthermore, OSM queries were performed using a web interface and an OSM-based R package. The geographical database returns data in JavaScript Object Notation (JSON), GeoJSON or osmar ('OpenStreetMap and R') format, which are standard representations for geographical data. Based on the structure of these data formats, information for each of the single POIs can be accessed efficiently via a hierarchical structure and subsequently processed. For the Google Maps results, each entry had to be queried again in order to get additional relevant information, e.g. on names, addresses and categorizations. Using the downloaded OSM data, additional information could be extracted directly from the previously described data formats without any additional query. An overview of the data formats and query possibilities is shown in Fig. 2. The return of the spatial databases was checked regarding consistency and plausibility. Important examinations were identifying POIs that were counted twice or more because of being listed within different categories and checking whether the return of the database lies completely within the pre-specified search area. Additionally, spatial POIs were visualized on maps in order to check coherence. The geographical data points were marked according to their factor category, and the search area was also plotted. An example of visualization of some factors for Area D can be found in Fig. 3 for OSM. The underlying code and the other codes regarding Area D are available on github [https://github.com/MAPraeger/GOcode. Accessed 23 April 2019].

Search area and download capacity

The shape of the downloaded regions was predefined by the geocoding services. OSM areas were rectangular, whereas Google Maps areas were circular. In order to make the shapes of OSM and Google Maps queries more comparable, we defined OSM search regions as quadratic. Further differences between the geocoding services affected the maximum downloadable data size. At the time of data download in 2017, Google Maps allowed up to 200 results per query and 1000 queries per day per person at zero costs [31], whereas OSM had fewer restrictions [32]. Depending on the API and the download tools used, areas of arbitrary size, whole so-called 'planet files' [33] or nearly arbitrary data sizes caused by memory overload within the statistical software, could be downloaded. Therefore, areas for the validation process were determined such that none of the above-mentioned restrictions took effect. Owing to the lower number of spatial POIs within the rural region (Area A), a wider area containing the whole municipality was chosen compared

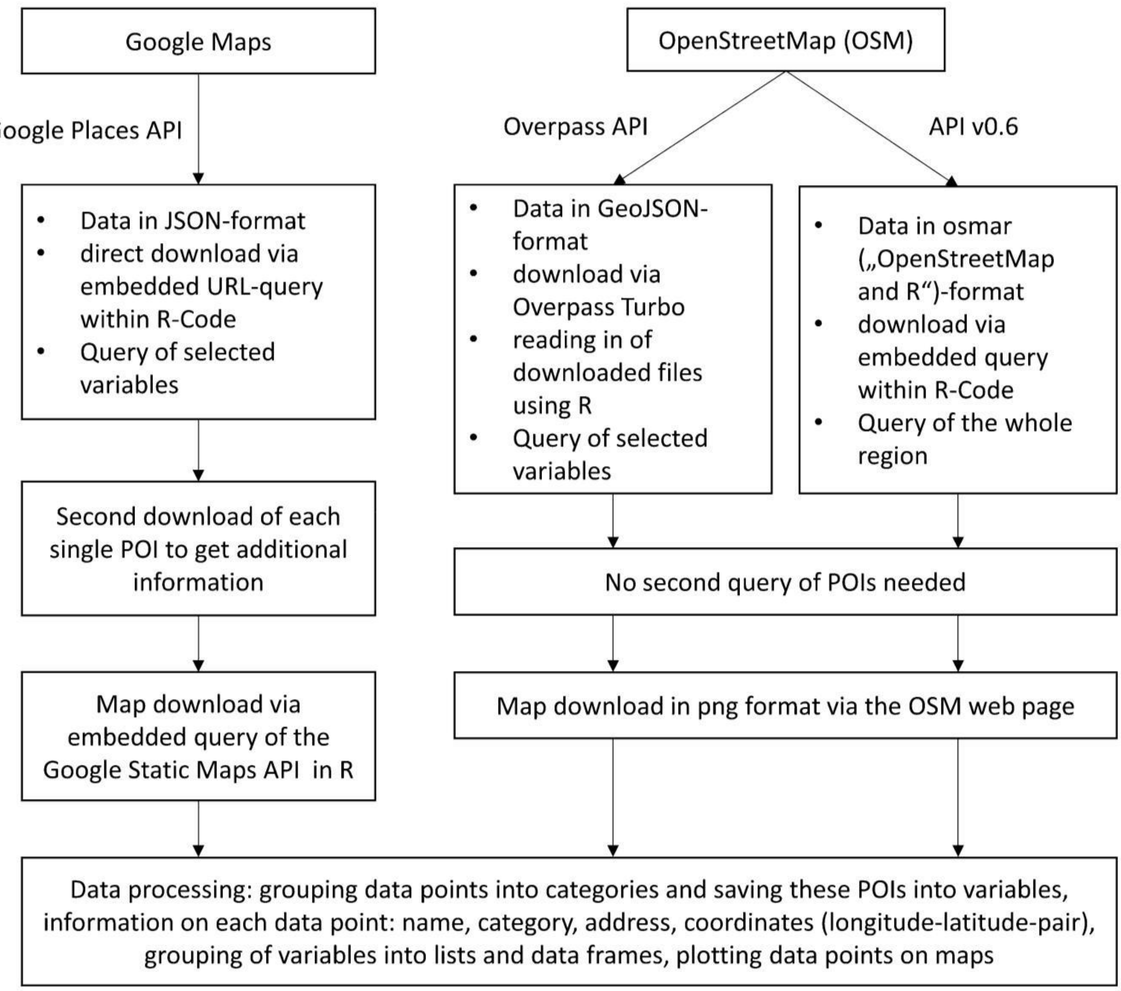

Fig. 2 Overview on the data and query structure. *JSON* JavaScript Object Notation, *API* application programming interface, *POI* point of interest

with the more urban areas (Areas B–D), for which the diameters of the circles and the edges of the squares were set to 200 metres.

Validation process

Four researchers in our team locally scanned the predefined validation areas looking for the existence of the downloaded POIs of Google Maps and OSM. We designed a template to standardize the recording process and used maps containing the data points to improve efficiency. The number of returned POIs of a database was called 'hits'. Each researcher documented the validation date, confirmation (true positive hit) or rejection (false positive hit) of existence of the POIs and new record of false negatives, i.e. data points discovered in the field that were not covered by Google Maps or OSM or both. After completion, the templates were digitalized.

If uncertainties regarding the existence of a POI were present during validation in the field, the researchers recorded their comments. If these notes indicated restrictions, e.g. regarding access to certain facilities during in-field validations, several online search engines were used to resolve these uncertainties. Further examples were incorrect categorization or implausible numbers of false positives at a certain place. To overcome these issues, we visited the home pages of the affected amenities and considered business directories (yellow pages).

Common summary statistics for the validation of geographical data points were calculated. For the quality assessment of the performance of a geocoding service for a given area, sensitivities, i.e. true positives divided by the sum of true positives and false negatives, and positive predictive values (PPVs), i.e. true positives divided by the sum of true positives and false positives, were calculated [34, 35].

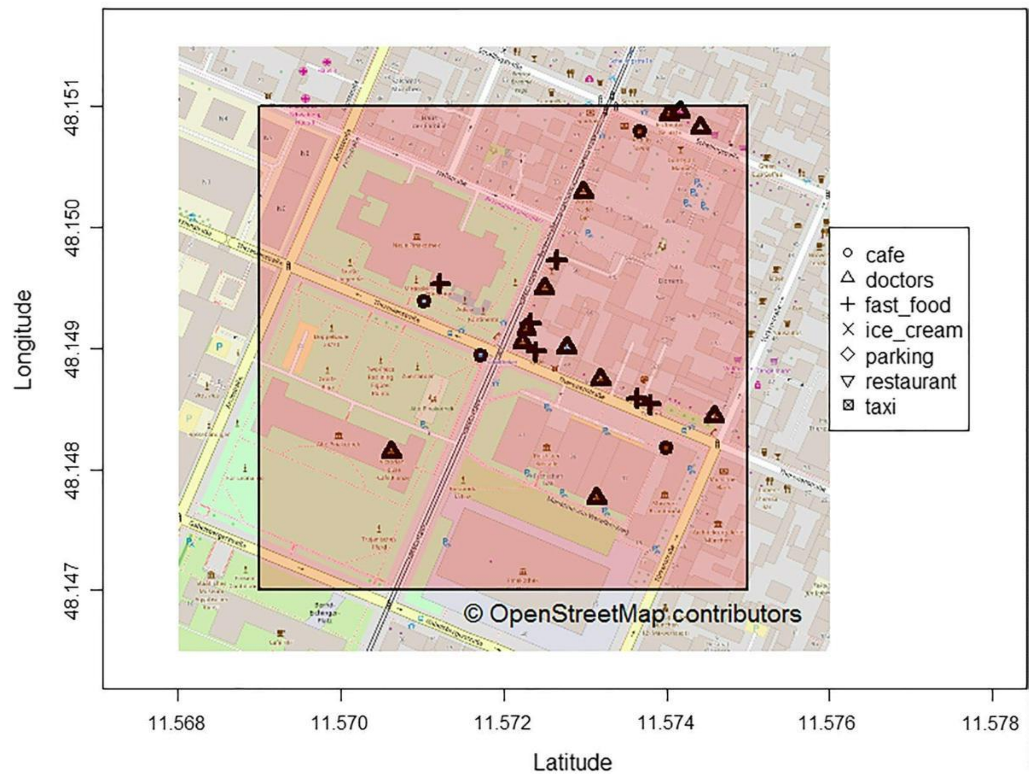

Fig. 3 Example of visualization of OpenStreetMap data points (Area D)

Software

We used the free software environment R, version 3.3.2, to implement code targeting the Google Places API via embedded URL query and for processing of the query results [36]. In order to download data from OSM, we applied an online tool for data filtering (Overpass Turbo) and the R package 'osmar' [37, 38]. For data processing, we used the packages 'geojsonR', 'jsonlite' and 'rgdal' and, for data visualization, the R packages 'ggmap' and 'ggplot2' [39–43].

Results

Literature search

An extensive list of environmental factors and the corresponding references (N = 256) can be found within Additional file 1: Table S1. The table contains the numbers of studies describing positive correlations, negative correlations and studies without significant associations for a given environmental factor. According to the definition of correlation within the methods section, overall positive correlations with weight status were discovered for the variables 'fast food', 'food retail', 'unhealthy food outlets', 'convenience store', 'rural areas', 'urban sprawl', 'county sprawl', 'traffic', 'transport' and 'poverty'. Overall negative correlations were found for the variables '(healthy) food outlets', 'restaurants', 'supermarkets', 'tree cover', 'fitness or physical activity facilities', 'forests', 'greenspace', 'longer way to school', 'open space', 'outdoor recreation', 'park', 'recreation centre', 'walkability', 'aesthetics', 'intersection density', 'land use mix', 'population density', 'safety', 'sidewalk completeness', 'street connectivity', 'education' and 'physician supply'.

Chosen variables from Google Maps and OSM

Tables 1 and 2 show the factors from Google Maps (N = 25 in total) and OSM (N = 126 in total) chosen for the validation process. Owing to the extent of the OSM variable pool, the relevant factors in the category

Table 1 Selected variables from the Google Maps pool

Bakery	Bar	Bus station	Cafe
Convenience store	Dentist	Doctor	Food
Grocery or supermarket	Gym	Hospital	Meal delivery
Meal takeaway	Park	Pharmacy	Physiotherapist
Restaurant	School	Spa	Stadium
Subway station	Taxi stand	Train station	Transit station
University			

Table 2 Selected OpenStreetMap (OSM) variables in the category 'amenity'

Bar	Bbq	Biergarten	Cafe	Fast food
Food court	Ice cream	Pub	Restaurant	College
School	Bicycle parking	Bicycle rental	Boat sharing	Bus station
Taxi	Clinic	Dentist	Doctors	Hospital
Nursing home	Pharmacy	Dive centre	Dojo	Ranger station
Beach resort	Dance	Fishing	Fitness centre	Garden
Golf course	Ice rink	Nature reserve	Park	Pitch
Playground	Sports centre	Stadium	Swimming area	Swimming pool
Track	Water park			

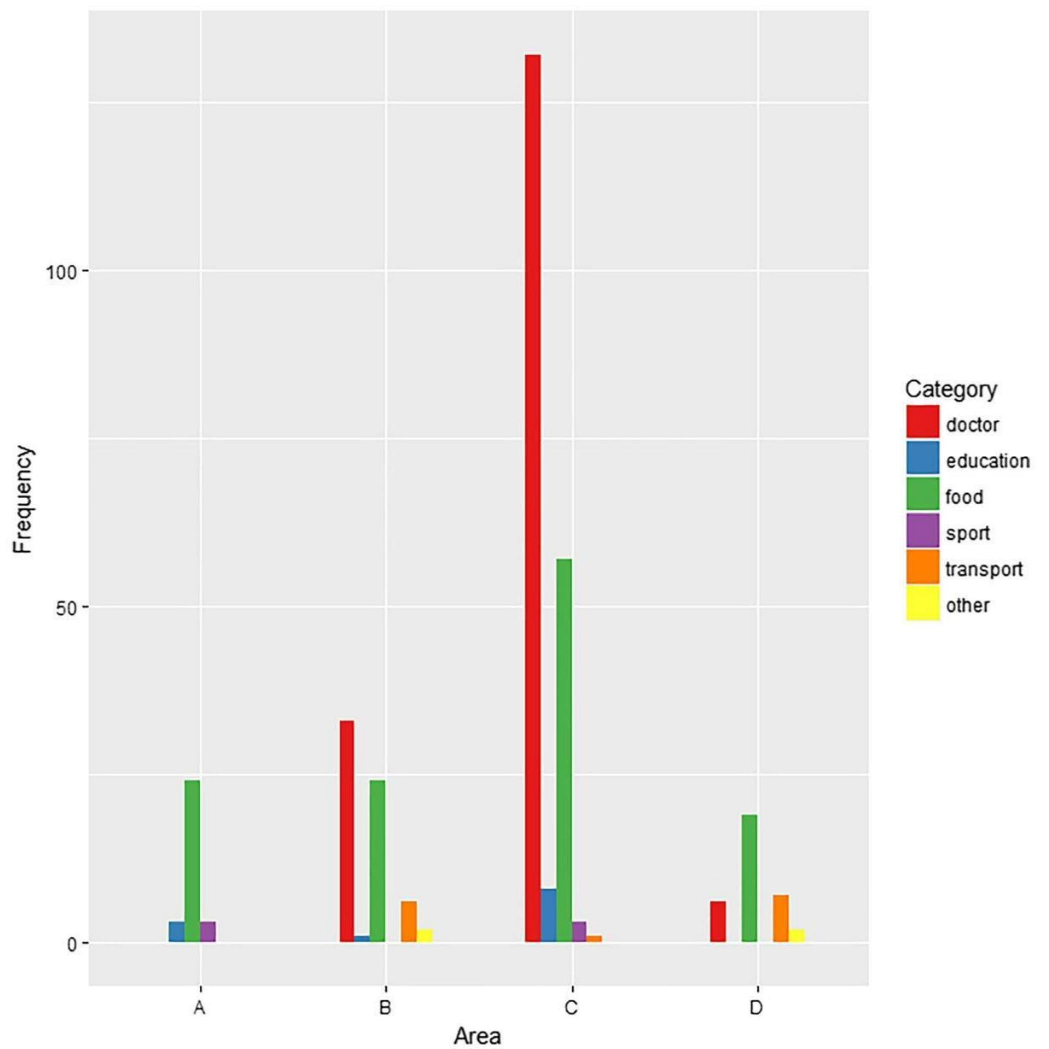

Fig. 4 Distribution of hits across variable categories using Google Maps. Area A: sparsely populated municipality in the south-west of Bavaria. Area B: street in a medium-sized populated major district town near Munich. Area C: area close to the centre within the densely populated city of Munich. Area D: area with a lower density of amenities within the densely populated city of Munich

'amenity' are shown within Table 2 (N=42). The full list of OSM variables is shown in Additional file 2: Table S2.

Distribution of database results across categories

Bar charts are shown for Google Maps (Fig. 4) and OSM (Fig. 5) in order to visualize the distribution of the database hits, i.e. the distribution of the sum of true-positive and false-positive entries of the geocoding services, across the six categories of 'doctor', 'education', 'food', 'sport', 'transport' and 'other' for the validation areas. Within the medium-sized populated Area B and the densely populated Area C, predominantly entries in the categories 'doctor' and 'food' account for most of the database hits in Google Maps. For the remaining areas, using Google Maps, 'food' was the most relevant category. Regarding OSM, the category 'food' was the most frequent category within Areas C and D of the city of Munich.

Validations

Tables 3 and 4 show the numbers of true positives and false positives, PPVs, numbers of false negatives and sensitivity values for each validation area. As shown in the table, absolute numbers of true hits were higher for Google Maps than the corresponding numbers for OSM, irrespective of the validation area under consideration. Furthermore, false positives were also higher for Google Maps compared with OSM. The PPVs of OSM hits, ranging between 81 and 100%, were higher than the PPVs of Google Maps hits, which were found to be between 65

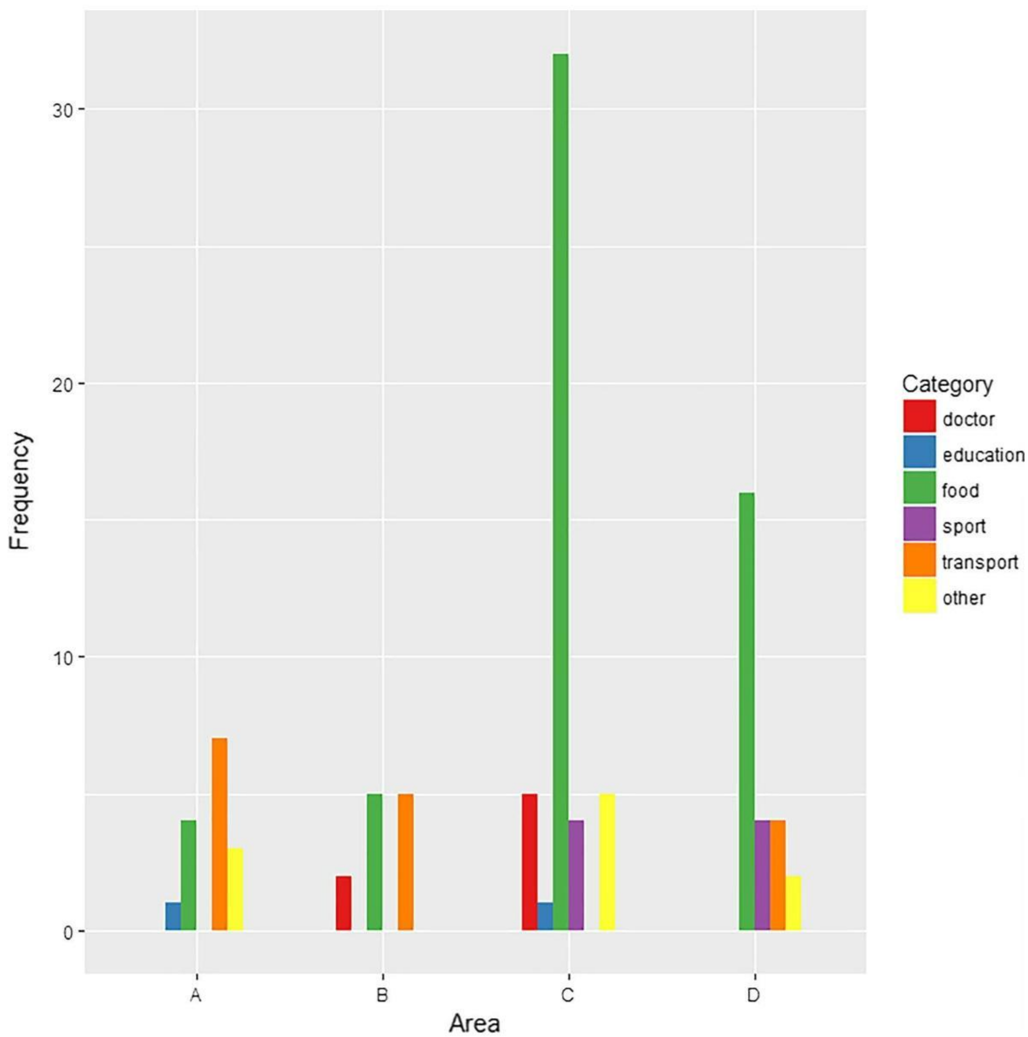

Fig. 5 Distribution of hits across variable categories using OpenStreetMap. Area A: sparsely populated municipality in the south-west of Bavaria. Area B: street in a medium-sized populated major district town near Munich. Area C: area close to the centre within the densely populated city of Munich. Area D: area with a lower density of amenities within the densely populated city of Munich

Table 3 Results of the field validation

Area	Geocoding service	True positives: N (% positive)[a]	False positives: N (% positive)	False negatives: N	Sensitivity[b]: %
A	Google Maps	19 (63.33)	11 (36.67)	13	59.38
A	OpenStreetMap	15 (88.24)	2 (11.76)	17	46.88
B	Google Maps	58 (89.23)	7 (10.77)	1	98.31
B	OpenStreetMap	12 (100)	0 (0)	47	20.34
C	Google Maps	144 (71.64)	57 (28.36)	63	69.57
C	OpenStreetMap	41 (87.23)	6 (12.77)	166	19.81
D	Google Maps	22 (64.71)	12 (35.29)	11	66.67
D	OpenStreetMap	21 (80.77)	5 (19.23)	12	63.64

Area A: sparsely populated municipality in the south-west of Bavaria

Area B: area in a medium-sized populated major district town near Munich

Area C: area close to the centre within the densely populated city of Munich

Area D: area with a lower density of amenities within the densely populated city of Munich

[a] The percentage of true positives is the positive predictive value (PPV) [PPV = true positives/(true positives + false positives)]

[b] Sensitivity = true positives/(true positives + false negatives)

Table 4 Results of the field validation without the category 'doctor'

Area	Geocoding service	True positives: N (% positive)[a]	False positives: N (% positive)	False negatives: N	Sensitivity[b]: %
A	Google Maps	18 (62.07)	11 (37.93)	13	58.06
A	OpenStreetMap	15 (88.24)	2 (11.76)	16	48.39
B	Google Maps	29 (90.63)	3 (9.38)	1	96.67
B	OpenStreetMap	10 (100)	0 (0)	20	33.33
C	Google Maps	48 (69.57)	21 (30.43)	30	61.54
C	OpenStreetMap	36 (85.71)	6 (14.29)	42	46.15
D	Google Maps	19 (67.86)	9 (32.14)	6	76.00
D	OpenStreetMap	21 (80.77)	5 (19.23)	4	84.00

Area A: sparsely populated municipality in the south-west of Bavaria

Area B: area in a medium-sized populated major district town near Munich

Area C: area close to the centre within the densely populated city of Munich

Area D: area with a lower density of amenities within the densely populated city of Munich

[a] The percentage of true positives is the positive predictive value (PPV) [PPV = true positives/(true positives + false positives)]

[b] Sensitivity = true positives/(true positives + false negatives)

and 89%. In contrast, sensitivities were higher in Google Maps (between 59 and 98%) than in OSM (between 20 and 64%). False negatives were higher for OSM within three of the four validation areas. An overall comparison between the four areas showed that Area C within the city of Munich had the highest numbers of false negatives for both geocoding services. For OSM, high numbers of false negatives were also discovered for Area B, i.e. for the major district town. Predominantly during the validation within Area C, it became evident that the data quality regarding the variable category 'doctor' had a fundamental influence on the validation results. Therefore, we recalculated Table 3 without the POIs belonging to this category. The results of this recalculation process can be found within Table 4. Having omitted the category 'doctor', sensitivities of OSM improved for Area B and Area C. Within Area D, sensitivities of OSM were higher than sensitivities of Google Maps.

Discussion

The aim of our study was to examine the feasibility of integrating information from online geocoding services for the assessment of environmental obesogenic factors that could potentially be used for diabetes surveillance. First, we identified variables correlated with obesogenic environments from the literature. Subsequently, we tested whether these variables could be reproduced using data from the online geocoding services Google Maps and OpenStreetMap (OSM). The results showed that this was possible given some restrictions, predominantly the diversity of the variable pools of the geocoding services and the complexity of the environmental factor to be projected. Maps created from the obesogenic and from protective data showed the geographical distribution of the environmental factors and were used within subsequent field validations. On the one hand, Google Maps showed greater completeness, i.e. lower proportion of false negatives, regarding POIs subsequently discovered in the field and the additional information assigned to them. Furthermore, the sensitivity of Google Maps was higher than the sensitivity of OSM. On the other hand, a higher PPV was seen for OSM in each of the validation areas.

Recently, the validity of the geocoding service Google Maps was tested using geoprocessing information [18]. Instead of using single geographical data points from the spatial databases of Google Maps and OSM, the authors compared virtual audit via Google Street View. Additionally, local field inspections were performed as the gold standard. It was shown that the validity and reliability of using Google Maps for the assessment of the built environment was high (Kappa of 78% and 80% respectively). Considering the German context, field inspections concerning the obesogenic environment have been performed in the past in order to record POIs [44]. Therefore, it was an important step within our study to inspect the database results of Google Maps and OSM locally.

PPVs of Google Maps and OSM found during our validation process were compared with each other. It became evident that the PPV of OSM was higher than the PPV for Google Maps in each region, because Google Maps showed considerably more false positives. Considering sensitivity, OSM showed lower values than Google Maps. Most influential variables regarding these comparisons were found within the category 'doctor'. The data quality regarding physicians was better for Google Maps compared to OSM. Therefore, within areas with a higher share of doctors (Area B and Area C) the differences in sensitivities between Google Maps and OSM were large. Deleting the category 'doctor' from the analysis thus moderated this difference. False positives of Google Maps within the densely populated Area C were also mainly caused by the category 'doctor'. The same category also contributed to the number of false negatives in OSM within this area and the sensitivity of OSM improved considerably after omitting POIs belonging to this category (see Table 4). To highlight the different influences of certain variable groups, it was an important step in our validation process to look for suitable stratification structures, such as the six categories 'doctor', 'food', 'sport', 'transport', 'education' and 'other'.

In our study, we calculated the sensitivities and PPVs of Google Maps and OSM hits. They can be compared with the PPVs of other POI databases that we found in the existing literature. For example, Clary and colleagues [34] validated a Canadian food outlet database in the field. Comparing their database results with the actual occurrences in the field, the authors found sensitivities between 54.5 and 65.5% as well as PPVs between 64.4 and 77.3%. Within our study, the PPV for OSM was markedly higher (between 81 and 100%), whereas Google Maps had a more similar PPV compared with the Canadian database (between 63 and 89%). Regarding sensitivity, OSM showed lower values within three of the four validation areas (between 20 and 64%), whereas Google Maps sensitivities were at least comparable (between 59 and 98%) with the food outlet database.

To evaluate features of the obesogenic environment, Bethlehem and colleagues [14] performed a virtual audit based on Google Earth (GE) and Google Street View (GSV). They assessed the aspects walking, cycling, public transport, aesthetics, land use mix, grocery stores, food outlets and recreational facilities using observers. Virtual audit was found to be a valid and reliable approach. Within our study, we used Google Maps and OSM APIs for the programmed download of POIs, which does not need individual assessment for data collection.

Within our analyses, it also became evident that new variable entries appear more frequently, but old entries were deleted with time lag within the Google Maps database. The more specific variables in the OSM pool made it possible to identify some POIs that could not be precisely queried by Google. For example, OSM made it possible to extract 'fast food' instead of the broader category 'food'. This feature nevertheless required taking into account all relevant specific factors describing a variable at a higher level in order to exhaust the OSM database completely.

Strengths and limitations

Our study is based on an extensive literature search extracting factors of obesogenic environments. We used freely available data from global geocoding services Google Maps and OSM and applied various methods for downloading and processing geographical data using new query codes in the R programming environment. Finally, we validated our results with in-field inspections. To evaluate both physical activity and food-related environmental factors, composite approaches are required, which have been performed rather infrequently in the past [12]. Within our approach, we combined the food environment and the physical activity environment into a single layer containing POIs of the obesogenic factors and POIs of the protective factors.

Some limitations of our study have to be mentioned. First, it is focused on evidence from the literature based on an energy imbalance model [45]. However, according to the recent literature, other etiological causes for the development of obesity have to be considered to fully understand the underlying mechanisms, e.g. the carbohydrate-insulin model of obesity (beyond 'calories in, calories out') [46] or dietary behaviour ('ultra-processed food vs unprocessed food') [47]. Second, the literature search had a broad scope by updating and complementing a systematic review; however, a large number of the identified studies originated from the US. Structural differences regarding the built environment in US and European cities may influence direct transferability to the European context. For example, cities in the US are much more car dependent than European cities, which results in expected different health effects of environmental factors associated with physical activity [48]. Furthermore, instead of unhealthy corner stores in the US, in European countries, healthy stores selling fresh fruit and vegetables exist more often and are more evenly distributed across the cities [49]. A third drawback regarding our literature search could be publication bias, which would influence the assessment of the overall correlation of an environmental factor [50, 51]. Fourth, a significant proportion of the environmental factors was not correlated with obesity in the same direction across studies. Given this restriction, we have summarized the correlations found in the literature based on expert decision. Fifth, the precision and feasibility of variable extraction fundamentally depend on the variable pool structure of the geocoding service. Differences in the definition of a variable across geocoding services hamper direct comparisons of variables. Within our study, we found that the variable pool of OSM contains many more variables than Google Maps for a large number of environmental factors. Finding broader categories for environmental factors within our analysis made it easier to compare variables across geocoding services. Sixth, our study was limited to a German environment; therefore, generalization of our findings needs further assessment in other countries. Seventh, we have downloaded spatial POIs at a certain point in time; thus, we cannot make inferences on time effects. However, this cross-section offers an important starting point for future analyses. Eighth, each geographical area was validated by a different researcher; therefore, interobserver variability could have appeared during validation. In order to counteract this kind of bias, prior instructions were defined as precisely as possible, and discussions between the observers took place both before and after the validations. Finally, some restrictions regarding access to certain facilities appeared during validations in the field, mostly concerning database hits of the category 'doctor'. Results of the validation process without this category are shown in Table 4. Within the analysis including the category 'doctor', this generated some uncertainties; therefore, we used the best available evidence, i.e. the home pages of these amenities and business directories (yellow pages). However, these uncertainties occurred only in a small number of cases and were discussed in detail during processing of the validation results.

The aim of our study was to examine the feasibility of using data from online geocoding services for diabetes surveillance. We were able to integrate information from these services by downloading, processing and visualizing their data on maps. The reliability of these variables was assessed within field validations and by search engines. Future examinations could test further types of variables from other research areas.

Conclusions

Based on an extensive literature search, environmental factors could be identified that are associated with obesity. These factors could be partly operationalized through the variables and data available from the online geocoding services Google Maps and OSM. Using APIs, spatial data points could be identified and subsequently visualized on maps. Our findings showed that the validity of data from online geocoding services was reasonable. Consequently, environmental obesogenic factors could be described with our methodology and potentially used within diabetes surveillance. Further validation studies are needed to investigate the importance of environmental obesogenic factors.

Additional files

Additional file 1: Table S1. Factors determined by literature search.

Additional file 2: Table S2. Complete list of chosen OSM variables.

Abbreviations

API: application programming interface; BMI: body mass index; GE: Google Earth; GIS: geographic information system; GSV: Google Street View; JSON: JavaScript Object Notation; OSM: OpenStreetMap; POI: point of interest; PPV: positive predictive value; T2DM: type 2 diabetes mellitus; URL: uniform resource locator; VGI: volunteered geographical information.

Acknowledgements

Not applicable.

Authors' contributions

WM initiated the cooperation and supervised the project. MP and JB were responsible for the literature search and extraction of environmental factors. MP programmed the database queries, processed the data and visualized the results on maps. CK provided statistical input. ML and WM provided health scientific input. MP, JB, CK and ML performed field validations. MP wrote the manuscript. All authors read previous versions of the manuscript, commented on and approved the final version of the manuscript for submission. All authors read and approved the final manuscript.

Funding
The study was supported by the Federal Ministry of Health, Germany (FKZ: GE20160324) as part of the German Diabetes Surveillance Project conducted by the Robert Koch Institute, Berlin.

Availability of data and materials
The data generated for this study were downloaded from Google Maps and OSM and the exemplary code is available on github [https://github.com/MAPraeger/GOcode. Accessed 23 April 2019] as described in the methods section.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Institute of Health Economics and Health Care Management, Helmholtz Zentrum München – German Research Center for Environmental Health (GmbH), Ingolstädter Landstraße 1, 85758 Neuherberg, Germany. [2] German Center for Diabetes Research, Neuherberg, Germany.

Received: 22 January 2019 Accepted: 29 May 2019
Published online: 07 June 2019

References
1. Ogden CL, Yanovski SZ, Carroll MD, Flegal KM. The epidemiology of obesity. Gastroenterology. 2007;132(6):2087–102.
2. Apovian CM. Obesity: definition, comorbidities, causes, and burden. Am J Managed Care. 2016;22(7 Suppl):s176–85.
3. Gregg EW, Shaw JE. Global health effects of overweight and obesity. N Engl J Med. 2017;377(1):80–1.
4. Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, Lee A, Marczak L, Mokdad AH, Moradi-Lakeh M, Naghavi M, et al. Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med. 2017;377(1):13–27.
5. Alam S, Lang JJ, Drucker AM, et al. Assessment of the burden of diseases and injuries attributable to risk factors in Canada from 1990 to 2016: an analysis of the Global Burden of Disease Study. CMAJ Open. 2019;7(1):E140–8.
6. Bischoff SC, Boirie Y, Cederholm T, Chourdakis M, Cuerda C, Delzenne NM, Deutz NE, Fouque D, Genton L, Gil C, et al. Towards a multidisciplinary approach to understand and manage obesity and related diseases. Clin Nutr. 2017;36(4):917–38.
7. Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus—present and future perspectives. Nat Rev Endocrinol. 2011;8(4):228–36.
8. Verma S, Hussain ME. Obesity and diabetes: an update. Diabetes Metab Syndr. 2017;11(1):73–9.
9. Bhatnagar A. Environmental determinants of cardiovascular disease. Circ Res. 2017;121(2):162–80.
10. Brisbon N, Plumb J, Brawer R, Paxman D. The asthma and obesity epidemics: the role played by the built environment—a public health perspective. J Allergy Clin Immunol. 2005;115(5):1024–8.
11. Garfinkel-Castro A, Kim K, Hamidi S, Ewing R. Obesity and the built environment at different urban scales: examining the literature. Nutr Rev. 2017;75(suppl 1):51–61.
12. Townshend T, Lake A. Obesogenic environments: current evidence of the built and food environments. Perspect Public Health. 2017;137(1):38–44.
13. Swinburn B, Egger G, Raza F. Dissecting obesogenic environments: the development and application of a framework for identifying and prioritizing environmental interventions for obesity. Prev Med. 1999;29(6 Pt 1):563–70.
14. Bethlehem JR, Mackenbach JD, Ben-Rebah M, Compernolle S, Glonti K, Bardos H, Rutter HR, Charreire H, Oppert JM, Brug J, et al. The SPOTLIGHT virtual audit tool: a valid and reliable tool to assess obesogenic characteristics of the built environment. Int J Health Geogr. 2014;13:52.
15. Lebel A, Daepp MI, Block JP, Walker R, Lalonde B, Kestens Y, Subramanian SV. Quantifying the foodscape: a systematic review and meta-analysis of the validity of commercially available business data. PLoS ONE. 2017;12(3):e0174417.
16. Thornton LE, Pearce JR, Kavanagh AM. Using geographic information systems (GIS) to assess the role of the built environment in influencing obesity: a glossary. Int J Behav Nutr Phys Act. 2011;8:71.
17. Lemke D, Mattauch V, Heidinger O, Hense HW. [Who hits the mark? A comparative study of the free geocoding services of Google and OpenStreetMap]. Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany)). 2015;77(8–9):e160–5.
18. Silva V, Grande AJ, Rech CR, Peccin MS. Geoprocessing via Google Maps for assessing obesogenic built environments related to physical activity and chronic noncommunicable diseases: validity and reliability. J Healthc Eng. 2015;6(1):41–54.
19. Seto E, Hua J, Wu L, Bestick A, Shia V, Eom S, Han J, Wang M, Li Y. The Kunming CalFit study: modeling dietary behavioral patterns using smartphone data. In: Conference proceedings: 2014 annual international conference of the IEEE engineering in medicine and biology society. 2014. p. 6884–7.
20. Feuillet T, Charreire H, Roda C, Ben Rebah M, Mackenbach JD, Compernolle S, Glonti K, Bardos H, Rutter H, De Bourdeaudhuij I, et al. Neighbourhood typology based on virtual audit of environmental obesogenic characteristics. Obes Rev. 2016;17(Suppl 1):19–30.
21. Li B, Wang J, Wu S, Jia Z, Li Y, Wang T, Zhou S. New method for improving spatial allocation accuracy of industrial energy consumption and implications for polycyclic aromatic hydrocarbon emissions in China. Environ Sci Technol. 2019;53(8):4326–34.
22. Schuurman N, Berube M, Crooks VA. Measuring potential spatial access to primary health care physicians using a modified gravity model. Can Geogr. 2010;54(1):29–45.
23. Ali MK, Siegel KR, Laxy M, Gregg EW. Advancing measurement of diabetes at the population level. Curr Diabetes Rep. 2018;18(11):108.
24. Mackenbach JD, Rutter H, Compernolle S, Glonti K, Oppert JM, Charreire H, De Bourdeaudhuij I, Brug J, Nijpels G, Lakerveld J. Obesogenic environments: a systematic review of the association between the physical environment and adult weight status, the SPOTLIGHT project. BMC Public Health. 2014;14:233.
25. Jia P, Cheng X, Xue H, Wang Y. Applications of geographic information systems (GIS) data and methods in obesity-related research. Obes Rev. 2017;18(4):400–11.
26. Google. Legal notices for Google Maps/Google Earth and Google Maps/Google Earth APIs. https://www.google.com/intl/en_ALL/help/legalnotices_maps.html. Accessed 11 Sept 2018.
27. Landesamt für Digitalisierung Breitband und Vermessung. https://www.ldbv.bayern.de/index.html. Accessed 11 Sept 2018.
28. Neis P, Zielstra D. Recent developments and future trends in volunteered geographic information research: the case of OpenStreetMap. Future Internet. 2014;6(1):76–106.
29. Feng J, Glass TA, Curriero FC, Stewart WF, Schwartz BS. The built environment and obesity: a systematic review of the epidemiologic evidence. Health Place. 2010;16(2):175–90.
30. Statistische Ämter des Bundes und der Länder. Gemeinsames Statistikportal. Gemeindeverzeichnis-Online. https://www.statistikportal.de/de/produkte/gemeindeverzeichnis. Accessed 11 Sept 2018.
31. Places API. https://developers.google.com/places/web-service/intro. Accessed 29 Aug 2017.
32. Downloading data. https://wiki.openstreetmap.org/wiki/Downloading_data. Accessed 27 Nov 2018.
33. Planet.osm. https://wiki.openstreetmap.org/wiki/Planet.osm. Accessed 27 Nov 2018.
34. Clary CM, Kestens Y. Field validation of secondary data sources: a novel measure of representativity applied to a Canadian food outlet database. Int J Behav Nutr Phys Act. 2013;10:77.

35. D'Angelo H, Fleischhacker S, Rose SW, Ribisl KM. Field validation of secondary data sources for enumerating retail tobacco outlets in a state without tobacco outlet licensing. Health Place. 2014;28:38–44.
36. R: a language and environment for statistical computing, Vienna, Austria. https://www.R-project.org/.
37. osmar: OpenStreetMap and R. R Journal. 2012. http://osmar.r-forge.r-project.org/RJpreprint.pdf. Accepted for publication on 2012-08-14.
38. Overpass turbo. https://overpass-turbo.eu/. Accessed 17 Jan 2019.
39. Lampros Mouselimis. geojsonR: a GeoJson processing toolkit. R package version 1.0.0. 2017. https://CRAN.R-project.org/package=geojsonR. Accessed 11 Sept 2018.
40. Jeroen Ooms. The jsonlite package: a practical and consistent mapping between JSON data and R objects. 2014. arXiv:1403.2805 [stat.CO]. https://arxiv.org/abs/1403.2805.
41. Bivand R, Keitt T, Rowlingson B. rgdal: bindings for the Geospatial Data Abstraction Library. R package version 1.2-4. 2016.
42. Kahle D, Wickham H. ggmap: spatial visualization with ggplot2. R J. 2013;5(1):144–61.
43. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer; 2009.
44. Schneider S, Gruber J. Neighbourhood deprivation and outlet density for tobacco, alcohol and fast food: first hints of obesogenic and addictive environments in Germany. Public Health Nutr. 2013;16(7):1168–77.
45. Economos CD, Hatfield DP, King AC, Ayala GX, Pentz MA. Food and physical activity environments: an energy balance approach for research and practice. Am J Prev Med. 2015;48(5):620–9.
46. Ludwig DS, Ebbeling CB. The carbohydrate-insulin model of obesity: beyond"Calories In, Calories Out". JAMA Intern Med. 2018;178(8):1098–103.
47. Hall KD, Ayuketah A, Brychta R, et al. Ultra-processed diets cause excess calorie intake and weight gain: an inpatient randomized controlled trial of ad libitum food intake. Cell Metab. 2019. https://doi.org/10.1016/j.cmet.2019.05.008.
48. Congdon P. Variations in obesity rates between US counties: impacts of activity access, food environments, and settlement patterns. Int J Environ Res Public Health. 2017 Sep 7;14(9):1023
49. Diez J, Bilal U, Cebrecos A, Buczynski A, Lawrence RS, Glass T, Escobar F, Gittelsohn J, Franco M. Understanding differences in the local food environment across countries: a case study in Madrid (Spain) and Baltimore (USA). Prev Med. 2016;89:237–44.
50. Roshandel S, Zheng Z, Washington S. Impact of real-time traffic characteristics on freeway crash occurrence: systematic review and meta-analysis. Accid Anal Prev. 2015;79:198–211.
51. Gebel K, Ding D, Foster C, Bauman AE, Sallis JF. Improving current practice in reviews of the built environment and physical activity. Sports Med. 2015;45(3):297–302.

Publisher's Note

Additional file 1

Table S1: Factors determined by literature search

Factor	Google Maps	OpenStreetMap	Country	Negative correlation	Positive correlation	No association
Factors regarding the food environment						
(Healthy) Food outlets, healthy food	food, health	"amenity"="restaurant", "shop"="farm", "shop"="greengrocer", "shop"="supermarket"	USA, UK, Australia	[1-8]	[9]	[10]
Convenience store	convenience_store	"shop"="convenience"	USA, UK, Japan, Canada	[3, 11-14]	[15-24]	[11, 25-27]
Farmer markets	---	"amenity"="marketplace", "shop"="farm"	USA	---	[18]	---
Fast food	food, meal_delivery, meal_takeaway	"amenity"="fast_food"	USA, New Zealand, UK, Japan, Canada, China, Germany	[3, 16, 28-32]	[12, 17, 20, 23, 25, 33-57]	[4, 13, 14, 26, 43, 54, 58-64]
Food advertising	---	---	USA	---	[65]	---
Food basket price / Relative food prices	---	---	USA	[66, 67]	---	---
Food retail	food	"amenity"="fast_food", "amenity"="restaurant", "shop"="convenience", "shop"="farm", "shop"="greengrocer", "shop"="mall",	USA, Canada, Australia, Ghana	[68]	[9, 31, 69-71]	[72]

		"shop"="supermarket"				
Grocery stores	grocery_or_supermarket	"shop"="greengrocer"	USA, UK, China	[14, 24, 30, 73-75]	[16, 25, 28, 35, 71, 72, 75-77]	[29, 77-80]
Healthy and unhealthy food	health	"amenity"="bbq", "amenity"="fast_food", "shop"="convenience", "shop"="greengrocer"	USA	[81]	---	---
Nutrition environments (high/low)	bakery, bar, food	"amenity"="bbq", "amenity"="food_court", "amenity"="ice_cream", "craft"="bakery", "craft"="caterer", "shop"="bakery", "shop"="beverages", "shop"="butcher", "shop"="cheese", "shop"="chocolate", "shop"="coffee", "shop"="confectionery", "shop"="convenience", "shop"="convenience", "shop"="dairy", "shop"="deli", "shop"="farm", "shop"="ice_cream", "shop"="nutrition_supplements", "shop"="pasta", "shop"="pastry", "shop"="seafood", "shop"="spices", "shop"="tea", "shop"="wine", "vending"="bread", "amenity"="fast_food", "shop"="greengrocer"	USA	[82, 83]	---	---
Other food stores	convenience_store, grocery_or_supermarket	"amenity"="fast_food", "amenity"="restaurant", "shop"="convenience", "shop"="farm", "shop"="greengrocer", "shop"="mall", "shop"="supermarket"	UK, France	[24, 84]	[24, 85]	---
Prepared food sites	food, meal_delivery, meal_takeaway	"amenity"="fast_food"	USA	---	[86, 87]	---

Quality and availability of daily shopping	shopping_mall	"shop"="beverages", "shop"="bicycle", "shop"="butcher", "shop"="cheese", "shop"="chocolate", "shop"="coffee", "shop"="confectionery", "shop"="convenience", "shop"="dairy", "shop"="deli", "shop"="farm", "shop"="fishing", "shop"="free_flying", "shop"="garden_centre", "shop"="garden_furniture", "shop"="greengrocer", "shop"="hunting", "shop"="ice_cream", "shop"="medical_supply", "shop"="nutrition_supplements", "shop"="outdoor", "shop"="pasta", "shop"="pastry", "shop"="scuba_diving", "shop"="seafood", "shop"="spices", "shop"="sports", "shop"="supermarket", "shop"="swimming_pool", "shop"="tea", "shop"="wine","shop"="bakery"	Netherlands	[88]	---	---
Restaurants	food, restaurant	"amenity"="biergarten", "amenity"="restaurant"	USA, China, Canada	[11, 14, 20, 28, 29, 39-41, 43, 48, 51, 73, 75, 89]	[42]	[11, 25, 27, 43, 90]
Side walk cafes	bar	"amenity"="bar", "amenity"="cafe", "amenity"="pub"	USA	[89, 91]	---	---
Supermarkets	convenience_store, grocery_or_supermarket	"shop"="mall", "shop"="supermarket"	USA, Japan, UK, Australia, Canada, France, Portugal	[4, 21, 26, 29, 30, 35, 58, 60, 92-98]	[17, 27, 71, 99]	[13, 100-107]
Unhealthy	food	"amenity"="fast_food"	UK, USA	---	[1, 3, 8]	[7]

food outlets						
Factors regarding the physical activity environment						
(Open) tree cover	---	"natural"="tree", "natural"="tree_row", "natural"="wood"	USA	[91, 108, 109]	---	---
Access to quality parks (larger)	park	"boundary"="national_park", "leisure"="park"	USA	[110]	---	---
Built environment pattern	park	"boundary"="national_park", "highway"="cycleway", "leisure"="park","natural"="wood"	France	---	---	[111]
Automobile dependency, commuting time	---	---	USA	---	[112]	---
Bikeability	---	"amenity"="bicycle_rental", "bicycle_road"="yes", "highway"="cycleway"	USA, USA	[113, 114]	---	[115]
Coastline, access to the beach	---	"natural"="coastline"	Australia, New Zealand	[116, 117]	---	---
Commuting time	---	---	USA	---	[112, 118]	---
Fitness facilities, physical activity facilities, sports	gym, spa	"amenity"="dive_centre", "amenity"="dojo", "leisure"="dance", "leisure"="golf_course", "leisure"="ice_rink", "leisure"="pitch", "leisure"="swimming_area", "leisure"="swimming_pool", "leisure"="track", "sport", "leisure"="fitness_centre",	USA, Australia, France, Spain, China, Canada	[29, 58, 64, 84, 86, 115, 116, 119-127]	[128, 129]	[27, 90, 94]

facilities		"leisure"="sports_centre", "route"="fitness_trail", "route"="hiking", "route"="running"				
Forests, access to forests	---	"natural"="wood"	USA	[108, 130, 131]	---	---
General physical activity level increase, no direct environmental factor	---	---	Canada	[132]	---	---
Greenness, green space	amusement_park, park	"boundary"="national_park", "leisure"="garden", "leisure"="nature_reserve", "leisure"="park", "natural"="fell", "natural"="grassland", "natural"="scrub", "natural"="tree", "natural"="tree_row", "natural"="wood"	USA, New Zealand, Europe (France, Germany, Slovakia, Hungary, Portugal, Italy, Switzerland, Lithuania), UK, Denmark, Canada, Australia, Netherlands, Finland, Germany	[88, 92, 96, 133-147]	[86, 141, 148-150]	[94, 133, 134, 143, 148, 149, 151-154]
Longer way (distance) to	school, university	"amenity"="school","building"="school",	Spain, USA	[49, 50, 129]	---	---

school						
NDVI (normalized difference vegetation index), only another (objective) measure of greenness	park	"natural"	USA	[145]	---	---
Negative perceptions of neighbourhood: crime, traffic, cleanness	---	---	Canada, USA,	---	[155, 156]	---
Neighbourhood activity supportiveness (residential density, number of parks, land-use mix, intersection density)	bicycle_store, bowling_alley, gym, park	"building"="…", "boundary"="national_park", "highway"="…", "landuse"=…", "leisure"="park", "store"="…"	USA	[157]	---	---
Open space	natural_feature	"landuse"="recreation_ground"	USA, China	[49, 50, 128]	[150]	---
Outdoor	amusement_park, rv_park, spa,	"boundary"="national_park", "highway"="cycleway", "highway"="footway",	Canada, USA, Sweden,	[27, 89, 115,	---	[132, 160,

recreation	zoo	"highway"="path", "highway"="rest_area", "leisure"="beach_resort", "leisure"="nature_reserve", "leisure"="park", "leisure"="playground", "leisure"="water_park", "natural"="water", "natural"="wood", "route"="fitness_trail", "route"="hiking", "route"="running", "tourism"="aquarium", "tourism"="camp_site", "tourism"="picnic_site", "tourism"="theme_park"	Australia	158, 159]		161]
Park characteristics, park land area	amusement_park, park	"boundary"="national_park", "leisure"="dog_park", "leisure"="park", "natural"="wood"	USA, UK, Chile, Canada, China, Australia	[22, 27, 30, 57, 110, 121, 134, 137, 162-167]	[23, 150]	[23, 152]
Parking quality and availability	parking	"amenity"="bicycle_parking", "amenity"="motorcycle_parking", "amenity"="parking"	Netherlands	---	[88]	---
Physical activity environment (high / low)	gym	"boundary"="national_park", "highway"="cycleway", "highway"="footway", "highway"="path", "leisure"="park", "natural"="wood", "route"="fitness_trail", "route"="hiking", "route"="running"	USA	[83, 168]	---	---
Playground	---	"leisure"="playground"	USA	---	---	[59]
Recreation centres, recreation facilities	beauty_salon, gym, spa	"amenity"="kneipp_water_cure", "leisure"="fitness_centre", "leisure"="sports_centre"	USA, UK, Australia	[6, 12, 169-171]	---	---

River	---	"amenity"="boat_sharing", "waterway"	China	[150]	---	---
Walkability	---	"boundary"="national_park", "highway"="cycleway", "highway"="footway", "highway"="living_street", "highway"="path", "highway"="pedestrian", "leisure"="park", "natural"="wood"	USA, Canada, Australia, Belgium, Germany, France	[6, 37, 68, 113, 114, 132, 161, 172-194]	[37, 160, 195]	[57, 190, 193, 196-199]
Well-connected landscape	intersection	---	USA	[108]	---	---
Factors regarding the urban form						
(More) rural areas	---	"boundaries", "place"	USA, Finland, Canada, Australia, Italy	[200]	[72, 118, 124, 199, 201-204]	[64, 204-206]
Aesthetics	---	---	USA, Netherlands	[88, 91, 120, 144, 147, 207-209]	[91]	[89]
Bus stop density	bus_station	"amenity"="bus_station", "highway"="bus_stop", "public_transport"="station"	USA, UK	[210]	[211]	[171]
County sprawl	---	"building"="commercial", "building"="residential", "landuse"="commercial", "landuse"="residential"	USA	---	[39, 212-217]	[218]
Destination intensity, destination accessibility	---	---	Australia, New Zealand	[219, 220]	---	---
Graffiti	---	---	Europe (France, Germany, Slovakia,	---	[135]	---

			Hungary, Portugal, Italy, Switzerland, Lithuania)			
Housing density, dwelling density	---	"boundaries", "building", "place"	USA, Australia, New Zealand, UK	[220]	[221]	[68, 171, 222]
Incivilities (breakdown of social order)	---	---	USA	---	[135, 208]	---
Infrastructure	bus_station, car_rental, taxi_stand, train_station	"amenity"="bus_station", "amenity"="car_rental", "amenity"="car_sharing", "amenity"="taxi", "building"="train_station", "highway"="bus_stop", "public_transport"="station"	USA	---	[223]	[120]
Intensity of development	---	---	USA	[179]	---	---
Intersection density	intersection	---	USA, France	[30, 94, 115, 124, 161, 166, 224, 225]	[14, 72, 226, 227]	[226]
Land use (mix)	---	"building"="commercial", "building"="residential", "landuse"="allotments", "landuse"="commercial", "landuse"="farmland", "landuse"="farmyard", "landuse"="forest", "landuse"="grass", "landuse"="greenfield", "landuse"="greenhouse_horticulture",	USA, Australia, UK, Canada, China, International (Australia, Belgium, Brazil,	[7, 47, 68, 125, 150, 161, 210, 226, 228-231]	[232]	[220, 226]

		"landuse"="meadow", "landuse"="orchard", "landuse"="plant_nursery", "landuse"="recreation_ground", "landuse"="residential", "landuse"="village_green", "landuse"="vineyard"	China, Colombia, Czech Republic, Denmark, Mexico, New Zealand, Spain, the UK and USA)			
Landmark buildings	church, museum	"historic"	USA	[91]	---	---
Natural amenities (access to open water, varied topography and mild climate)	natural_feature	"natural"	USA	[233]	---	---
Residential density, population density	---	"building"="commercial", "building"="residential", "landuse"="commercial", "landuse"="residential"	USA, UK, Canada, New Zealand, France, Portugal	[7, 14, 30, 64, 94, 97, 161, 166, 168, 178, 189, 201, 210, 225, 230, 234-237]	[166]	[72, 90, 106, 226]
Route exposure characteristics	route	"route"	USA	---	---	[148]

Safety, trust, no crime	fire_station, police	"amenity"="fire_station", "amenity"="police", "amenity"="prison"	USA, International (Australia, Belgium, Brazil, China, Colombia, Czech Republic, Denmark, Mexico, New Zealand, Spain, the UK and USA), Spain, Portugal, Australia, France	[6, 23, 27, 89, 120, 125, 137, 151, 204, 228, 238-240]	[91, 94, 175]	[29, 37, 59, 90, 164, 223, 240, 241]
Side walk completeness	---	"sidewalk"="both / left / right / no"	USA, UK, Australia	[120, 161, 169, 207, 209]	[211]	---
Street connectivity, road density	---	"highway"="motorway", "highway"="primary", "highway"="residential", "highway"="secondary", "highway"="tertiary", "highway"="trunk", "highway:attribute = noexit = yes"	USA, UK, Canada, Australia, New Zealand, France	[56, 134, 220, 225, 230, 234, 242]	---	[57, 94, 164, 178, 201, 230, 243, 244]
Streetscape	---	---	New Zealand	[220]	---	---
Subway station density, railway	subway_station, train_station	"public_transport"	USA	[50, 226]	---	[226]
Traffic	airport, bus_station,	---	Canada, USA,	[88, 161]	[143, 156, 209, 228,	[30, 143]

	car_rental, taxi_stand, train_station		Germany		245, 246]	
Transport	airport, bus_station, car_rental, taxi_stand, train_station	"aeroway"="aerodrome", "aeroway"="heliport", "amenity"="bus_station", "amenity"="car_rental", "amenity"="car_sharing", "amenity"="taxi", "building"="train_station", "highway"="bus_stop", "public_transport"="station", "railway"="station"	USA	[72, 225]	[176, 225, 229, 236]	---
Urban sprawl	---	"building"="commercial", "building"="residential", "landuse"="commercial", "landuse"="residential"	USA, Canada, Australia, International (Australia, Belgium, Brazil, China, Colombia, Czech Republic, Denmark, Mexico, New Zealand, Spain, the UK and USA)	---	[222, 228, 247-249]	[214, 218, 222]
Other factors						
Education	school, university	"amenity"="college", "amenity"="school"	USA, Canada, Finland, Portugal	[20, 30, 33, 37, 56, 64, 102, 106, 170, 202, 222, 248, 250-253]	---	[72, 216]
Immigration	---	---	Canada	[222]	---	[222]
Poverty	---	---	USA	---	[14, 37, 39,	[37, 254,

					56, 57, 183, 187, 248, 253]	255]
(Primary care) physician supply	doctor	"amenity" = "doctors"	USA	[64, 170, 256]	---	---

References

1. Cetateanu A, Jones A: **Understanding the relationship between food environments, deprivation and childhood overweight and obesity: evidence from a cross sectional England-wide study**. *Health & place* 2014, **27**:68-76.
2. Hosler AS: **Retail food availability, obesity, and cigarette smoking in rural communities**. *The Journal of rural health : official journal of the American Rural Health Association and the National Rural Health Care Association* 2009, **25**(2):203-210.
3. Jennings A, Welch A, Jones AP, Harrison F, Bentham G, van Sluijs EM, Griffin SJ, Cassidy A: **Local food outlets, weight status, and dietary intake: associations in children aged 9-10 years**. *American journal of preventive medicine* 2011, **40**(4):405-410.
4. Larsen K, Cook B, Stone MR, Faulkner GEJ: **Food access and children's BMI in Toronto, Ontario: assessing how the food environment relates to overweight and obesity**. *International Journal of Public Health* 2015, **60**(1):69-77.
5. Miller LJ, Joyce S, Carter S, Yun G: **Associations between childhood obesity and the availability of food outlets in the local environment: a retrospective cross-sectional study**. *American journal of health promotion : AJHP* 2014, **28**(6):e137-145.
6. Moore K, Diez Roux AV, Auchincloss A, Evenson KR, Kaufman J, Mujahid M, Williams K: **Home and work neighbourhood environments in relation to body mass index: the Multi-Ethnic Study of Atherosclerosis (MESA)**. *Journal of epidemiology and community health* 2013, **67**(10):846-853.
7. Rundle A, Neckerman KM, Freeman L, Lovasi GS, Purciel M, Quinn J, Richards C, Sircar N, Weiss C: **Neighborhood food environment and walkability predict obesity in New York City**. *Environmental health perspectives* 2009, **117**(3):442-447.
8. Stark JH, Neckerman K, Lovasi GS, Konty K, Quinn J, Arno P, Viola D, Harris TG, Weiss CC, Bader MD *et al*: **Neighbourhood food environments and body mass index among New York City adults**. *Journal of epidemiology and community health* 2013, **67**(9):736-742.
9. Gantner LA, Olson CM, Frongillo EA: **Relationship of Food Availability and Accessibility to Women's Body Weights in Rural Upstate New York**. *Journal of Hunger and Environmental Nutrition* 2013, **8**(4):490-505.
10. Griffiths C, Frearson A, Taylor A, Radley D, Cooke C: **A cross sectional study investigating the association between exposure to food outlets and childhood obesity in Leeds, UK**. *International Journal of Behavioral Nutrition and Physical Activity* 2014, **11**.

11. Fan JX, Hanson HA, Zick CD, Brown BB, Kowaleski-Jones L, Smith KR: **Geographic scale matters in detecting the relationship between neighbourhood food environments and obesity risk: an analysis of driver license records in Salt Lake County, Utah**. *BMJ open* 2014, **4**(8):e005458.
12. Larson NI, Wall MM, Story MT, Neumark-Sztainer DR: **Home/family, peer, school, and neighborhood correlates of obesity in adolescents**. *Obesity (Silver Spring, Md)* 2013, **21**(9):1858-1869.
13. Mejia N, Lightstone AS, Basurto-Davila R, Morales DM, Sturm R: **Neighborhood Food Environment, Diet, and Obesity Among Los Angeles County Adults, 2011**. *Preventing chronic disease* 2015, **12**:E143.
14. Zick CD, Smith KR, Fan JX, Brown BB, Yamada I, Kowaleski-Jones L: **Running to the store? The relationship between neighborhood environments and the risk of obesity**. *Social science & medicine (1982)* 2009, **69**(10):1493-1500.
15. Berge JM, Wall M, Larson N, Forsyth A, Bauer KW, Neumark-Sztainer D: **Youth dietary intake and weight status: healthful neighborhood food environments enhance the protective role of supportive family home environments**. *Health & place* 2014, **26**:69-77.
16. Fraser LK, Edwards KL, Tominitz M, Clarke GP, Hill AJ: **Food outlet availability, deprivation and obesity in a multi-ethnic sample of pregnant women in Bradford, UK**. *Social science & medicine (1982)* 2012, **75**(6):1048-1056.
17. Hanibuchi T, Kondo K, Nakaya T, Nakade M, Ojima T, Hirai H, Kawachi I: **Neighborhood food environment and body mass index among Japanese older adults: results from the Aichi Gerontological Evaluation Study (AGES)**. *International journal of health geographics* 2011, **10**:43.
18. Jilcott SB, Wade S, McGuirt JT, Wu Q, Lazorick S, Moore JB: **The association between the food environment and weight status among eastern North Carolina youth**. *Public health nutrition* 2011, **14**(9):1610-1617.
19. Laska MN, Hearst MO, Forsyth A, Pasch KE, Lytle L: **Neighbourhood food environments: are they associated with adolescent dietary intake, food purchases and weight status?** *Public health nutrition* 2010, **13**(11):1757-1763.
20. Minaker LM, Raine KD, Wild TC, Nykiforuk CIJ, Thompson ME, Frank LD: **Objective Food Environments and Health Outcomes**. *American journal of preventive medicine* 2013, **45**(3):289-296.
21. Morland K, Diez Roux AV, Wing S: **Supermarkets, other food stores, and obesity: the atherosclerosis risk in communities study**. *American journal of preventive medicine* 2006, **30**.
22. Ohri-Vachaspati P, Lloyd K, Delia D, Tulloch D, Yedidia MJ: **A closer examination of the relationship between children's weight status and the food and physical activity environment**. *Preventive medicine* 2013, **57**(3):162-167.
23. Prince SA, Kristjansson EA, Russell K, Billette JM, Sawada MC, Ali A, Tremblay MS, Prud'homme D: **Relationships between neighborhoods, physical activity, and obesity: a multilevel analysis of a large Canadian City**. *Obes (Silver Spring)* 2012, **20**.
24. Yan R, Bastian ND, Griffin PM: **Association of food environment and food retailers with obesity in US adults**. *Health & place* 2015, **33**:19-24.
25. Block JP, Christakis NA, O'Malley AJ, Subramanian SV: **Proximity to food establishments and body mass index in the Framingham Heart Study offspring cohort over 30 years**. *American journal of epidemiology* 2011, **174**(10):1108-1114.
26. Bodor JN, Rice JC, Farley TA, Swalm CM, Rose D: **The association between obesity and urban food environments**. *Journal of urban health : bulletin of the New York Academy of Medicine* 2010, **87**(5):771-781.
27. Wall MM, Larson NI, Forsyth A, Van Riper DC, Graham DJ, Story MT, Neumark-Sztainer D: **Patterns of obesogenic neighborhood features and adolescent weight: a comparison of statistical approaches**. *American journal of preventive medicine* 2012, **42**(5):e65-75.

28. Ahern M, Brown C, Dukas S: **A National Study of the Association Between Food Environments and County-Level Health Outcomes**. *The Journal of Rural Health* 2011, **27**(4):367-379.
29. Black JL, Macinko J, Dixon LB, Fryer GE, Jr.: **Neighborhoods and obesity in New York City**. *Health & place* 2010, **16**(3):489-499.
30. Huang R, Moudon AV, Cook AJ, Drewnowski A: **The spatial clustering of obesity: does the built environment matter?** *Journal of human nutrition and dietetics : the official journal of the British Dietetic Association* 2015, **28**(6):604-612.
31. Jilcott SB, McGuirt JT, Imai S, Evenson KR: **Measuring the retail food environment in rural and urban North Carolina counties**. *Journal of Public Health Management and Practice* 2010, **16**(5):432-440.
32. Pearce J, Hiscock R, Blakely T, Witten K: **A national study of the association between neighbourhood access to fast-food outlets and the diet and weight of local residents**. *Heal Place* 2009, **15**.
33. Burgoine T, Forouhi NG, Griffin SJ, Brage S, Wareham NJ, Monsivais P: **Does neighborhood fast-food outlet exposure amplify inequalities in diet and obesity? A cross-sectional study**. *The American journal of clinical nutrition* 2016, **103**(6):1540-1547.
34. Burgoine T, Forouhi NG, Griffin SJ, Wareham NJ, Monsivais P: **Associations between exposure to takeaway food outlets, takeaway food consumption, and body weight in Cambridgeshire, UK: population based, cross sectional study**. *BMJ (Clinical research ed)* 2014, **348**:g1464.
35. Chen H-J, Wang Y: **Changes in the neighborhood food store environment and children's body mass index at peripuberty in the United States**. *Journal of Adolescent Health* 2016, **58**(1):111-118.
36. Chen SE, Florax RJ, Snyder SD: **Obesity and fast food in urban markets: a new approach using geo-referenced micro data**. *Health economics* 2013, **22**(7):835-856.
37. Christman Z, Pruchno R, Cromley E, Wilson-Genderson M, Mir I: **A Spatial Analysis of Body Mass Index and Neighborhood Factors in Community-Dwelling Older Men and Women**. *International journal of aging & human development* 2016, **83**(1):3-25.
38. Fraser LK, Edwards KL: **The association between the geography of fast food outlets and childhood obesity rates in Leeds, UK**. *Health & place* 2010, **16**(6):1124-1128.
39. Gregson J: **Poverty, sprawl, and restaurant types influence body mass index of residents in California counties**. *Public health reports (Washington, DC : 1974)* 2011, **126 Suppl 1**:141-149.
40. Hollands S, Campbell MK, Gilliland J, Sarma S: **A spatial analysis of the association between restaurant density and body mass index in Canadian adults**. *Preventive medicine* 2013, **57**(4):258-264.
41. Hollands S, Campbell MK, Gilliland J, Sarma S: **Association between neighbourhood fast-food and full-service restaurant density and body mass index: a cross-sectional study of Canadian adults**. *Canadian journal of public health = Revue canadienne de sante publique* 2014, **105**(3):e172-178.
42. Inagami S, Cohen DA, Brown AF, Asch SM: **Body mass index, neighborhood fast food and restaurant concentration, and car ownership**. *Journal of urban health : bulletin of the New York Academy of Medicine* 2009, **86**.
43. Jeffery RW, Baxter J, McGuire M, Linde J: **Are fast food restaurants an environmental risk factor for obesity?** *Int J Behav Nutr Phys Act* 2006, **3**.
44. Kruger DJ, Greenberg E, Murphy JB, DiFazio LA, Youra KR: **Local concentration of fast-food outlets is associated with poor nutrition and obesity**. *American journal of health promotion : AJHP* 2014, **28**(5):340-343.
45. Lakes T, Burkart K: **Childhood overweight in Berlin: intra-urban differences and underlying influencing factors**. *International journal of health geographics* 2016, **15**.

46. Li F, Harmer P, Cardinal BJ, Bosworth M, Johnson-Shelton D: **Obesity and the built environment: does the density of neighborhood fast-food outlets matter?** *American journal of health promotion : AJHP* 2009, **23**(3):203-209.
47. Li F, Harmer PA, Cardinal BJ, Bosworth M, Acock A, Johnson-Shelton D, Moore JM: **Built environment, adiposity, and physical activity in adults aged 50-75**. *American journal of preventive medicine* 2008, **35**(1):38-46.
48. Mehta N, Chang VW: **Weight status and restaurant availability: a multilevel analysis**. *American journal of preventive medicine* 2008, **34**.
49. Oreskovic NM, Kuhlthau KA, Romm D, Perrin JM: **Built environment and weight disparities among children in high- and low-income towns**. *Academic pediatrics* 2009, **9**(5):315-321.
50. Oreskovic NM, Winickoff JP, Kuhlthau KA, Romm D, Perrin JM: **Obesity and the built environment among Massachusetts children**. *Clinical pediatrics* 2009, **48**(9):904-912.
51. Polsky JY, Moineddin R, Dunn JR, Glazier RH, Booth GL: **Absolute and relative densities of fast-food versus other restaurants in relation to weight status: Does restaurant mix matter?** *Preventive medicine* 2016, **82**:28-34.
52. Polsky JY, Moineddin R, Glazier RH, Dunn JR, Booth GL: **Local restaurant environment in relation to obesity among adults in Southern Ontario: Does restaurant mix matter?** *Canadian Journal of Diabetes* 2015, **39**:S25.
53. Pruchno R, Wilson-Genderson M, Gupta AK: **Neighborhood food environment and obesity in community-dwelling older adults: individual and neighborhood effects**. *American journal of public health* 2014, **104**(5):924-929.
54. Reitzel LR, Regan SD, Nguyen N, Cromley EK, Strong LL, Wetter DW, McNeill LH: **Density and proximity of fast food restaurants and body mass index among African Americans**. *American journal of public health* 2014, **104**(1):110-116.
55. Williams J, Scarborough P, Townsend N, Matthews A, Burgoine T, Mumtaz L, Rayner M: **Associations between Food Outlets around Schools and BMI among Primary Students in England: A Cross-Classified Multi-Level Analysis**. *PloS one* 2015, **10**(7):e0132930.
56. Xu Y, Wang F: **Built environment and obesity by urbanicity in the U.S**. *Health & place* 2015, **34**:19-29.
57. Xu Y, Wen M, Wang F: **Multilevel built environment features and individual odds of overweight and obesity in Utah**. *Applied geography (Sevenoaks, England)* 2015, **60**:197-203.
58. Boone-Heinonen J, Diez-Roux AV, Goff DC, Loria CM, Kiefe CI, Popkin BM, Gordon-Larsen P: **The neighborhood energy balance equation: does neighborhood food retail environment + physical activity environment = obesity? The CARDIA study**. *PloS one* 2013, **8**(12):e85141.
59. Burdette HL, Whitaker RC: **Neighborhood playgrounds, fast food restaurants, and crime: relationships to overweight in low-income preschool children**. *Preventive medicine* 2004, **38**(1):57-63.
60. Hattori A, An R, Sturm R: **Neighborhood food outlets, diet, and obesity among California adults, 2007 and 2009**. *Preventing chronic disease* 2013, **10**:E35.
61. Jiao J, Moudon AV, Kim SY, Hurvitz PM, Drewnowski A: **Health Implications of Adults' Eating at and Living near Fast Food or Quick Service Restaurants**. *Nutrition & Diabetes* 2015, **5**.
62. Mackenbach JD, Lakerveld J, Brug J, Nijpels G: **Understanding how access to fast food outlets is associated with obesity. The SPOTLIGHT project**. *Diabetologia* 2016, **59**(1 Supplement 1):S144.
63. Richardson AS, Meyer KA, Howard AG, Boone-Heinonen J, Popkin BM, Evenson KR, Shikany JM, Lewis CE, Gordon-Larsen P: **Multiple pathways from the neighborhood food environment to increased body mass index through dietary behaviors: A structural equation-based analysis in the CARDIA study**. *Health & place* 2015, **36**:74-87.

64. Slack T, Myers CA, Martin CK, Heymsfield SB: **The geographic concentration of US adult obesity prevalence and associated social, economic, and environmental factors**. *Obesity (Silver Spring, Md)* 2014, **22**(3):868-874.
65. Lesser LI, Zimmerman FJ, Cohen DA: **Outdoor advertising, obesity, and soda consumption: a cross-sectional study**. *BMC Public Health* 2013, **13**.
66. Lear SA, Gasevic D, Schuurman N: **Association of supermarket characteristics with the body mass index of their shoppers**. *Nutrition Journal* 2013, **12**.
67. Xu X, Variyam JN, Zhao Z, Chaloupka FJ: **Relative food prices and obesity in US Metropolitan areas: 1976-2001**. *PloS one* 2014, **9**(12):e114707.
68. Gebel K, Bauman AE, Sugiyama T, Owen N: **Mismatch between perceived and objectively assessed neighborhood walkability attributes: prospective relationships with walking and weight gain**. *Heal Place* 2011, **17**.
69. Dake FA, Thompson AL, Ng SW, Agyei-Mensah S, Codjoe SN: **The Local Food Environment and Body Mass Index among the Urban Poor in Accra, Ghana**. *Journal of urban health : bulletin of the New York Academy of Medicine* 2016, **93**(3):438-455.
70. Spence JC, Cutumisu N, Edwards J, Raine KD, Smoyer-Tomic K: **Relation between local food environments and obesity among adults**. *BMC Public Health* 2009, **9**.
71. Wang MC, Kim S, Gonzalez AA, MacLeod KE, Winkleby MA: **Socioeconomic and food-related physical characteristics of the neighbourhood environment are associated with body mass index**. *Journal of epidemiology and community health* 2007, **61**.
72. Yamada I, Brown BB, Smith KR, Zick CD, Kowaleski-Jones L, Fan JX: **Mixed land use and obesity: an empirical comparison of alternative land use measures and geographic scales**. *Prof Geogr* 2011, **64**.
73. Cerin E, Frank LD, Sallis JF, Saelens BE, Conway TL, Chapman JE, Glanz K: **From neighborhood design and food options to residents' weight status**. *Appetite* 2011, **56**(3):693-703.
74. Kapinos KA, Yakusheva O, Eisenberg D: **Obesogenic environmental influences on young adults: evidence from college dormitory assignments**. *Economics and human biology* 2014, **12**:98-109.
75. Zhang J, Wang HJ, Wang ZH, Zhang JG, Du WW, Su C, Jiang HR, Jia XF, Huang FF, Zhang B: **Influence of proximities to food outlets on daily energy intake among Chinese adults, evidence from china health and nutrition survey, 2009-2011**. *Obesity Reviews* 2016, **17**:159.
76. Gibson DM: **The neighborhood food environment and adult weight status: estimates from longitudinal data**. *American journal of public health* 2011, **101**(1):71-78.
77. Zenk SN, Mentz G, Schulz AJ, Johnson-Lawrence V, Gaines CR: **Longitudinal Associations Between Observed and Perceived Neighborhood Food Availability and Body Mass Index in a Multiethnic Urban Sample**. *Health education & behavior : the official publication of the Society for Public Health Education* 2017, **44**(1):41-51.
78. Gase LN, DeFosset AR, Smith LV, Kuo T: **The Association between Self-Reported Grocery Store Access, Fruit and Vegetable Intake, Sugar-Sweetened Beverage Consumption, and Obesity in a Racially Diverse, Low-Income Population**. *Frontiers in public health* 2014, **2**:229.
79. Mendez DD, Gary-Webb TL, Goode R, Zheng Y, Imes CC, Fabio A, Duell J, Burke LE: **Neighborhood factors and six-month weight change among overweight individuals in a weight loss intervention**. *Preventive medicine reports* 2016, **4**:569-573.
80. Tung EL, Peek ME, Makelarski JA, Lindau ST: **The built and social environment: Neighborhood obesity in vulnerable populations**. *Journal of General Internal Medicine* 2015, **30**:S267.
81. Hutchinson PL, Nicholas Bodor J, Swalm CM, Rice JC, Rose D: **Neighbourhood food environments and obesity in southeast Louisiana**. *Heal Place* 2012, **18**.

82. Paquet C, Coffee NT, Haren MT, Howard NJ, Adams RJ, Taylor AW, Daniel M: **Food environment, walkability, and public open spaces are associated with incident development of cardio-metabolic risk factors in a biomedical cohort**. *Health & place* 2014, **28**:173-176.
83. Saelens BE, Sallis JF, Frank LD, Couch SC, Zhou C, Colburn T, Cain KL, Chapman J, Glanz K: **Obesogenic neighborhood environments, child and parent obesity: the Neighborhood Impact on Kids study**. *American journal of preventive medicine* 2012, **42**(5):e57-64.
84. Casey R, Chaix B, Weber C, Schweitzer B, Charreire H, Salze P, Badariotti D, Banos A, Oppert J, Simon C: **Spatial accessibility to physical activity facilities and to food outlets and overweight in French youth**. *International journal of obesity* 2012, **36**(7):914-919.
85. Macdonald L, Ellaway A, Ball K, Macintyre S: **Is proximity to a food retail store associated with diet and BMI in Glasgow, Scotland?** *BMC Public Health* 2011, **11**.
86. Nies MA, Weber KT, Holmes J, Peterson T, Serr K, Arias J, Lim WY, Force R: **Spatial and Census Data to Evaluate Obese Persons and their Environment (SCOPE)**. *American journal of health behavior* 2015, **39**(4):582-588.
87. von Hippel P, Benson R: **Obesity and the natural environment across US counties**. *American journal of public health* 2014, **104**(7):1287-1293.
88. Putrik P, van Amelsvoort L, De Vries NK, Mujakovic S, Kunst AE, van Oers H, Jansen M, Kant I: **Neighborhood Environment is Associated with Overweight and Obesity, Particularly in Older Residents: Results from Cross-Sectional Study in Dutch Municipality**. *Journal of urban health : bulletin of the New York Academy of Medicine* 2015, **92**(6):1038-1051.
89. Drewnowski A, Aggarwal A, Rehm CD, Cohen-Cline H, Hurvitz PM, Moudon AV: **Environments perceived as obesogenic have lower residential property values**. *American journal of preventive medicine* 2014, **47**(3):260-274.
90. Zhao Z, Kaestner R, Xu X: **Spatial mobility and environmental effects on obesity**. *Economics and human biology* 2014, **14**:128-140.
91. Lovasi GS, Bader MD, Quinn J, Neckerman K, Weiss C, Rundle A: **Body mass index, safety hazards, and neighborhood attractiveness**. *American journal of preventive medicine* 2012, **43**(4):378-384.
92. Hsieh S, Klassen AC, Curriero FC, Caulfield LE, Cheskin LJ, Davis JN, Goran MI, Weigensberg MJ, Spruijt-Metz D: **Built environment associations with adiposity parameters among overweight and obese Hispanic youth**. *Preventive medicine reports* 2015, **2**:406-412.
93. Lamb KE, White SR: **Categorisation of built environment characteristics: the trouble with tertiles**. *Int J Behav Nutr Phys Act* 2015, **12**:19.
94. Leal C, Bean K, Thomas F, Chaix B: **Multicollinearity in associations between multiple environmental features and body weight and abdominal fat: using matching techniques to assess whether the associations are separable**. *American journal of epidemiology* 2012, **175**(11):1152-1162.
95. Li Y, Robinson LE, Carter WM, Gupta R: **Childhood obesity and community food environments in Alabama's Black Belt region**. *Child: care, health and development* 2015, **41**(5):668-676.
96. Liu GC, Wilson JS, Qi R, Ying J: **Green neighborhoods, food retail and childhood overweight: differences by population density**. *American journal of health promotion : AJHP* 2007, **21**(4 Suppl):317-325.
97. Lopez RP: **Neighborhood risk factors for obesity**. *Obesity (Silver Spring, Md)* 2007, **15**(8):2111-2119.
98. Viola D, Arno PS, Maroko AR, Schechter CB, Sohler N, Rundle A, Neckerman KM, Maantay J: **Overweight and obesity: can we reconcile evidence about supermarkets and fast food retailers for public health policy?** *Journal of public health policy* 2013, **34**(3):424-438.
99. Fiechtner L, Block J, Duncan DT, Gillman MW, Gortmaker SL, Melly SJ, Rifas-Shiman SL, Taveras EM: **Proximity to supermarkets associated with higher body mass index among overweight and obese preschool-age children**. *Preventive medicine* 2013, **56**(3-4):218-221.
100. Cummins S, Flint E, Matthews SA: **New neighborhood grocery store increased awareness of food access but did not alter dietary habits or obesity**. *Health affairs (Project Hope)* 2014, **33**(2):283-291.

101. Drewnowski A, Aggarwal A, Hurvitz PM, Monsivais P, Moudon AV: **Obesity and supermarket access: proximity or price?** *Am J Public Heal* 2012, **102**.

102. Drewnowski A, Moudon AV, Jiao J, Aggarwal A, Charreire H, Chaix B: **Food environment and socioeconomic status influence obesity rates in Seattle and in Paris**. *International journal of obesity (2005)* 2014, **38**(2):306-314.

103. Dubowitz T, Ghosh-Dastidar M, Cohen DA, Beckman R, Steiner ED, Hunter GP, Florez KR, Huang C, Vaughan CA, Sloan JC *et al*: **Diet And Perceptions Change With Supermarket Introduction In A Food Desert, But Not Because Of Supermarket Use**. *Health Affairs* 2015, **34**(11):1858-1868.

104. Dubowitz T, Zenk SN, Ghosh-Dastidar B, Cohen DA, Beckman R, Hunter G, Steiner ED, Collins RL: **Healthy food access for urban food desert residents: examination of the food environment, food purchasing practices, diet and BMI**. *Public health nutrition* 2015, **18**(12):2220-2230.

105. Hobbs M, Green M, Griffiths C, Jordan H, Saunders J, McKenna J: **How different data sources and definitions of neighbourhood influence the association between food outlet availability and body mass index: a cross-sectional study**. *Perspectives in public health* 2016.

106. Santana P, Santos R, Nogueira H: **The link between local environment and obesity: a multilevel analysis in the Lisbon Metropolitan Area, Portugal**. *Social science & medicine (1982)* 2009, **68**(4):601-609.

107. Zhang YT, Laraia BA, Mujahid MS, Blanchard SD, Warton EM, Moffet HH, Karter AJ: **Is a reduction in distance to nearest supermarket associated with BMI change among type 2 diabetes patients?** *Health & place* 2016, **40**:15-20.

108. Kim JH, Lee C, Olvara NE, Ellis CD: **The role of landscape spatial patterns on obesity in Hispanic children residing in inner-city neighborhoods**. *Journal of physical activity & health* 2014, **11**(8):1449-1457.

109. Ulmer JM, Wolf KL, Backman DR, Tretheway RL, Blain CJ, O'Neil-Dunne JP, Frank LD: **Multiple health benefits of urban tree canopy: The mounting evidence for a green prescription**. *Health & place* 2016, **42**:54-62.

110. Rundle A, Quinn J, Lovasi G, Bader MD, Yousefzadeh P, Weiss C, Neckerman K: **Associations between body mass index and park proximity, size, cleanliness and recreational facilities**. *Am J Heal Promot* 2013, **27**.

111. Charreire H, Weber C, Chaix B, Salze P, Casey R, Banos A, Badariotti D, Kesse-Guyot E, Hercberg S, Simon C: **Identifying built environmental patterns using cluster analysis and GIS: Relationships with walking, cycling and body mass index in French adults**. *International Journal of Behavioral Nutrition and Physical Activity* 2012, **9**(1):59.

112. Zhang XY, Holt JB, Lu H, Onufrak S, Yang JW, French SP, Sui DZ: **Neighborhood commuting environment and obesity in the United States: An urban-rural stratified multilevel analysis**. *Preventive medicine* 2014, **59**:31-36.

113. Brown BB, Smith KR, Hanson H, Fan JX, Kowaleski-Jones L, Zick CD: **Neighborhood design for walking and biking: physical activity and body mass index**. *American journal of preventive medicine* 2013, **44**.

114. Horacek TM, Dede Yildirim E, Kattelmann K, Brown O, Byrd-Bredbenner C, Colby S, Greene G, Hoerr S, Kidd T, Koenings MM *et al*: **Path Analysis of Campus Walkability/Bikeability and College Students' Physical Activity Attitudes, Behaviors, and Body Mass Index**. *American journal of health promotion : AJHP* 2016.

115. Timperio A, Jeffery RW, Crawford D, Roberts R, Giles-Corti B, Ball K: **Neighbourhood physical activity environments and adiposity in children and mothers: a three-year longitudinal study**. *Int J Behav Nutr Phys Act* 2010, **7**:18.

116. Abbott G, Backholer K, Peeters A, Thornton L, Crawford D, Ball K: **Explaining educational disparities in adiposity: the role of neighborhood environments**. *Obesity (Silver Spring, Md)* 2014, **22**(11):2413-2419.

117. Witten K, Hiscock R, Pearce J, Blakely T: **Neighbourhood access to open spaces and the physical activity of residents: a national study**. *Preventive medicine* 2008, **47**(3):299-303.
118. Kowaleski-Jones L, Wen M: **Community and child energy balance: differential associations between neighborhood environment and overweight risk by gender**. *International journal of environmental health research* 2013, **23**(5):434-445.
119. Black JL, Macinko J: **The changing distribution and determinants of obesity in the neighborhoods of New York City, 2003-2007**. *American journal of epidemiology* 2010, **171**(7):765-775.
120. Catlin TK, Simoes EJ, Brownson RC: **Environmental and policy factors associated with overweight among adults in Missouri**. *Am J Heal Promot* 2003, **17**.
121. Colabianchi N, Coulton CJ, Hibbert JD, McClure SM, Ievers-Landis CE, Davis EM: **Adolescent self-defined neighborhoods and activity spaces: spatial overlap and relations to physical activity and obesity**. *Health & place* 2014, **27**:22-29.
122. Ellaway A, Lamb KE, Ferguson NS, Ogilvie D: **Associations between access to recreational physical activity facilities and body mass index in Scottish adults**. *BMC Public Health* 2016, **16**:756.
123. Gordon-Larsen P, Nelson MC, Page P, Popkin BM: **Inequality in the built environment underlies key health disparities in physical activity and obesity**. *Pediatrics* 2006, **117**(2):417-424.
124. Hoehner CM, Allen P, Barlow CE, Marx CM, Brownson RC, Schootman M: **Understanding the independent and joint associations of the home and workplace built environments on cardiorespiratory fitness and body mass index**. *American journal of epidemiology* 2013, **178**(7):1094-1105.
125. Mobley LR, Root ED, Finkelstein EA, Khavjou O, Farris RP, Will JC: **Environment, obesity, and cardiovascular disease risk in low-income women**. *American journal of preventive medicine* 2006, **30**.
126. Troped PJ, Starnes HA, Puett RC, Tamura K, Cromley EK, James P, Ben-Joseph E, Melly SJ, Laden F: **Relationships between the built environment and walking and weight status among older women in three U.S. States**. *Journal of aging and physical activity* 2014, **22**(1):114-125.
127. Wilson DK, Ainsworth BE, Bowles H: **Body mass index and environmental supports for physical activity among active and inactive residents of a U. S. southeastern county**. *Heal Psychol* 2007, **26**.
128. Almeida FA, Wall SS, You W, Harden SM, Hill JL, Krippendorf BE, Estabrooks PA: **The association between worksite physical environment and employee nutrition, and physical activity behavior and weight status**. *Journal of occupational and environmental medicine* 2014, **56**(7):779-784.
129. Gutierrez-Zornoza M, Sanchez-Lopez M, Garcia-Hermoso A, Gonzalez-Garcia A, Chillon P, Martinez-Vizcaino V: **Active commuting to school, weight status, and cardiometabolic risk in children from rural areas: the Cuenca study**. *Health education & behavior : the official publication of the Society for Public Health Education* 2015, **42**(2):231-239.
130. Boncinelli F, Riccioli F, Marone E: **Do forests help to keep my body mass index low?** *Forest Policy and Economics* 2015, **54**:11-17.
131. Tsai WL, Floyd MF, Leung YF, McHale MR, Reich BJ: **Urban Vegetative Cover Fragmentation in the U.S.: Associations With Physical Activity and BMI**. *American journal of preventive medicine* 2016, **50**(4):509-517.
132. Berry TR, Spence JC, Blanchard CM, Cutumisu N, Edwards J, Selfridge G: **A longitudinal and cross-sectional examination of the relationship between reasons for choosing a neighbourhood, physical activity and body mass index**. *Int J Behav Nutr Phys Act* 2010, **7**:57.
133. Astell-Burt T, Feng X, Kolt GS: **Greener neighborhoods, slimmer people? Evidence from 246,920 Australians**. *International journal of obesity (2005)* 2014, **38**(1):156-159.

134. Coombes E, Jones AP, Hillsdon M: **The relationship of physical activity and overweight to objectively measured green space accessibility and use**. *Social science & medicine (1982)* 2010, **70**(6):816-822.
135. Ellaway A, Macintyre S, Bonnefoy X: **Graffiti, greenery, and obesity in adults: secondary analysis of European cross sectional survey**. *BMJ (Clinical research ed)* 2005, **331**.
136. Halonen JI, Kivimaki M, Pentti J, Stenholm S, Kawachi I, Subramanian SV, Vahtera J: **Green and blue areas as predictors of overweight and obesity in an 8-year follow-up study**. *Obesity (Silver Spring, Md)* 2014, **22**(8):1910-1917.
137. Mathis AL, Rooks RN, Tawk RH, Kruger DJ: **Neighborhood Influences and BMI in Urban Older Adults**. *Journal of applied gerontology : the official journal of the Southern Gerontological Society* 2015.
138. Nielsen TS, Hansen KB: **Do green areas affect health? Results from a Danish survey on the use of green areas and health indicators**. *Health & place* 2007, **13**.
139. Pearson AL, Bentham G, Day P, Kingham S: **Associations between neighbourhood environmental characteristics and obesity and related behaviours among adult New Zealanders**. *BMC Public Health* 2014, **14**:553.
140. Pereira G, Christian H, Foster S, Boruff BJ, Bull F, Knuiman M, Giles-Corti B: **The association between neighborhood greenness and weight status: an observational study in Perth Western Australia**. *Environmental health : a global access science source* 2013, **12**:49.
141. Prince SA, Kristjansson EA, Russell K, Billette J, Sawada M, Ali A, Tremblay MS, Prud D: **A multilevel analysis of neighbourhood built and social environments and adult self-reported physical activity and body mass index in Ottawa**. *Canada* 2011, **7600**.
142. Sanders T, Feng X, Fahey PP, Lonsdale C, Astell-Burt T: **Greener neighbourhoods, slimmer children? Evidence from 4423 participants aged 6 to 13 years in the Longitudinal Study of Australian children**. *International Journal of Obesity* 2015, **39**(8):1224-1229.
143. Schule SA, Fromme H, Bolte G: **Built and socioeconomic neighbourhood environments and overweight in preschool aged children. A multilevel study to disentangle individual and contextual relationships**. *Environmental research* 2016, **150**:328-336.
144. Stark JH, Neckerman K, Lovasi GS, Quinn J, Weiss CC, Bader MD, Konty K, Harris TG, Rundle A: **The impact of neighborhood park access and quality on body mass index among adults in New York City**. *Preventive medicine* 2014, **64**:63-68.
145. Tilt JH, Unfried TM, Roca B: **Using objective and subjective measures of neighborhood greenness and accessible destinations for understanding walking trips and BMI in Seattle.Washington**. *Am J Heal Promot* 2007, **21**.
146. Toftager M, Ekholm O, Schipperijn J, Stigsdotter U, Bentsen P, Gronbaek M, Randrup TB, Kamper-Jorgensen F: **Distance to green space and physical activity: a Danish national representative survey**. *J Phys Act Heal* 2011, **8**.
147. Van Der Zwaard BC, Schalkwijk A, Elders PJM, Nijpels G, Platt L: **Is access to the outdoors associated with childhood overweight and obesity?** *Diabetologia* 2015, **58**(1 SUPPL. 1):S94.
148. Burgoine T, Jones AP, Namenek Brouwer RJ, Benjamin Neelon SE: **Associations between BMI and home, school and route environmental exposures estimated using GPS and GIS: do we see evidence of selective daily mobility bias in children?** *International journal of health geographics* 2015, **14**:8.
149. Cummins S, Fagg J: **Does greener mean thinner? Associations between neighbourhood greenspace and weight status among adults in England**. *Int J Obes* 2012, **36**.
150. Ying Z, Ning LD, Xin L: **Relationship between built environment, physical activity, adiposity, and health in adults aged 46–80 in Shanghai, China**. *Journal of physical activity & health* 2015, **12**(4):569-578.

151. Lovasi GS, Schwartz-Soicher O, Quinn JW, Berger DK, Neckerman KM, Jaslow R, Lee KK, Rundle A: **Neighborhood safety and green space as predictors of obesity among preschool children from low-income families in New York City**. *Preventive medicine* 2013, **57**(3):189-193.
152. Potestio ML, Patel AB, Powell CD, McNeil DA, Jacobson RD, McLaren L: **Is there an association between spatial access to parks/green space and childhood overweight/obesity in Calgary, Canada?** *Int J Behav Nutr Phys Act* 2009, **6**:77.
153. Richardson EA, Pearce J, Mitchell R, Kingham S: **Role of physical activity in the relationship between urban green space and health**. *Public Health* 2013, **127**.
154. Ward JS, Duncan JS, Jarden A, Stewart T: **The impact of children's exposure to greenspace on physical activity, cognitive development, emotional wellbeing, and ability to appraise risk**. *Health & place* 2016, **40**:44-50.
155. Aggarwal A, Rehm C, Cohen-Cline H, Hurvitz P, Moudon A, Drewnowski A: **Residential property values: a novel GIS-based metric to characterize obesogenic environments**. *Faseb J* 2014, **28**(1):no pagination.
156. Berry TR, Spence JC, Blanchard C, Cutumisu N, Edwards J, Nykiforuk C: **Changes in BMI over 6 years: the role of demographic and neighborhood characteristics**. *International journal of obesity (2005)* 2010, **34**.
157. Carlson JA, Remigio-Baker RA, Anderson CA, Adams MA, Norman GJ, Kerr J, Criqui MH, Allison M: **Walking mediates associations between neighborhood activity supportiveness and BMI in the Women's Health Initiative San Diego cohort**. *Health & place* 2016, **38**:48-53.
158. Bjork J, Albin M, Grahn P, Jacobsson H, Ardo J, Wadbro J, Ostergren PO: **Recreational values of the natural environment in relation to neighbourhood satisfaction, physical activity, obesity and wellbeing**. *Journal of epidemiology and community health* 2008, **62**(4):e2.
159. Sullivan SM, Brashear MM, Broyles ST, Rung AL: **Neighborhood environments and obesity among Afro-Caribbean, African American, and Non-Hispanic white adults in the United States: results from the National Survey of American Life**. *Preventive medicine* 2014, **61**:1-5.
160. Adams MA, Todd M, Kurka J, Conway TL, Cain KL, Frank LD, Sallis JF: **Patterns of Walkability, Transit, and Recreation Environment for Physical Activity**. *American journal of preventive medicine* 2015, **49**(6):878-887.
161. Duncan DT, Sharifi M, Melly SJ, Marshall R, Sequist TD, Rifas-Shiman SL, Taveras EM: **Characteristics of walkable built environments and BMI z-scores in children: evidence from a large electronic health record database**. *Environmental health perspectives* 2014, **122**(12):1359-1365.
162. Armstrong B, Lim CS, Janicke DM: **Park Density Impacts Weight Change in a Behavioral Intervention for Overweight Rural Youth**. *Behavioral Medicine* 2015, **41**(3):123-130.
163. Mena C, Fuentes E, Ormazabal Y, Palomo-Velez G, Palomo I: **Role of access to parks and markets with anthropometric measurements, biological markers, and a healthy lifestyle**. *International journal of environmental health research* 2015, **25**(4):373-383.
164. Scott MM, Dubowitz T, Cohen DA: **Regional differences in walking frequency and BMI: what role does the built environment play for Blacks and Whites?** *Heal Place* 2009, **15**.
165. Veitch J, Abbott G, Kaczynski AT, Wilhelm Stanis SA, Besenyi GM, Lamb KE: **Park availability and physical activity, TV time, and overweight and obesity among women: Findings from Australia and the United States**. *Health & place* 2016, **38**:96-102.
166. Wen M, Kowaleski-Jones L: **The built environment and risk of obesity in the United States: racial-ethnic disparities**. *Health & place* 2012, **18**(6):1314-1322.
167. West ST, Shores KA, Mudd LM: **Association of Available Parkland, Physical Activity, and Overweight in America’s Largest Cities**. *J Public Heal Manag Pr* 2012, **18**.

168. Compernolle S, Oppert JM, Mackenbach JD, Lakerveld J, Charreire H, Glonti K, Bardos H, Rutter H, De Cocker K, Cardon G *et al*: **Mediating role of energy-balance related behaviors in the association of neighborhood socio-economic status and residential area density with BMI: The SPOTLIGHT study**. *Preventive medicine* 2016, **86**:84-91.
169. Giles-Corti B, Macintyre S, Clarkson JP, Pikora T, Donovan RJ: **Environmental and lifestyle factors associated with overweight and obesity in Perth, Australia**. *American journal of health promotion : AJHP* 2003, **18**(1):93-102.
170. Myers CA, Slack T, Martin CK, Broyles ST, Heymsfield SB: **Change in Obesity Prevalence across the United States Is Influenced by Recreational and Healthcare Contexts, Food Environments, and Hispanic Populations**. *PloS one* 2016, **11**(2):e0148394.
171. Sarkar C, Gallacher J, Webster C: **Built environment configuration and change in body mass index: the Caerphilly Prospective Study (CaPS)**. *Health & place* 2013, **19**:33-44.
172. Casagrande SS, Gittelsohn J, Zonderman AB, Evans MK, Gary-Webb TL: **Association of walkability with obesity in Baltimore City, Maryland**. *Am J Public Heal* 2011, **101**.
173. Chiu M, Shah B, Rezai MR, Austin P, Tu J: **Neighbourhood walkability and risk of obesity**. *Canadian Journal of Diabetes* 2014, **38**:S39.
174. Creatore MI, Glazier RH, Moineddin R, Fazli GS, Johns A, Gozdyra P, Matheson FI, Kaufman-Shriqui V, Rosella LC, Manuel DG *et al*: **Association of Neighborhood Walkability With Change in Overweight, Obesity, and Diabetes**. *Jama* 2016, **315**(20):2211-2220.
175. Doyle S, Kelly-schwartz A, Schlossberg M, Stockard J: **Active Community Environments and Health: The Relationship of Walkable and Safe Communities to Individual Health**. *J Am Plan Assoc* 2006, **72**.
176. Frank LD, Saelens BE, Powell KE, Chapman JE: **Stepping towards causation: do built environments or neighborhood and travel preferences explain physical activity, driving, and obesity?** *Social science & medicine (1982)* 2007, **65**.
177. Frank LD, Sallis JF, Conway TL, Chapman JE, Saelens BE, Bachman W: **Many pathways from land use to health - Associations between neighborhood walkability and active transportation, body mass index, and air quality**. *J Am Plan Assoc* 2007, **72**.
178. Glazier RH, Creatore MI, Weyman JT, Fazli G, Matheson FI, Gozdyra P, Moineddin R, Shriqui VK, Booth GL: **Density, Destinations or Both? A Comparison of Measures of Walkability in Relation to Transportation Behaviors, Obesity and Diabetes in Toronto, Canada**. *PloS one* 2014, **9**(1).
179. Hirsch JA, Diez Roux AV, Moore KA, Evenson KR, Rodriguez DA: **Change in walking and body mass index following residential relocation: the multi-ethnic study of atherosclerosis**. *American journal of public health* 2014, **104**(3):e49-56.
180. Hoehner CM, Handy SL, Yan Y, Blair SN, Berrigan D: **Association between neighborhood walkability, cardiorespiratory fitness and body-mass index**. *Social science & medicine* 2011, **73**(12):1707-1716.
181. Jilcott Pitts SB, McGuirt JT, Carr LJ, Wu Q, Keyserling TC: **Associations between body mass index, shopping behaviors, amenity density, and characteristics of the neighborhood food environment among female adult Supplemental Nutrition Assistance Program (SNAP) participants in eastern North Carolina**. *Ecology of food and nutrition* 2012, **51**(6):526-541.
182. King AC, Sallis JF, Frank LD, Saelens BE, Cain K, Conway TL, Chapman JE, Ahn DK, Kerr J: **Aging in neighborhoods differing in walkability and income: associations with physical activity and obesity in older adults**. *Social science & medicine (1982)* 2011, **73**(10):1525-1533.
183. Lim S, Harris TG: **Neighborhood Contributions to Racial and Ethnic Disparities in Obesity Among New York City Adults**. *American journal of public health* 2015, **105**(1):159-165.
184. Morinis J, Lebovic G, Gozdyra P, Khovratovich M, Carsley S, Creatore M, Moheddin R, Booth G, Maguire J, Parkin P *et al*: **The weight of place-the role of the neighbourhood in childhood obesity**. *Paediatrics and Child Health (Canada)* 2014, **19**(6):e40.

185. Muller-Riemenschneider F, Pereira G, Villanueva K, Christian H, Knuiman M, Giles-Corti B, Bull FC: **Neighborhood walkability and cardiometabolic risk factors in Australian adults: an observational study**. *BMC Public Health* 2013, **13**:755.
186. Oishi S, Saeki M, Axt J: **Are People Living in Walkable Areas Healthier and More Satisfied with Life?** *Applied psychology Health and well-being* 2015, **7**(3):365-386.
187. Sallis JF, Saelens BE, Frank LD, Conway TL, Slymen DJ, Cain KL, Chapman JE, Kerr J: **Neighborhood built environment and income: examining multiple health outcomes**. *Social science & medicine (1982)* 2009, **68**.
188. Smith KR, Brown BB, Yamada I, Kowaleski-Jones L, Zick CD, Fan JX: **Walkability and body mass index density, design, and new diversity measures**. *American journal of preventive medicine* 2008, **35**(3):237-244.
189. Smith KR, Zick CD, Kowaleski-Jones L, Brown BB, Fan JX, Yamada I: **Effects of neighborhood walkability on healthy weight: assessing selection and causal influences**. *Soc Sci Res* 2011, **40**.
190. Sriram U, LaCroix AZ, Barrington WE, Corbie-Smith G, Garcia L, Going SB, LaMonte MJ, Manson JE, Sealy-Jefferson S, Stefanick ML *et al*: **Neighborhood Walkability and Adiposity in the Women's Health Initiative Cohort**. *American journal of preventive medicine* 2016, **51**(5):722-730.
191. Van Cauwenberg J, Van Holle V, De Bourdeaudhuij I, Van Dyck D, Deforche B: **Neighborhood walkability and health outcomes among older adults: The mediating role of physical activity**. *Health & place* 2016, **37**:16-25.
192. Van Dyck D, Cerin E, Cardon G, Deforche B, Sallis JF, Owen N, de Bourdeaudhuij I: **Physical activity as a mediator of the associations between neighborhood walkability and adiposity in Belgian adults**. *Heal Place* 2010, **16**.
193. Wasfi RA, Dasgupta K, Orpana H, Ross NA: **Neighborhood Walkability and Body Mass Index Trajectories: Longitudinal Study of Canadians**. *American journal of public health* 2016, **106**(5):934-940.
194. Zick CD, Hanson H, Fan JX, Smith KR, Kowaleski-Jones L, Brown BB, Yamada I: **Re-visiting the relationship between neighbourhood environment and BMI: an instrumental variables approach to correcting for residential selection bias**. *Int J Behav Nutr Phys Act* 2013, **10**.
195. McDonald KN, Oakes JM, Forsyth A: **Effect of street connectivity and density on adult BMI: results from the Twin Cities Walking Study**. *Journal of epidemiology and community health* 2012, **66**.
196. Duncan GE, Cash SW, Horn EE, Turkheimer E: **Quasi-causal associations of physical activity and neighborhood walkability with body mass index: a twin study**. *Preventive medicine* 2015, **70**:90-95.
197. Gose M, Plachta-Danielzik S, Willie B, Johannsen M, Landsberg B, Muller MJ: **Longitudinal influences of neighbourhood built and social environment on children's weight status**. *International journal of environmental research and public health* 2013, **10**(10):5083-5096.
198. Michael YL, Gold R, Perrin N, Hillier TA: **Built environment and change in body mass index in older women**. *Health & place* 2013, **22**:7-10.
199. Sugiyama T, Niyonsenga T, Howard NJ, Coffee NT, Paquet C, Taylor AW, Daniel M: **Residential proximity to urban centres, local-area walkability and change in waist circumference among Australian adults**. *Preventive medicine* 2016, **93**:39-45.
200. Donatiello E, Dello Russo M, Formisano A, Lauria F, Nappo A, Reineke A, Sparano S, Barba G, Russo P, Siani A: **Physical activity, adiposity and urbanization level in children: Results for the italian cohort of the IDEFICS study**. *Public Health* 2013, **127**(8):761-765.
201. Burgoine T, Alvanides S, Lake AA: **Assessing the obesogenic environment of North East England**. *Heal Place* 2011, **17**.
202. Nayha S, Lankila T, Rautio A, Koiranen M, Tammelin TH, Taanila A, Rusanen J, Laitinen J: **Body mass index and overweight in relation to residence distance and population density: experience from the Northern Finland birth cohort 1966**. *BMC Public Health* 2013, **13**:938.

203. Penney TL, Rainham DG, Dummer TJ, Kirk SF: **A spatial analysis of community level overweight and obesity**. *Journal of human nutrition and dietetics : the official journal of the British Dietetic Association* 2014, **27 Suppl 2**:65-74.
204. Poortinga W: **Perceptions of the environment, physical activity, and obesity**. *Social science & medicine (1982)* 2006, **63**.
205. Lahti-Koski M, Taskinen O, Simila M, Mannisto S, Laatikainen T, Knekt P, Valsta LM: **Mapping geographical variation in obesity in Finland**. *European journal of public health* 2008, **18**(6):637-643.
206. Moore JB, Brinkley J, Crawford TW, Evenson KR, Brownson RC: **Association of the built environment with physical activity and adiposity in rural and urban youth**. *Preventive medicine* 2013, **56**(2):145-148.
207. Boehmer TK, Hoehner CM, Deshpande AD, Brennan Ramirez LK, Brownson RC: **Perceived and observed neighborhood indicators of obesity among urban adults**. *Int J Obes* 2007, **31**.
208. Knapp EA, Nau C, Brandau S, DeWalle J, Hirsch AG, Bailey-Davis L, Schwartz BS, Glass TA: **Community Audit of Social, Civil, and Activity Domains in Diverse Environments (CASCADDE)**. *American journal of preventive medicine* 2017.
209. Powell-Wiley TM, Ayers CR, de Lemos JA, Lakoski SG, Vega GL, Grundy S, Das SR, Banks-Richard K, Albert MA: **Relationship between perceptions about neighborhood environment and prevalent obesity: data from the Dallas Heart Study**. *Obesity (Silver Spring, Md)* 2013, **21**(1):E14-21.
210. Rundle A, Diez Roux AV, Free LM, Miller D, Neckerman KM, Weiss CC: **The urban built environment and obesity in New York City: a multilevel analysis**. *Am J Heal Promot* 2007, **21**.
211. Duncan DT, Castro MC, Gortmaker SL, Aldstadt J, Melly SJ, Bennett GG: **Racial differences in the built environment--body mass index relationship? A geospatial analysis of adolescents in urban neighborhoods**. *International journal of health geographics* 2012, **11**:11.
212. Arcaya M, James P, Rhodes JE, Waters MC, Subramanian SV: **Urban sprawl and body mass index among displaced Hurricane Katrina survivors**. *Preventive medicine* 2014, **65**:40-46.
213. Ewing R, Meakins G, Hamidi S, Nelson AC: **Relationship between urban sprawl and physical activity, obesity, and morbidity - update and refinement**. *Health & place* 2014, **26**:118-126.
214. Ewing R, Schmid T, Killingsworth R, Zlot A, Raudenbush S: **Relationship between urban sprawl and physical activity, obesity, and morbidity**. *Am J Heal Promot* 2003, **18**.
215. James P, Troped PJ, Hart JE, Joshu CE, Colditz GA, Brownson RC, Ewing R, Laden F: **Urban Sprawl, Physical Activity, and Body Mass Index: Nurses’ Health Study and Nurses' Health Study II**. *Am J Public Heal* 2012, **103**.
216. Joshu CE, Boehmer TK, Brownson RC, Ewing R: **Personal, neighbourhood and urban factors associated with obesity in the United States**. *J Epidemiol Community Heal* 2008, **62**.
217. Plantinga AJ, Bernell S: **The association between urban sprawl and obesity: Is it a two-way street?** *J Reg Sci* 2007, **47**.
218. Eid J, Overman HG, Puga D, Turner MA: **Fat city: Questioning the relationship between urban sprawl and obesity**. *J Urban Econ* 2008, **63**.
219. King TL, Bentley RJ, Thornton LE, Kavanagh AM: **Using kernel density estimation to understand the influence of neighbourhood destinations on BMI**. *BMJ open* 2016, **6**(2).
220. Oliver M, Witten K, Blakely T, Parker K, Badland H, Schofield G, Ivory V, Pearce J, Mavoa S, Hinckson E *et al*: **Neighbourhood built environment associations with body size in adults: mediating effects of activity and sedentariness in a cross-sectional study of New Zealand adults**. *BMC Public Health* 2015, **15**:956.

221. Norman GJ, Adams MA, Kerr J, Ryan S, Frank LD, Roesch SC: **A latent profile analysis of neighborhood recreation environments in relation to adolescent physical activity, sedentary time, and obesity**. *Journal of public health management and practice : JPHMP* 2010, **16**(5):411-419.
222. Ross NA, Tremblay S, Khan S, Crouse D, Tremblay M, Berthelot J-M: **Body mass index in urban Canada: neighborhood and metropolitan area effects**. *American journal of public health* 2007, **97**(3):500-508.
223. Cunningham-Myrie CA, Theall KP, Younger NO, Mabile EA, Tulloch-Reid MK, Francis DK, McFarlane SR, Gordon-Strachan GM, Wilks RJ: **Associations between neighborhood effects and physical activity, obesity, and diabetes: The Jamaica Health and Lifestyle Survey 2008**. *Journal of clinical epidemiology* 2015, **68**(9):970-978.
224. Leonardi C, Simonsen NR, Yu Q, Park C, Scribner RA: **Street Connectivity and Obesity Risk: Evidence From Electronic Health Records**. *American journal of preventive medicine* 2017, **52**(1s1):S40-s47.
225. Samimi A, Mohammadian A, Madanizadeh S: **Effects of transportation and built environment on general health and obesity**. *Transp Res Part D-Transport Environ* 2009, **14**.
226. Brown BB, Yamada I, Smith KR, Zick CD, Kowaleski-Jones L, Fan JX: **Mixed land use and walkability: Variations in land use measures and relationships with BMI, overweight, and obesity**. *Health & place* 2009, **15**(4):1130-1141.
227. James P, Berrigan D, Hart JE, Hipp JA, Hoehner CM, Kerr J, Major JM, Oka M, Laden F: **Effects of buffer size and shape on associations between the built environment and energy balance**. *Health & place* 2014, **27**:162-170.
228. De Bourdeaudhuij I, Van Dyck D, Salvo D, Davey R, Reis RS, Schofield G, Sarmiento OL, Mitas J, Christiansen LB, MacFarlane D *et al*: **International study of perceived neighbourhood environmental attributes and Body Mass Index: IPEN Adult study in 12 countries**. *Int J Behav Nutr Phys Act* 2015, **12**:62.
229. Frank LD, Andresen MA, Schmid TL: **Obesity relationships with community design, physical activity, and time spent in cars**. *American journal of preventive medicine* 2004, **27**.
230. Pouliou T, Elliott SJ: **Individual and socio-environmental determinants of overweight and obesity in Urban Canada**. *Health & place* 2010, **16**(2):389-398.
231. Pouliou T, Elliott SJ, Paez A, Newbold KB: **Building obesity in Canada: understanding the individual- and neighbourhood-level determinants using a multi-level approach**. *Geospatial health* 2014, **9**(1):45-55.
232. Rutt CD, Coleman KJ: **Examining the relationships among built environment, physical activity, and body mass index in El Paso, TX**. *Preventive medicine* 2005, **40**(6):831-841.
233. Littenberg B, Bonnell LN, LeBruin AS, Lubetkin DA, Troy AR, Zia A: **The Relationship Between Access to Natural Environmental Amenities and Obesity**. *Cureus* 2015, **7**(11):e377.
234. Bodea TD, Garrow LA, Meyer MD, Ross CL: **Socio-demographic and built environment influences on the odds of being overweight or obese: The Atlanta experience**. *Transp Res Part a-Policy Pract* 2009, **43**.
235. King AC, Salvo D, Banda JA, Ahn DK, Gill TM, Miller M, Newman AB, Fielding RA, Siordia C, Moore S *et al*: **An observational study identifying obese subgroups among older adults at increased risk of mobility disability: do perceptions of the neighborhood environment matter?** *Int J Behav Nutr Phys Act* 2015, **12**:157.
236. Pendola R, Gen S: **BMI, auto use, and the urban environment in San Francisco**. *Health & place* 2007, **13**.
237. Zhao Z, Kaestner R: **Effects of urban sprawl on obesity**. *J Heal Econ* 2010, **29**.

238. Brown BB, Werner CM, Smith KR, Tribby CP, Miller HJ: **Physical activity mediates the relationship between perceived crime safety and obesity**. *Preventive medicine* 2014, **66**:140-144.
239. Christian H, Giles-Corti B, Knuiman M, Timperio A, Foster S: **The influence of the built environment, social environment and health behaviors on body mass index. results from RESIDE**. *Preventive medicine* 2011, **53**(1-2):57-60.
240. Nogueira H, Ferrao M, Gama A, Mourao I, Marques VR, Padez C: **Perceptions of neighborhood environments and childhood obesity: Evidence of harmful gender inequities among Portuguese children**. *Health & place* 2013, **19**:69-73.
241. Sanchez-Cruz JJ, de Ruiter I, Jimenez-Moleon JJ: **Individual, family and environmental factors associated with pediatric excess weight in Spain: a cross-sectional study**. *Bmc Pediatrics* 2014, **14**.
242. Li K, Wen M, Henry KA: **Residential racial composition and black-white obesity risks: differential effects of neighborhood social and built environment**. *International journal of environmental research and public health* 2014, **11**(1):626-642.
243. Ball K, Lamb K, Travaglini N, Ellaway A: **Street connectivity and obesity in Glasgow, Scotland: impact of age, sex and socioeconomic position**. *Health & place* 2012, **18**(6):1307-1313.
244. Wang F, Wen M, Xu Y: **Population-Adjusted Street Connectivity, Urbanicity and Risk of Obesity in the U.S**. *Applied geography (Sevenoaks, England)* 2013, **41**:1-14.
245. Li W, Dorans KS, Wilker EH, Rice MB, Schwartz J, Coull BA, Koutrakis P, Gold DR, Fox CS, Mittleman MA: **Residential proximity to major roadways, fine particulate matter, and adiposity: The framingham heart study**. *Obesity (Silver Spring, Md)* 2016, **24**(12):2593-2599.
246. Van Hulst A, Gauvin L, Kestens Y, Barnett TA: **Neighborhood built and social environment characteristics: a multilevel analysis of associations with obesity among children and their parents**. *International journal of obesity (2005)* 2013, **37**(10):1328-1335.
247. Garden FL, Jalaludin BB: **Impact of urban sprawl on overweight, obesity, and physical activity in Sydney, Australia**. *J Urban Heal* 2008, **86**.
248. Lopez R: **Urban sprawl and risk for being overweight or obese**. *American journal of public health* 2004, **94**(9):1574-1579.
249. O'Dare Wilson K: **Place matters: Mitigating obesity with the person-in-environment approach**. *Social work in health care* 2016, **55**(3):214-230.
250. McTigue KM, Cohen ED, Moore CG, Hipwell AE, Loeber R, Kuller LH: **Urban Neighborhood Features and Longitudinal Weight Development in Girls**. *American journal of preventive medicine* 2015, **49**(6):902-911.
251. Morris DS, Main EC, Harris JK, Moland A, Cude C: **State-Issued Identification Cards Reveal Patterns in Adult Weight Status**. *International journal of environmental research and public health* 2015, **12**(6):6388-6402.
252. Oka M, Link CL, Kawachi I: **Area-based variations in obesity are more than a function of the food and physical activity environment : area-based variations in obesity**. *Journal of urban health : bulletin of the New York Academy of Medicine* 2013, **90**(3):442-463.
253. Rundle A, Field S, Park Y, Freeman L, Weiss CC, Neckerman K: **Personal and neighborhood socioeconomic status and indices of neighborhood walk-ability predict body mass index in New York City**. *Social science & medicine (1982)* 2008, **67**.
254. Alfonzo M, Guo Z, Lin L, Day K: **Walking, obesity and urban design in Chinese neighborhoods**. *Preventive medicine* 2014, **69 Suppl 1**:S79-85.
255. Leonard T, Powell-Wiley TM, Ayers C, Murdoch JC, Yin W, Pruitt SL: **Property Values as a Measure of Neighborhoods: An Application of Hedonic Price Theory**. *Epidemiology (Cambridge, Mass)* 2016, **27**(4):518-524.
256. Gaglioti AH, Petterson S, Bazemore A, Phillips R: **Access to Primary Care in US Counties Is Associated with Lower Obesity Rates**. *Journal of the American Board of Family Medicine : JABFM* 2016, **29**(2):182-190.

Additional file 2: Complete list of chosen OSM variables

Table S2: Complete list of chosen OSM variables (Key = Value)

"amenity"="bar"

"amenity"="bbq"

"amenity"="biergarten"

"amenity"="cafe"

"amenity"="fast_food"

"amenity"="food_court"

"amenity"="ice_cream"

"amenity"="pub"

"amenity"="restaurant"

"amenity"="college"

"amenity"="school"

"amenity"="bicycle_parking"

"amenity"="bicycle_rental"

"amenity"="boat_sharing"

"amenity"="bus_station"

"amenity"="motorcycle_parking"

"amenity"="parking"

"amenity"="taxi"

"amenity"="clinic"

"amenity"="dentist"

"amenity"="doctors"

"amenity"="hospital"

"amenity"="nursing_home"

"amenity"="pharmacy"

"amenity"="dive_centre"

"amenity"="dojo"

"amenity"="hunting_stand"

"amenity"="kneipp_water_cure"

"amenity"="marketplace"

"amenity"="ranger_station"

"amenity"="vending_machine"

"leisure"="beach_resort"

"leisure"="dance"

"leisure"="dog_park"

"leisure"="fishing"

"leisure"="fitness_centre"

"leisure"="garden"

"leisure"="golf_course"

"leisure"="ice_rink"

"leisure"="nature_reserve"

"leisure"="park"

"leisure"="pitch"

"leisure"="playground"

"leisure"="sports_centre"

"leisure"="stadium"

"leisure"="swimming_area"

"leisure"="swimming_pool"

"leisure"="track"

"leisure"="water_park"

"sport"

"landuse"="allotments"

"landuse"="farmland"

"landuse"="farmyard"

"landuse"="forest"

"landuse"="grass"

"landuse"="greenfield"

"landuse"="greenhouse_horticulture"

"landuse"="meadow"

"landuse"="orchard"

"landuse"="plant_nursery"

"landuse"="recreation_ground"

"landuse"="village_green"

"landuse"="vineyard"

"natural"="wood"

"natural"="tree_row"

"natural"="tree"

"natural"="scrub"

"natural"="grassland"

"natural"="fell"

"natural"="water"

"office"="therapist"

"shop"="bakery"

"shop"="beverages"

"shop"="butcher"

"shop"="cheese"

"shop"="chocolate"

"shop"="coffee"

"shop"="confectionery"
"shop"="convenience"
"shop"="deli"
"shop"="dairy"
"shop"="farm"
"shop"="greengrocer"
"shop"="ice_cream"
"shop"="pasta"
"shop"="pastry"
"shop"="seafood"
"shop"="spices"
"shop"="tea"
"shop"="wine"
"shop"="supermarket"
"shop"="medical_supply"
"shop"="nutrition_supplements"
"shop"="garden_centre"
"shop"="garden_furniture"
"shop"="bicycle"
"shop"="fishing"
"shop"="free_flying"
"shop"="hunting"
"shop"="outdoor"
"shop"="scuba_diving"
"shop"="sports"
"shop"="swimming_pool"
"vending"="bicycle_tube"
"vending"="bread"
"vending"="chemist"
"vending"="chewing_gums"
"vending"="coffee"
"vending"="drinks"
"vending"="first_aid"
"vending"="fishing_tackle"
"vending"="food"
"vending"="ice_cream"
"vending"="milk"
"vending"="sweets"
"craft"="bakery"
"craft"="caterer"
"tourism"="aquarium"
"tourism"="camp_site"
"tourism"="picnic_site"
"tourism"="theme_park"
"highway"="bus_stop"
"highway"="rest_area"
"railway"="station"
"aeroway"="heliport"
"aeroway"="aerodrome"

Note: Key mention without corresponding value means that all values listed within this category were chosen

Licensing information of Article 1

The article

Using data from online geocoding services for the assessment of environmental obesogenic factors: a feasibility study
Original authors: Maximilian Präger, Christoph Kurz, Julian Böhm, Michael Laxy and Werner Maier

was published in the scientific journal "International Journal of Health Geographics" under the terms of the following license:

3.2 Article 2:
A spatial obesity risk score for describing the obesogenic environment using kernel density estimation: development and parameter variation

Published in:

BMC Med Res Methodol. 2023 Mar 17;23(1):65. doi: 10.1186/s12874-023-01883-y.

Maximilian Präger, Christoph Kurz, Rolf Holle, Werner Maier, Michael Laxy

https://doi.org/10.1186/s12874-023-01883-y
BMC Medical Research Methodology

RESEARCH **Open Access**

A spatial obesity risk score for describing the obesogenic environment using kernel density estimation: development and parameter variation

Maximilian Präger[1,2*], Christoph Kurz[3], Rolf Holle[2], Werner Maier[2] and Michael Laxy[1,4]

Abstract

Background Overweight and obesity are severe public health problems worldwide. Obesity can lead to chronic diseases such as type 2 diabetes mellitus. Environmental factors may affect lifestyle aspects and are therefore expected to influence people's weight status. To assess environmental risks, several methods have been tested using geographic information systems. Freely available data from online geocoding services such as OpenStreetMap (OSM) can be used to determine the spatial distribution of these obesogenic factors. The aim of our study was to develop and test a spatial obesity risk score (SORS) based on data from OSM and using kernel density estimation (KDE).

Methods Obesity-related factors were downloaded from OSM for two municipalities in Bavaria, Germany. We visualized obesogenic and protective risk factors on maps and tested the spatial heterogeneity via Ripley's K function. Subsequently, we developed the SORS based on positive and negative KDE surfaces. Risk score values were estimated at 50 random spatial data points. We examined the bandwidth, edge correction, weighting, interpolation method, and numbers of grid points. To account for uncertainty, a spatial bootstrap (1000 samples) was integrated, which was used to evaluate the parameter selection via the ANOVA F statistic.

Results We found significantly clustered patterns of the obesogenic and protective environmental factors according to Ripley's K function. Separate density maps enabled ex ante visualization of the positive and negative density layers. Furthermore, visual inspection of the final risk score values made it possible to identify overall high- and low-risk areas within our two study areas. Parameter choice for the bandwidth and the edge correction had the highest impact on the SORS results.

Discussion The SORS made it possible to visualize risk patterns across our study areas. Our score and parameter testing approach has been proven to be geographically scalable and can be applied to other geographic areas and in other contexts. Parameter choice played a major role in the score results and therefore needs careful consideration in future applications.

Keywords Risk score, Kernel density estimation, Obesity, Spatial, OpenStreetMap

*Correspondence:
Maximilian Präger
maximilian.praeger@tum.de
Full list of author information is available at the end of the article

Background

Overweight and obesity are severe problems worldwide, causing a number of diseases such as type 2 diabetes, and thus reducing expected life years and quality of life [1, 2]. In Germany, for example, the prevalence of being overweight or obese among adults was 54.0% according to the GEDA (GEDA, German Health Update) study from the Robert Koch Institute (a national public health institute in Germany) in 2014/2015, with men being affected more often than women [3]. Other personal aspects affecting obesity besides gender were low education and higher age according to Mader et al. [4]. In addition, several German cohort studies have shown that the average weight in middle-aged populations increased slightly during recent years [5].

Obesity has become a major public health concern, and recent studies describe regional heterogeneity [6, 7]. In obesity-related research, the term "obesogenic environment" describes environmental influences such as green space or fast food restaurants on the development of obesity [8, 9], which has been investigated intensively in the past [10]. Several approaches have been developed in order to analyze the effect of the environment on the risk of obesity. Examples include obesogeneity assessment via questionnaires [11] and via data visualization tools for obesity policy [12].

Some geographic modeling approaches were used to characterize the accumulation of environmental factors. Common techniques include kernel density estimation (KDE), a density method that allows for the estimation of a continuous risk surface [13, 14], as well as hot spot mapping [15] and further geographic information system (GIS) methods [16]. These methods can be used to develop obesity risk scores [17].

Online geocoding services offer low-cost geographic data for researchers that can be downloaded and used for spatial statistical analyses. Their validity has been investigated in the past with reasonable results regarding completeness of environmental factors and positional accuracy of their coordinates [18, 19]. Therefore, they offer a rich database on which geographic tools can be built. However, geocoding services such as Google Maps offer data only in a limited way. In contrast, geodata from OpenStreetMap (OSM) contain geographic information provided by volunteers and thus are less restricted [20]. In a recent study, we performed an extensive literature search to identify obesity-related environmental factors [18]. Furthermore, we operationalized and downloaded corresponding points of interest (POIs).

The aim of our study extends this approach by developing and testing the spatial obesity risk score (SORS) based on data from OSM. The SORS calculates the obesity risk for each geographic point in a given region based on the local density of positive and/or negative obesity-related environmental factors. In our study, we developed a methodological framework for risk score estimation using KDE and tested the influence of five KDE parameters on the SORS values: (1) bandwidth, (2) edge correction via the size of the download area, (3) number of grid points, (4) risk interpolation method, and (5) weighting scheme of the environmental factors.

Methods

Overview of the study approach

We describe the general strategy of obesity risk score estimation, which consists of several steps, by applying it to two regions in Bavaria, Germany. First, we chose our study area and downloaded POIs related to obesogenic environmental factors (cf. [18]). Second, the data were processed to adjust for some imprecisions and minimize redundancy, e.g., we adapted the coordinate units to represent the same length. Third, we analyzed the spatial heterogeneity of the study area and inspected the study area visually to get first insights regarding the distribution of POIs. Fourth, we presented the basic risk score estimation approach, as well as a deterministic sensitivity analysis, in which a spatial POI resampling approach was integrated. This approach made it finally possible in a last step to evaluate the deterministic parameter selection via the ANOVA F statistic. An overview of the risk score estimation process is shown in Fig. 1. Further details regarding each step are given below.

Area selection and data download

We based our analysis on the overall obesogenic and protective environmental factors identified in our previous study [18]. The list of chosen variables for the score was derived via an extensive literature review, i.e., previous work by Mackenbach et al. [10], Jia et al. [21], and enriched with our own further searches.

We chose two areas in the south of Bavaria, Germany, to illustrate our approach and develop the SORS. Our aim was to base the geographical analysis on different levels of urbanity. The first area was the city of Augsburg with about 300,000 inhabitants covering an area of 147 km^2. A more rural region, the town of Meitingen with 11,000 inhabitants and a size of 30 km^2 lying 20 km north of Augsburg, was chosen as the second area. Information on these regions was available from the German Federal Statistical Office [22]. We selected the region around the city of Augsburg, as it is well known within obesity- and diabetes-related research [23].

Spatial POIs related to the selected variables, such as fast food restaurants and parks (compare Additional file 1), were downloaded from OSM using the online data filtering tool Overpass turbo [24]. Data regarding the area

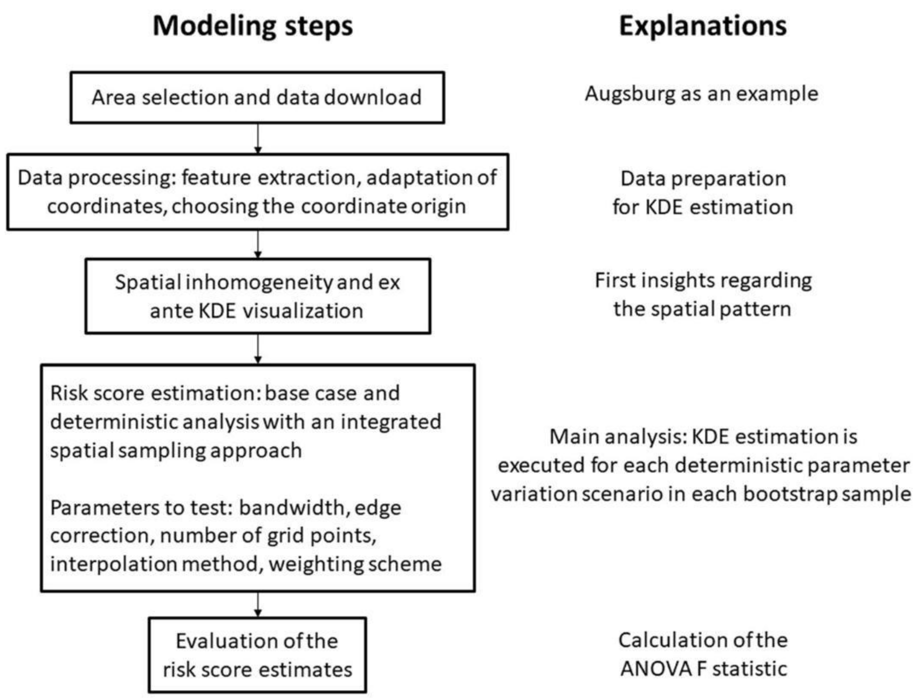

Fig. 1 Modelling process for obesity risk score estimation. ANOVA = analysis of variances, KDE = kernel density estimation

borders of Augsburg and Meitingen were downloaded in shape file format from a geographic online portal provided by the Bavarian government [25]. We directly downloaded maps intended for graphical visualization of results from the OSM web page [26]. Additional file 2 contains further information regarding data download.

Data processing

The downloaded GeoJSON files contained information on each POI regarding name, type of obesity-related environmental factor, spatial coordinates in latitude and longitude format, and other characteristics such as opening hours and street addresses. The relevant information, i.e., name, coordinates, and type of environmental factor, was extracted for each spatial POI and processed into lists and data frames. Table A1 of Additional file 1 gives an overview of the downloaded and processed variables from OSM.

As a further pre-processing step, we introduced a synthetic origin to the south-west of both study areas and adapted the length of a longitudinal coordinate unit to the length of a latitudinal coordinate unit. The schematic structure after the introduction of the origin and the coordinate adaptation is shown in Table 1. The single

Table 1 Schematic example of six processed POIs

Categorization			Coordinates[a]	
Type of environmental factor	Category of the POI[b]	Subcategory of the POI[c]	Longitude	Latitude
Obesogenic	Unhealthy_food	Pastry	0.2595	0.3189
Obesogenic	Unhealthy_food	Pastry	0.2488	0.3161
Obesogenic	Unhealthy_food	Sweets	0.2303	0.2657
Protective	Physical_activity	Canoe	0.2789	0.2966
Protective	Physical_activity	Climbing	0.2404	0.2867
Protective	Physical_activity	Climbing	0.3009	0.2084

POI Point of interest

[a] Coordinates after equidistant transformation and relative to the synthetic origin

[b] Categories were derived from the literature [10, 18]

[c] The subcategories were derived from OpenStreetMap map features [27]

bus stops were reduced from POIs to centroids of dense regions of bus stops. Further details regarding these additional pre-processing steps can be found within Additional file 2.

Spatial inhomogeneity and ex ante KDE visualization

Ripley's K function was calculated for the unweighted obesogenic and protective POIs separately to describe the spatial inhomogeneity of the two study areas. The K function is a second-order moment function that is based on the variance of the radial interpoint distance r around each POI [28]. It compares the cumulative number of actual POIs with the number of expected POIs under random distribution assumption [29]. This random comparison process is realized via the Poisson process, which has a K value of $r^2\pi$ [30]. K functions lying above the K values of a random process therefore represent clustered patterns, whereas smaller values indicate regular processes [31]. We examined the spatial inhomogeneity to investigate ex ante whether clusters are expected to occur in our subsequent POI analysis. The isotropic edge correction, which is a method based on weighting of the POIs according to the probability of their next neighbors being within the study area, was applied to the K function [32]. In order to test whether the K functions of the POIs were significantly different from the K function of a random point pattern, we estimated bootstrap confidence bands around the K functions of the POIs based on the method of Loh with 1000 simulations [33].

In order to visualize the spatial distribution of the POIs, we created KDEs separately for positive and negative spatial data points [34]. These density layers were superimposed and shown together on a single map. Within this process, we defined certain parameter choices as the base case, which were then changed as part of the sensitivity analysis in a later step.

Risk score estimation

We estimated the SORS based on the integration of obesogenic and protective kernel densities into a combined score. The aim of KDE is to provide a smooth and continuous estimation of the accumulation of spatial data points based on a sliding window technique. The geographic plane is represented by two dimensions and the estimated densities account for a third dimension, which thus leads to a three-dimensional mountainous structure, a so-called "risk surface". To visualize this structure on a map, the level of the density coordinate can be plotted by contour lines or via coloring [35]. An overview of this method is provided by Hastie et al., for example, as well as by King et al. [36, 37].

To estimate risk score values, several steps have been performed. First, the positive spatial data points were included into a single positive spatial data layer. Analogously, the negative spatial data points were integrated into a data layer. Second, an observation window together with a grid of suitable size was set up and laid on the respective study area. For each of the following calculations, the same grid was used. Third, KDEs were performed to generate a risk surface based only on positive environmental factors and a second risk surface based only on negative environmental factors. Following this process, a density value based on positive factors and a value based on negative factors were generated at each grid point. Fourth, these positive and negative estimates were set against each other by taking the difference, which results in the final score values at the grid points [34]. Finally, to determine the risk value at the exact desired spatial location, interpolation methods were applied.

Deterministic analysis

The procedure described above implies several parameter choices within its different steps. We tested alternative values for five of these parameters: bandwidth, edge correction, number of grid points, interpolation method, and an alternative weighting scheme (see overview in Table 2). For each given parameter variation, the remaining KDE parameters were set to their base case values

Table 2 Overview on the sensitivity parameters

Parameter	Base case	Deterministic sensitivity analysis
1) Bandwidth	Method of Terrell [38]	1/3, 2/3, 4/3, 5/3 of the base case bandwidth
2) Size of download area (edge correction)	1.4 * side length of the exact box	x * side length of the exact box, with x ϵ {1.8, 2.2, 2.6, 3.0}
3) Number of grid points	35 × 35 grid points	70 × 70 grid points, 105 × 105 grid points
4) Interpolation method	Automatic interpolation with the R function"interp.surface"	Inverse distance weighting density of the nearest grid cell ordinary Kriging
5) Weighting scheme of the environmental factors	Equal weighting with unity	Double weighting of supermarkets and gyms

(see also Table 2). Further explanations regarding the parameters are provided below.

Bandwidth selection

The bandwidth is an important parameter in KDE that determines the degree of smoothing. An increased bandwidth results in a higher smoothing level. For the base case, the oversmoothing bandwidth proposed by Terrell et al. [38] was chosen, which can be described as the maximum smoothing degree that can be suitably applied to a given data set. Within the deterministic scenarios, alternative higher and lower values in steps of 1/3 of the base case bandwidth were tested. We used a pooled bandwidth as described by Davies and colleagues [39].

Size of the download area (edge correction)

Restricting the observational window for KDE to the area boundaries would lead to an underestimation of the true densities at the borders. These effects are especially high if the observation window has a complex structure. To correct for edge effects, we defined a rectangular observation window around the area borders, which simultaneously served as the POI download area and as the KDE area. As a first step, a window was determined from the maximum latitude, maximum longitude, minimum latitude, and minimum longitude of the city boundaries. This minimum bounding rectangle around the study area will be called the "exact box". For the base case, each side length of the rectangle was increased by 40%, and the resulting rectangle was held in a concentric orientation to the smaller rectangle from the step before. The whole observation window was used for the creation of the risk surface. However, risk score values were only evaluated at locations that lay within the boundaries of the respective study area. To determine the effects of the window size, we gradually increased the download area in steps of 40% of the exact box.

Number of grid points

A further central parameter of KDE is the number of grid points. These points were distributed equally on the estimation rectangle. Therefore, KDEs were generated for grid points lying both inside and outside the study area, and both types of grid points were used to estimate the risk score inside the borders of the study areas. A higher number of grid points means that the amount of interpolation is reduced, as more spatial points exist at which an exact estimation is known. To preserve the location of and distance between the inner grid points within the edge correction scenarios, we increased the number of grid points in 40% steps according to the increase in the side length of the download area. Setting the number of grid points to 25×25 for the exact box, this led to a base case grid of 35×35. For each of the following 40% steps of edge correction, again 10 additional grid points were added to each grid point dimension. Within the remaining grid point sensitivity analyses, increased numbers of 70×70 and 105×105 grid points were tested.

Interpolation method

The automatic interpolation function "interp.surface" of the R package "fields" [40] was applied within the base case. As an alternative interpolation approach, we used inverse distance weighting using the four grid points surrounding a targeted evaluation point. Furthermore, extraction of the score value of the grid point with the minimum distance to the targeted evaluation point was implemented as the third interpolation method [41]. As a fourth scenario, we chose ordinary Kriging which has proven its reliability for interpolating surfaces [42].

Weighting scheme of the environmental factors

Several approaches exist for the design of proper weighting schemes. One approach, for example, would be to weight factors according to the strength of evidence for a positive or negative correlation. Regarding our analysis, we chose an equal weighting scheme as a base case scenario. To test an alternative weighting scenario, we followed the approach of Jones-Smith et al. [34]. These authors gave a higher weighting to factors that generally reach more people because of their longer opening hours or size. For the deterministic sensitivity analysis, we therefore tested a double weighting of supermarkets and gyms.

Resampling approach to account for uncertainty in the distribution of the POIs

Uncertainty concerning POIs was integrated into density score estimations via a spatial bootstrapping method. We generated 1000 bootstrap samples for the positive and 1000 bootstrap samples for the negative environmental POIs at random with replacement. For each of the samples, the parameter variation described above was executed. This made it possible to integrate probabilistic variation of POIs into SORS value estimation for each deterministic scenario. These uncertainty estimates were used for the calculation of the ANOVA F statistic, which we applied to compare the deterministic sensitivity scenarios for a given parameter. We describe further details regarding the sampling process in Additional file 2.

Evaluation of the risk score estimates

A random sample ($N = 50$) of spatial data points (evaluation points, EPs) was drawn from each of the two areas for which we calculated and compared risk score results across the scenarios defined above. Our aim was to develop a robust and stable score that accounts for uncertainty and exhibits discriminatory power. To describe how much the choice of parameter values influences the discrimination of data points with low and high risk, we performed analysis of variance (ANOVA) between all 50 EPs based on their bootstrap replications. Thus, for each EP, 1000 bootstrap replicates of the SORS constituted an ANOVA group of estimates for a given parameter scenario. We calculated the F statistic in order to determine the degree of separation between the 50 groups of estimates with higher F values indicating higher degrees of separation. Therefore, the highest value of the F statistic within a given parameter variation analysis in this sense indicates the best result. The relative influence of the parameters on model results is estimated based on a normalization of the F statistic values. The algorithm used to implement deterministic and probabilistic analysis is shown in Fig. 2. Finally, we created heat maps based on the risk score estimates of the base case. For this purpose, the KDE values were again transferred back to the original map dimensions that were used in the pre-visualization step. To compare this base case risk map to an alternative visualization using a common geographical methodology, we estimated an inhomogeneous cluster point process with polynomial trend of degree two for positive and negative POIs that is designed to provide similar cluster structures compared to our KDE approach. Subsequently, we derived intensities for risk score plotting in a comparable way as it was done for the kernel density approach, i.e., via setting off the surfaces. We tested several point process types, such as "Thomas" and MatClust", and chose the model with the best fit based on the Akaike Information Criterion [43]. The code for data download, data processing, and analysis of the scenarios defined above can be found in the supplementary information.

Software

The spatial POIs were processed and analyzed using R version 4.0.2 [44]. For graphical visualization, R packages "ggplot2" [45], "graphics" [44], and "fields" [40] were used. We applied the "spatstat" [46] package to estimate

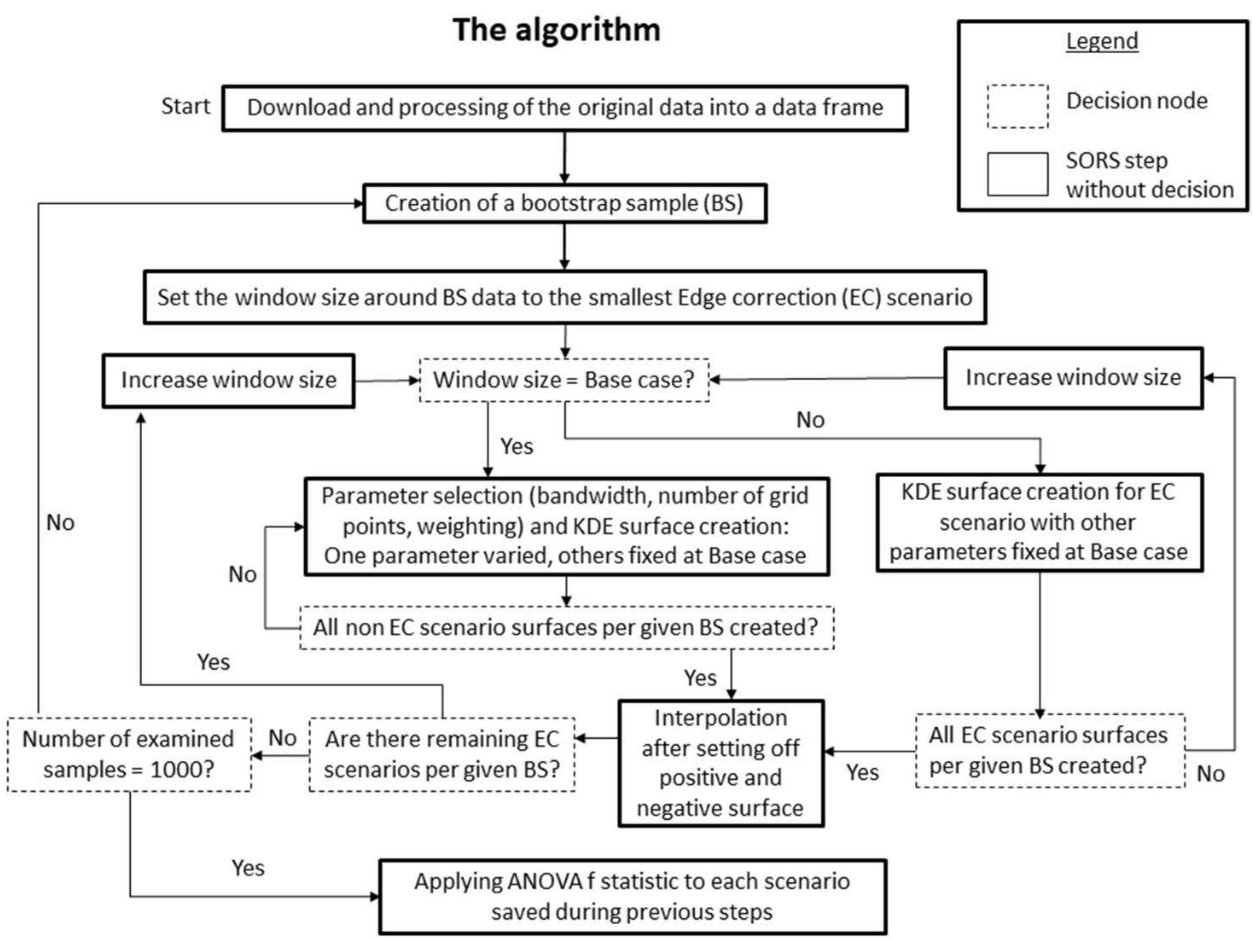

Fig. 2 Algorithm describing the combination of deterministic and probabilistic analysis, BS = bootstrap sample, EC = edge correction

Ripley's K function and the bootstrap confidence bands around it. For KDEs, the packages "MASS" [47] and "sparr" [39] were chosen. Spatial data objects were built and handled via the packages "rgdal" [48], "geojsonR" [49], "prob" [50], and "spatstat" [46]. For interpolation and for the generation of risk score maps, the packages "fields" [40] and "gstat" [51] were used. To plot spatial objects on maps, we used the package "png" [52]. Finally, the DBSCAN algorithm was applied using the package "dbscan" [53].

Results

Spatial inhomogeneity and visual inspection of the study area

The upper part of Fig. 3 shows estimates of Ripley's K function for Augsburg, separately for the obesogenic and the protective POIs. The point pattern for Augsburg was significantly clustered, as the lower confidence bands of the K functions lay above the random Poisson processes for each interpoint distance r, which means that the actual number of POIs within a distance r was greater than the number of expected POIs under random distribution assumptions [31]. This underlined the importance of subsequent KDE analysis, as the spatial pattern was suitable for clustering tasks. Estimates of the obesogenic and the protective K functions for Meitingen also revealed significantly clustered patterns, as the lower confidence bands lay above the random pattern for all or at least for several interpoint distances r, which is shown in the bottom part of Fig. 3.

The separate obesogenic and protective risk surfaces are shown in Fig. 4 for Augsburg and Meitingen. The obesogenic and protective kernel densities for Augsburg accumulated within a region lying to the

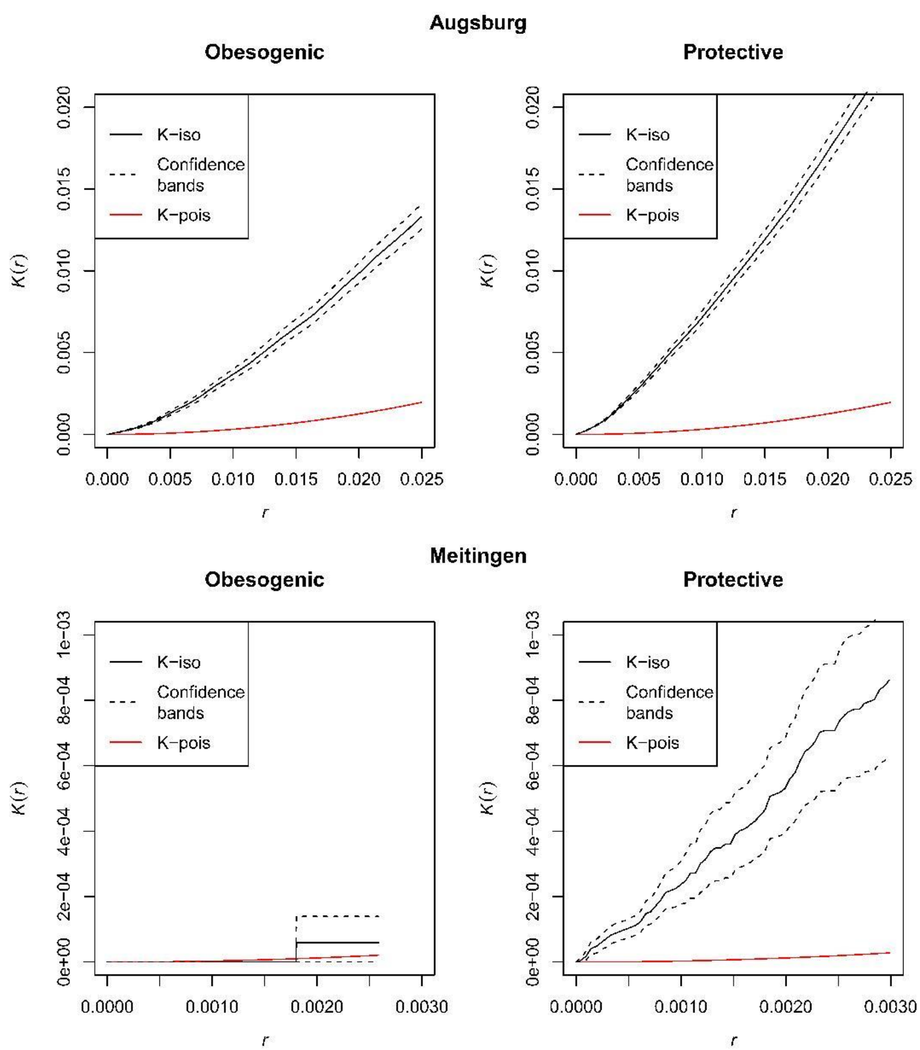

Fig. 3 Spatial inhomogeneity measured via Ripley's K function for Augsburg (top) and Meitingen (bottom), r = interpoint distance, K(r) = Ripley's K function, iso = isotropic edge correction, pois = Poisson point process

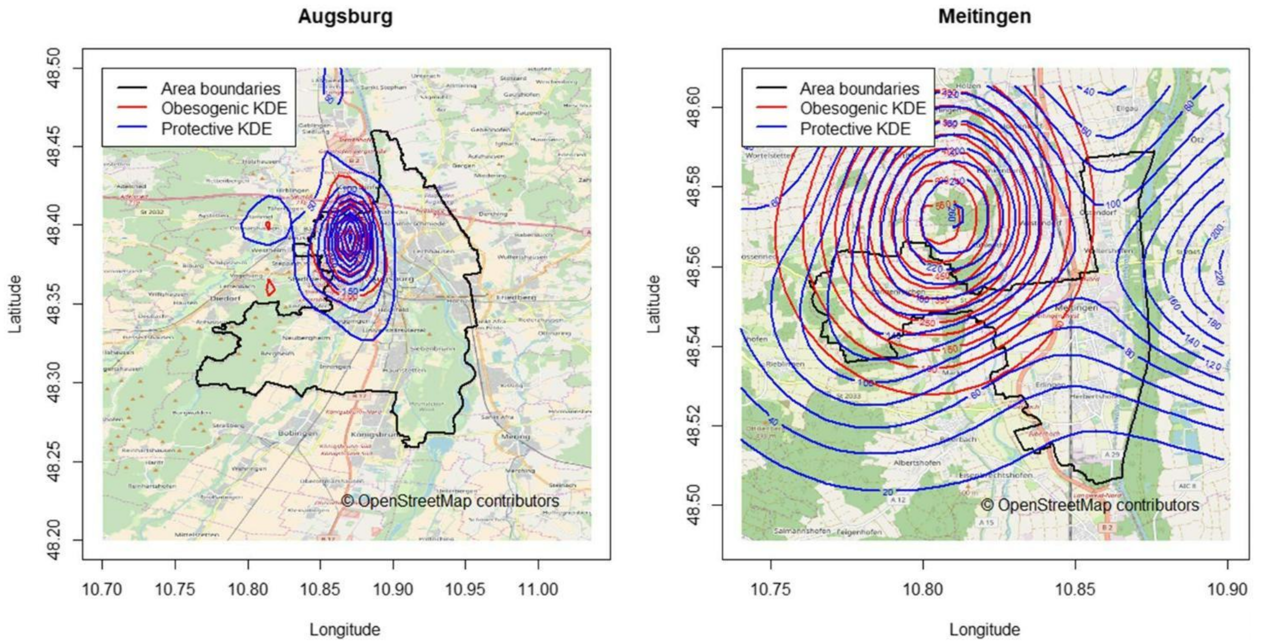

Fig. 4 Contour lines of obesogenic and protective kernel densities in Augsburg (left) and Meitingen (right). Values of KDE are constant on a contour line with increasing values toward the respective KDE center

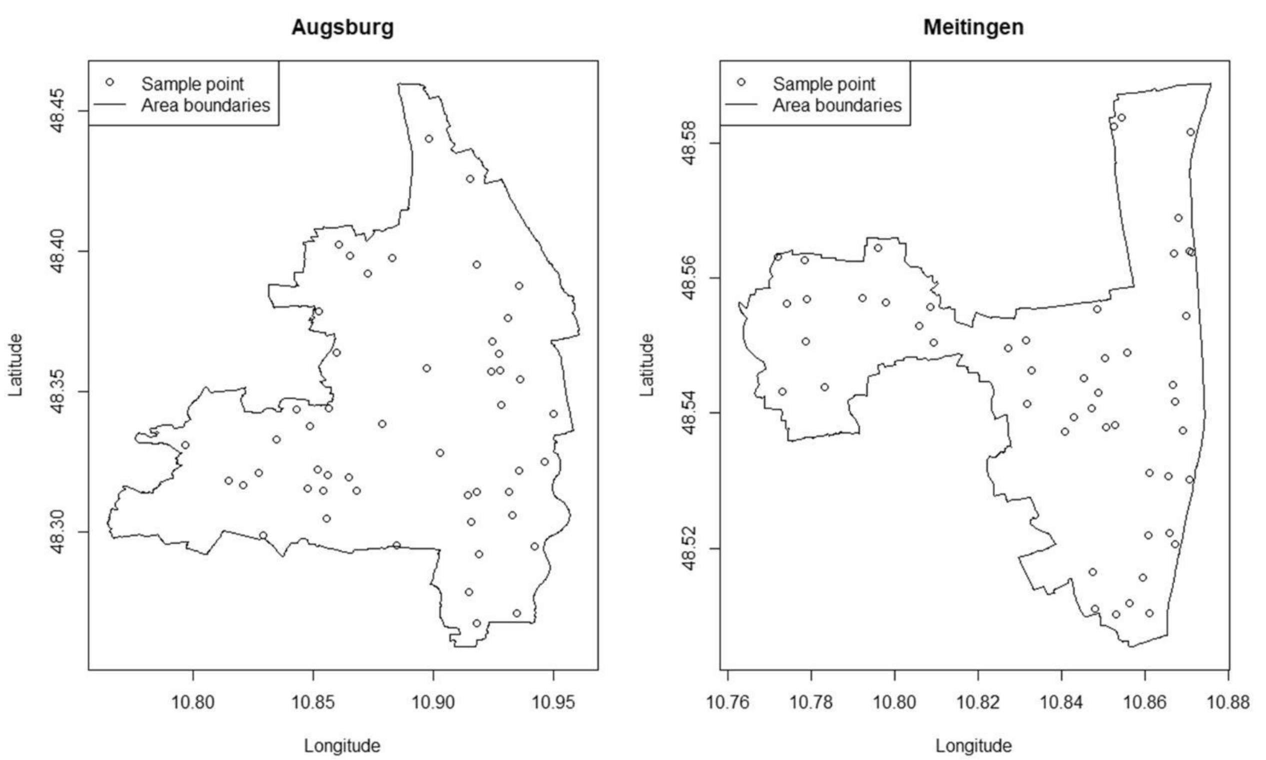

Fig. 5 Final set of randomly chosen spatial data points for Augsburg (left, $N = 50$) and Meitingen (right, $N = 50$)

northwest within the city boundaries, whereas the eastern and the southern areas showed no dense region. In contrast, kernel densities for Meitingen showed several dense regions outside the town borders.

Randomly drawn sample points

The final set of randomly chosen EPs for the evaluation of the risk score for Augsburg ($N = 50$) and Meitingen (N=50) is presented in Fig. 5. As seen within the graphics, the sample points generated via the random drawing process inside the study areas widely covered the respective regions under consideration.

Effect of parameter variation within KDE estimation process

Table 3 summarizes the values of the ANOVA F statistic for each scenario of base case and deterministic sensitivity analysis. The higher the F statistic for a given parameter variation, the higher the degree of separation between the sample point groups, and thus higher values were more preferable. Values of the F statistic for Augsburg and Meitingen increased with the amount of bandwidth for the first three scenarios. This trend continued for Meitingen with decreasing slope, whereas it showed a rather inverted u-shaped functional behavior for Augsburg. Regarding edge correction for Augsburg and Meitingen, increasing the study area to some degree led to the highest F statistic, but this effect was not permanently observed with increasing amounts of edge correction. For Augsburg, the second grid point scenario was preferred according to the ANOVA value, in contrast to Meitingen, for which the base case was preferred. Furthermore, the inverse distance weighting had the highest F value in Meitingen and Augsburg,. Finally, the second weighting scenario had a higher F value than the equal weighting scenario in Augsburg, whereas the opposite was seen for Meitingen. Overall, the bandwidth and the edge correction had the highest influence on the values of the F statistic.

Table 3 ANOVA F statistics for Augsburg and Meitingen

Augsburg		Meitingen	
Scenario[a]	F statistic[b]	Scenario[a]	F statistic[b]
BW1	871	BW1	1460
BW2	1095	BW2	3014
BW3 (BC)	1135	BW3 (BC)	4531
BW4	848	BW4	4779
BW5	657	BW5	4079
EC1 (BC)	1135	EC1 (BC)	4531
EC2	956	EC2	8372
EC3	724	EC3	9158
EC4	854	EC4	7047
EC5	1156	EC5	7912
GP1 (BC)	1135	GP1 (BC)	4531
GP2	1495	GP2	4293
GP3	1398	GP3	4530
INT1 (BC)	1135	INT1 (BC)	4531
INT2	1274	INT2	4628
INT3	1121	INT3	4306
INT4	875	INT4	6
WT1 (BC)	1135	WT1 (BC)	4531
WT2	1180	WT2	4335

BC Base case, *BW* Bandwidth, *EC* Edge correction, *GP* Grid points, *INT* Interpolation, *WT* Weighting

[a] bandwidth, edge correction, and number of grid points presented in ascending order, i.e., with the first scenario describing the least amount of bandwidth, edge correction, and number of grid points respectively

[b] The F statistic refers to the ANOVA F statistic between the groups of estimated data points at the 50 evaluation points for a given area (1000 data points at each evaluation point), calculated as follows: F = between-group variability/within-group variability; higher values of the F statistic reflect more preferable parameter values for a given scenario. The groups are generated based on the POI bootstrap replications

Obesity risk score map

Figure 6 depicts the risk score maps for Augsburg and Meitingen in which the five parameters bandwidth, edge correction, number of grid points, interpolation method and weighting, were set to their base case. The score values of the SORS are depicted as incremental densities resulting from subtracting the negative surface from the positive surface. The risk score map shows a composite picture of the separate estimates illustrated in Fig. 4. There was little heterogeneity in risk scores for the region of Augsburg except for a small area with higher obesogenic scores at the northwestern city boundary. The risk score map for Meitingen showed high risk scores for the northwestern part of the town as well as for the area north of the town borders. The risk score map based on point processes for comparison purposes is shown in Fig. 7. A region with a low obesity level is present in Figs. 6 and 7 for both study areas partly at similar places, however, obesity hotspots could only be derived from the SORS KDE plot in Fig. 6.

Discussion

We developed the SORS based on KDE using freely available data from online geocoding services. We tested several parameters which could potentially influence the final score values. Our tests showed that the SORS depended on the choice of bandwidth and the amount of edge correction applied to the KDE; the latter, however, for only one of the two study areas. In contrast, the interpolation method, the numbers of grid points, and an alternative weighting scenario had a small influence on the results.

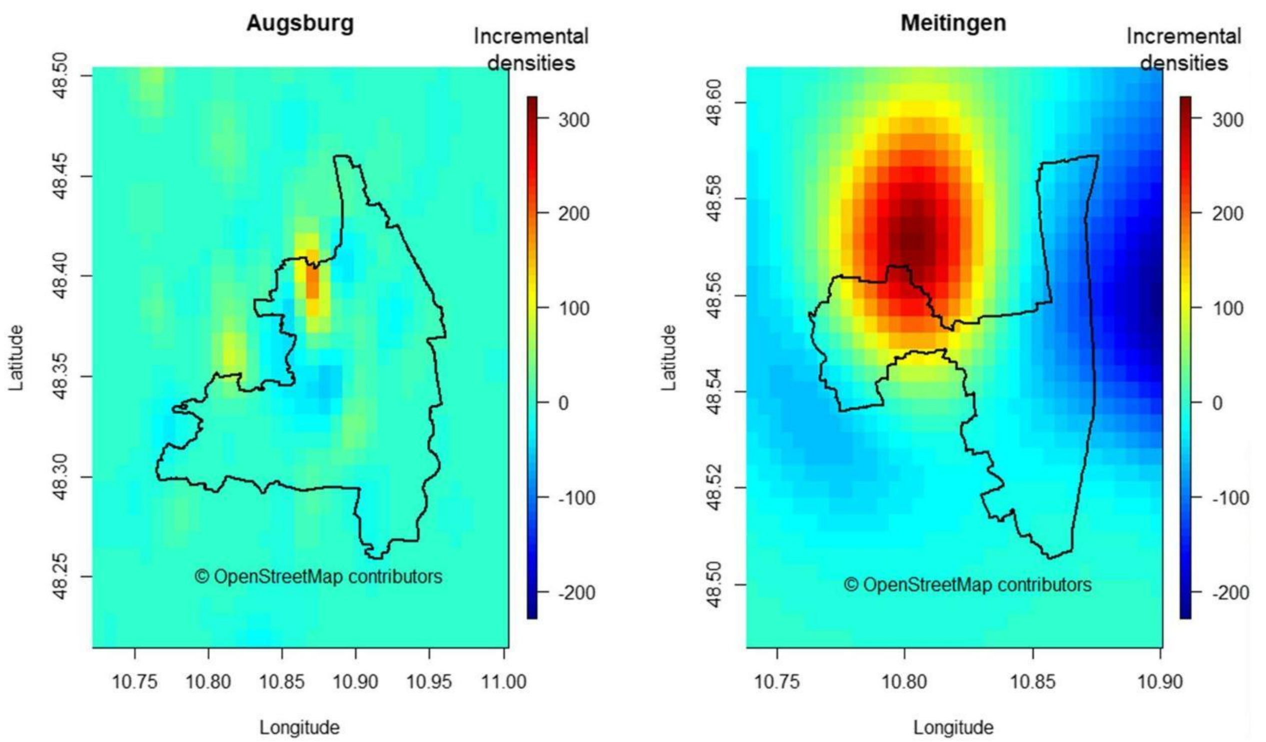

Fig. 6 Risk score maps for Augsburg (left) and Meitingen (right) showing the base case with the following parameters: bandwidth: method of Terrell [38]; edge correction: area side length: 1.4 * side length of the exact box; number of grid points: 35 × 35; weighting scheme of the environmental factors: equal weighting with unity, interpolation here via the"image.plot"function

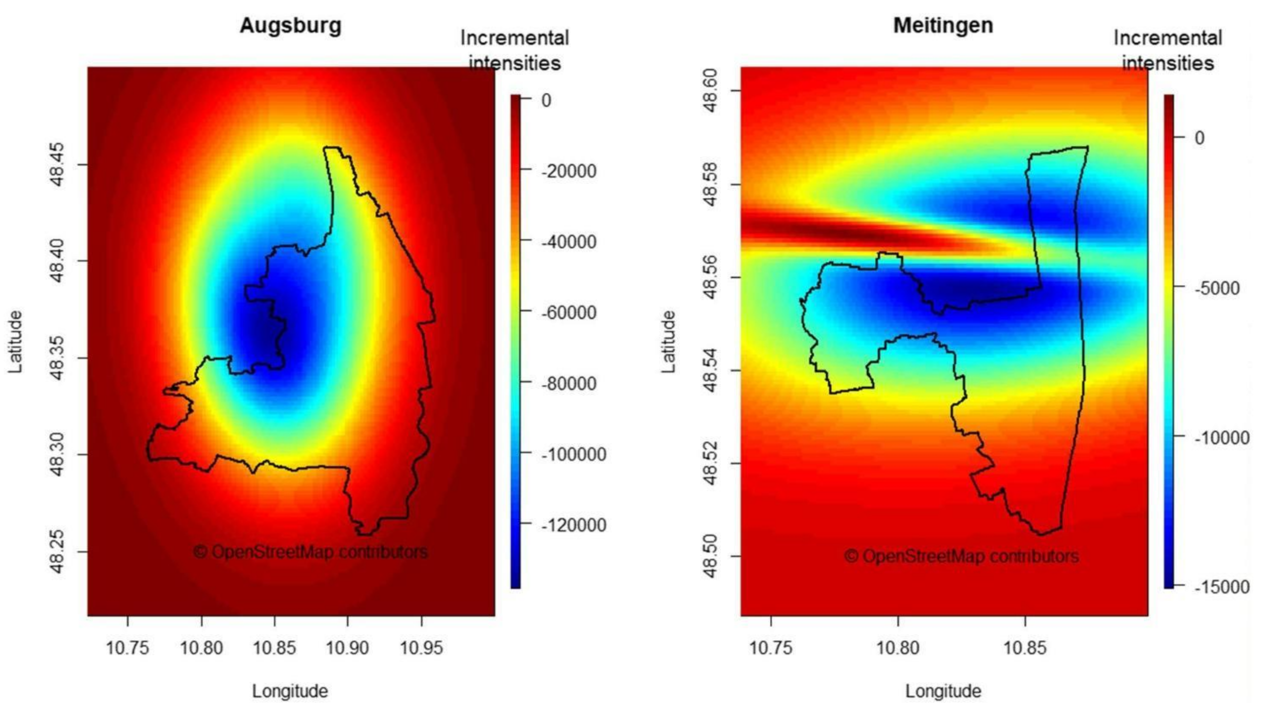

Fig. 7 Risk Score map based on incremental intensities derived from inhomogeneous cluster point processes

The SORS was calculated by taking the difference of the positive and the negative kernel density surface. We followed a similar approach to that in the work of Jones-Smith et al. [34]. They estimated correlations of their score with obesity. In contrast, the aim of our study was to investigate the effect of parameter variation on the robustness of our score. Furthermore, we covered an extensive list of obesogenic and protective environmental factors that expanded the approach of a food score to a more comprehensive measure, also including the physical activity environment. Some alternative approaches using KDE were based on the division of the kernel density surfaces [54]. A major drawback of these quotients is the issue with division by zero leading to values approaching infinity and thus leading to instability. Our approach to the SORS avoids this and also the need for some adjustments correcting for the instability.

Previous studies used various approaches to estimate risk scores based on kernel techniques, both in obesity-related research areas and elsewhere. Fitzpatrick and colleagues [55], for example, developed the keeping score based on KDE to characterize crime patterns, which has often been used by the police. Crime heat maps can be generated with this technique. This approach is based on the locations of past events instead of geolocated environmental factors, and the authors assumed that the pattern of these historic events would be maintained in the future.

Some studies created kernel density surfaces based on POIs and extracted density estimations from these surfaces in order to investigate the association with weight status. Rundle et al. (2009) analyzed the effects of environmental factors on body mass index (BMI). Results of KDE analysis concerning healthy and unhealthy food outlets were used to classify the neighborhood environment of each individual within the study based on a quintile approach [56]. Furthermore, walkability, land use mix, and population density were considered. These variables could not be implemented in our study based on the chosen POI approach with OSM data.

The five chosen SORS parameters, bandwidth, edge correction, grid points, interpolation, and weights, have also been investigated in the literature. Laraia et al. (2017) used a business software and ArcGIS to geocode the information from the study data [57]. As in our analysis, several bandwidths were tested within their KDE approach, which was found to be a sensitive model parameter. Similarly, we also found a fundamental influence of bandwidth on the results.

Effects at the edge of the study area were estimated in a simulation study concerning cluster models for food outlets [58]. Estimations at the boundaries were biased, and the authors came to the conclusion that edge effects should be corrected in studies considering measures of availability and accessibility. This underlined the importance of edge correction, which was also a major topic in our study. In addition, extending the study area has been proven to be a valuable edge correction method.

Finding the optimal number of grid points was also discussed in the literature. Some authors suggested that a choice between 100 and 500 grid points gives reasonable results [59]. In our analysis, we chose 25×25 points for the minimum bounding rectangle, i.e., 625 grid points, and chose some additional amount of edge correction for the base case. In addition, we performed some adjustments to preserve the distance between the grid points for the edge correction scenarios. In this case, the number of grid points was extended proportionally to the amount of edge correction applied, i.e., to the amount of study area extension. This made it possible to analyze grid point and edge effects separately. The choice of grid points in our base case and sensitivity analysis was chosen in accordance with default grid sizes implemented in KDE packages.

An inverse distance weighting method was applied in the past in KDE estimation regarding homicide locations as a parameter of area safety [60]. This method could be used to estimate effects at specific locations. We used such an inverse distance method in our model as an alternative to the automatic interpolation function of the base case. As a further common method, linear interpolation has been applied within the literature [61]. The "interp. surface" function applied to our SORS model was based on bilinear weights.

It was challenging to find a suitable weighting scheme applicable within our analysis. For the base case, we assumed that each factor has the same positive or negative weight, although this might look different in reality. Additionally, we tested an example from the literature [34]. We found that double weighting of supermarkets and physical activity facilities had little effect on the results. Owing to several possible weighting methods for spatial POIs, it is necessary to test further alternatives within future studies.

Finally, the SORS was graphically compared to a risk score that was derived from incremental intensities of inhomogeneous spatial point processes. Although the methodology applied here changed from KDE-based to intensity-based estimations, similar visual patterns could be derived from the two score approaches for protective patterns, which further underlines the robustness of our chosen algorithm.

Implications of the SORS on obesity-related research and policy

The SORS is a helpful tool to understand the spatial distribution of health-related harmful environmental factors in relation to health-promoting environmental factors. Risk score maps allow for an overall intuitive view on summarized structures, which can be a valuable help in obesity-related research and also within policy. Although the actual use of those structures might look different in reality, it nevertheless gives a composite simplifying measure of the environment and can be further extended to a more comprehensive tool accounting for several health dimensions affecting individuals simultaneously.

Strengths and limitations

Several strengths exist regarding our study. The automated processing of data and the automated testing of several important KDE parameters makes it possible to repeat the application of risk score estimation for other areas efficiently, given that the spatial data points and the shape files of the city or town boundaries have been downloaded before. This enables the user to describe, compare, and monitor (if done repeatedly) risk scores as well as the influence of relevant risk score parameters within several areas of interest, within other regions worldwide, and also on a larger geographic scale. For example, the analysis could be performed for a whole country in order to identify national inequalities regarding environmental obesity risks or to guide and prioritize prevention efforts that concentrate on the food and the physical activity environment. To achieve this on a regional scale, the data download area simply has to be increased to cover a larger area for the subsequent data download from OSM. The data files would be of a manageable size, as only a small number of features are important for this kind of analysis. For Augsburg, i.e., for the larger of our two study areas, the data file size was 8 MB. For larger areas, e.g., for Germany, other portals such as Geofabrik should be used. In this case, no query process is needed, and the data files are directly ready for download. The data size for Germany, for example, would be 3.1 gigabytes in this case [62]. Furthermore, using so-called planet OSM files, data disk space of around one terabyte (compressed 89 GB) or less is required [63].

We integrated uncertainty into our analysis by performing a spatial bootstrap. Subsequently, we used the samples directly for the evaluation of our method. This allowed us to assess the stability of the score values against POI variations and helped us to compare deterministic parameter scenarios based on the ANOVA F statistic. On the one hand, the impact of each parameter on score results could be assessed. In addition, the values of the F statistic could be used to find optimal parameter combinations for the SORS.

We checked the robustness of the score and repeated our analysis several times for a given area. Results were qualitatively equivalent, i.e., for each given parameter variation, the repeated analysis could be used to rank the scenarios in the same order.

However, there are also some limitations regarding the study. First, some of the environmental factors discovered during the literature search could not be implemented based on spatial POIs, especially complex constructs such as land use mix or walkability.

Second, the categorization of positive and negative obesogenic factors was based on data from pre-existing literature, and it is not known whether POIs categorized as “positive” or “negative” are really positively or negatively associated with obesogenic health (behavior). Further studies could compare the SORS with external data sources, such as walk scores in a given region, in order to test these associations [64].

As the content of OSM is generated by users, it is necessary to assess the data quality within validation studies. Within our previous work, we calculated sensitivity, specificity, and positive predictive values for OSM and compared the results with the corresponding values for Google Maps [18]. It became evident that both geocoding services performed adequately. OSM had higher positive predictive value but, in contrast, lower sensitivities than Google Maps.

Conclusion

KDE has been proven to be a useful methodology in the development of an obesity risk score, predominantly on account of the nature of the continuous estimation approach enabling efficient generation of risk score maps. However, some parameters of KDE have a large effect on score results. Parameter optimization should therefore play a major role during score model development.

Abbreviations

ANOVA	Analysis of variances
API	Application Programming Interface
BC	Base case
BS	Bootstrap sample
BW	Bandwidth
EC	Edge correction
EP	Evaluation point
GEDA	(GEDA, German Health Update)
GIS	Geographic information systems
GP	Grid points
INT	Interpolation
JSON	JavaScript Object Notation
KDE	Kernel density estimation
OSM	OpenStreetMap
POI	Point of interest
SORS	Spatial obesity risk score
WT	Weighting

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12874-023-01883-y.

Additional file 1. Complete list of chosen variables.

Additional file 2. Methodological details.

Additional file 3.

Acknowledgments

This study is partly based on a grant by the Federal Ministry of Health, Germany (FKZ: GE20160324) as part of the German Diabetes Surveillance Project conducted by the Robert Koch Institute, Berlin.

Authors' contributions

WM initiated the cooperation. RH and WM supervised the project. MP programmed the database queries, processed the data, and visualized the results on maps. RH and CK provided statistical input. All authors were involved in the development and evaluation concept of the SORS algorithm. MP programmed the SORS code. ML and WM provided health scientific input. MP wrote the manuscript. All authors read previous versions of the manuscript, commented on, and approved the final version of the manuscript for submission.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Availability of data and materials

The data generated for this study were downloaded from OSM using the tool Overpass turbo [https://overpass-turbo.eu/. Accessed 11 Aug 2022], and the exemplary code is included in this published article.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflicts of interests.

Author details

[1]Department of Sport and Health Sciences, Technical University of Munich, Munich, Germany. [2]Institute for Medical Information Processing, Biometry, and Epidemiology (IBE), Ludwig-Maximilians-University Munich, Munich, Germany. [3]Munich School of Management and Munich Center of Health Sciences, Ludwig-Maximilians-University Munich, Munich, Germany. [4]German Center for Diabetes Research, Neuherberg, Germany.

Received: 2 August 2022 Accepted: 6 March 2023
Published online: 17 March 2023

References

1. Bluher M. Obesity: global epidemiology and pathogenesis. Nat Rev Endocrinol. 2019;15:288–98.
2. Elbe AM, Elsborg P, Dandanell S, Helge JW. Correlates and predictors of obesity-specific quality of life of former participants of a residential intensive lifestyle intervention. Obes Sci Pract. 2018;4:188–93.
3. Schienkiewitz A, Mensink G, Kuhnert R, Lange C. Overweight and obesity among adults in Germany; 2017.
4. Mader S, Rubach M, Schaecke W, Röger C, Feldhoffer I, Thalmeier E-M. Healthy nutrition in Germany: a survey analysis of social causes, obesity and socioeconomic status. Public Health Nutr. 2020;23:2109–23.
5. Haftenberger M, Mensink GB, Herzog B, Kluttig A, Greiser KH, Merz B, et al. Changes in body weight and obesity status in German adults: results of seven population-based prospective studies. Eur J Clin Nutr. 2016;70:300–5.
6. Merlo J, Wagner P, Leckie G. A simple multilevel approach for analysing geographical inequalities in public health reports: the case of municipality differences in obesity. Health Place. 2019;58:102145.
7. Zhou Y, Wu K, Shen H, Zhang J, Deng H-W, Zhao L-J. Geographical differences in osteoporosis, obesity, and sarcopenia related traits in white American cohorts. Sci Rep. 2019;9:12311.
8. Elinder LS, Jansson M. Obesogenic environments--aspects on measurement and indicators. Public Health Nutr. 2009;12:307–15.
9. Hobbs M, Griffiths C, Green MA, Jordan H, Saunders J, McKenna J. Associations between the combined physical activity environment, socioeconomic status, and obesity: a cross-sectional study. Perspect Public Health. 2018;138:169–72.
10. Mackenbach JD, Rutter H, Compernolle S, Glonti K, Oppert JM, Charreire H, et al. Obesogenic environments: a systematic review of the association between the physical environment and adult weight status, the SPOTLIGHT project. BMC Public Health. 2014;14:233.
11. Rendina D, Campanozzi A, De Filippo G. Methodological approach to the assessment of the obesogenic environment in children and adolescents: a review of the literature. Nutr Metab Cardiovasc Dis. 2019;29:561–71.
12. Monsivais P, Francis O, Lovelace R, Chang M, Strachan E, Burgoine T. Data visualisation to support obesity policy: case studies of data tools for planning and transport policy in the UK. Int J Obes. 2018;42:1977–86.
13. Daley D, Bachmann M, Bachmann BA, Pedigo C, Bui MT, Coffman J. Risk terrain modeling predicts child maltreatment. Child Abuse Negl. 2016;62:29–38.
14. Caplan JM, Kennedy LW, Barnum JD, Piza EL. Risk terrain modeling for spatial risk assessment. Cityscape. 2015;17:7–16.
15. Ripoche M, Lindsay LR, Ludwig A, Ogden NH, Thivierge K, Leighton PA. Multi-scale clustering of Lyme disease risk at the expanding leading edge of the range of Ixodes scapularis in Canada. Int J Environ Res Public Health. 2018;15:603.
16. Lafontaine SJV, Sawada M, Kristjansson E. A direct observation method for auditing large urban centers using stratified sampling, mobile GIS technology and virtual environments. Int J Health Geogr. 2017;16:6.
17. Cebrecos A, Diez J, Gullon P, Bilal U, Franco M, Escobar F. Characterizing physical activity and food urban environments: a GIS-based multicomponent proposal. Int J Health Geogr. 2016;15:35.
18. Präger M, Kurz C, Bohm J, Laxy M, Maier W. Using data from online geocoding services for the assessment of environmental obesogenic factors: a feasibility study. Int J Health Geogr. 2019;18:13.
19. Lemke D, Mattauch V, Heidinger O, Hense HW. Who Hits the Mark? A Comparative Study of the Free Geocoding Services of Google and OpenStreetMap. Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany)). 2015;77:e160–5.
20. Mocnik F-B, Mobasheri A, Zipf A. Open source data mining infrastructure for exploring and analysing OpenStreetMap. Open Geospatial Data, Software and Standards. 2018;3:7.
21. Jia P, Cheng X, Xue H, Wang Y. Applications of geographic information systems (GIS) data and methods in obesity-related research. Obes Rev. 2017;18:400–11.
22. Statistische Ämter des Bundes und der Länder. Gemeinsames Statistikportal. Gemeindeverzeichnis-Online. https://www.statistikportal.de/de/produkte/gemeindeverzeichnis. Accessed 10 Jun 2020.
23. Meisinger C, Döring A, Thorand B, Heier M, Löwel H. Body fat distribution and risk of type 2 diabetes in the general population: are there differences between men and women? The MONICA/KORA Augsburg cohort study. Am J Clin Nutr. 2006;84:483–9.
24. Overpass turbo. https://overpass-turbo.eu/. Accessed 11 Aug 2021.
25. Bayerische Vermessungsverwaltung. OpenData. https://www.ldbv.bayern.de/produkte/weitere/opendata.html. Accessed 11 Aug 2021.
26. Export | OpenStreetMap. https://www.openstreetmap.org/export. Accessed 11 Aug 2021.
27. Map features. https://wiki.openstreetmap.org/wiki/Map_features. Accessed 20 Jul 2021.
28. Ripley BD. Modelling spatial patterns. J R Stat Soc Series B Stat Methodol. 1977;39:172–92.

29. Dai D, Taquechel E, Steward J, Strasser S. The impact of built environment on pedestrian crashes and the identification of crash clusters on an urban university campus. West J Emerg Med. 2010;11:294–301.
30. Haase P. Spatial pattern analysis in ecology based on Ripley's K-function: introduction and methods of edge correction. J Veg Sci. 1995;6:575–82.
31. ExplainKplot. http://spatstat.org/explainKplot.html. Accessed 11 Aug 2021.
32. Gabriel E. Estimating second-order characteristics of inhomogeneous Spatio-temporal point processes. Methodol Comput Appl Probab. 2014;16:411–31.
33. Loh JM. A valid and fast spatial bootstrap for correlation functions. Astrophys J. 2008;681:726.
34. Jones-Smith JC, Karter AJ, Warton EM, Kelly M, Kersten E, Moffet HH, et al. Obesity and the food environment: income and ethnicity differences among people with diabetes: the diabetes study of northern California (DISTANCE). Diabetes Care. 2013;36:2697–705.
35. Klemelä JS. Smoothing of multivariate data: density estimation and visualization: John Wiley & Sons; 2009.
36. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction: Springer; 2009.
37. King TL, Bentley RJ, Thornton LE, Kavanagh AM. Using kernel density estimation to understand the influence of neighbourhood destinations on BMI. BMJ Open. 2016;6:e008878.
38. Terrell GR. The maximal smoothing principle in density estimation. J Am Stat Assoc. 1990;85:470–7.
39. Davies TM, Marshall JC, Hazelton ML. Tutorial on kernel estimation of continuous spatial and spatiotemporal relative risk. Stat Med. 2018;37:1191–221.
40. Nychka D, Furrer R, Paige J, Sain S. fields: Tools for spatial data; 2017. https://doi.org/10.5065/D6W957CT. (URL: http://doi.org/10.5065/D6W957CT), R package version 9.6, <URL: www.image.ucar.edu/~nychka/Fields>
41. Wong DW, Yuan L, Perlin SA. Comparison of spatial interpolation methods for the estimation of air quality data. J Expo Anal Environ Epidemiol. 2004;14:404–15.
42. Kethireddy SR, Tchounwou PB, Ahmad HA, Yerramilli A, Young JH. Geospatial interpolation and mapping of tropospheric ozone pollution using geostatistics. Int J Environ Res Public Health. 2014;11:983–1000.
43. Wang X, Wiegand T, Wolf A, Howe R, Davies SJ, Hao Z. Spatial patterns of tree species richness in two temperate forests. J Ecol. 2011;99:1382–93.
44. R Core Team. R: a language and environment for statistical computing. Vienna: Austria; 2016.
45. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2009.
46. Baddeley A, Rubak E, Turner R. Spatial point patterns: methodology and applications with R. London: Chapman and hall/CRC press; 2015. URL http://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/9781482210200/
47. Venables WN, Ripley BD. Modern Applied Statistics with S. Fourth Edition. New York: Springer; 2002. ISBN 0-387-95457-0.
48. Bivand R, Keitt T, Rowlingson B. Rgdal: bindings for the geospatial data abstraction library. R package version 1; 2016. p. 2–4.
49. Mouselimis L. geojsonR: a GeoJson processing toolkit. R package version 1.0.0; 2017.
50. G. Jay Kerns (2018). Prob: elementary probability on finite sample spaces. R package version 1.0–1. https://CRAN.R-project.org/package=prob.
51. Pebesma EJ. Multivariable geostatistics in S: the gstat package. Comput Geosci. 2004;30:683–91.
52. Simon Urbanek (2013). Png: read and write PNG images. R package version 0.1–7. https://CRAN.R-project.org/package=png.
53. Michael Hahsler and Matthew Piekenbrock (2017). Dbscan: density based clustering of applications with noise (DBSCAN) and related algorithms. R package version 1.1–1. https://CRAN.R-project.org/package=dbscan.
54. Loo BP, Yao S, Wu J. Spatial point analysis of road crashes in Shanghai: a GIS-based network kernel density method. In: 2011 19th international conference on geoinformatics IEEE. 2011. p. 1–6.
55. Fitzpatrick DJ, Gorr WL, Neill DB. Keeping score: predictive analytics in policing. Annu Rev Criminol. 2019;2:473–91.
56. Rundle A, Neckerman KM, Freeman L, Lovasi GS, Purciel M, Quinn J, et al. Neighborhood food environment and walkability predict obesity in new York City. Environ Health Perspect. 2009;117:442–7.
57. Laraia BA, Downing JM, Zhang YT, Dow WH, Kelly M, Blanchard SD, et al. Food environment and weight change: does residential mobility matter?: the diabetes study of northern California (DISTANCE). Am J Epidemiol. 2017;185:743–50.
58. Van Meter EM, Lawson AB, Colabianchi N, Nichols M, Hibbert J, Porter DE, et al. An evaluation of edge effects in nutritional accessibility and availability measures: a simulation study. Int J Health Geogr. 2010;9:40.
59. Raykar VC, Duraiswami R, Zhao LH. Fast computation of kernel estimators. J Comput Graph Stat. 2010;19:205–20.
60. Weiss CC, Purciel M, Bader M, Quinn JW, Lovasi G, Neckerman KM, et al. Reconsidering access: park facilities and neighborhood disamenities in new York City. J Urban Health. 2011;88:297–310.
61. Van Kerm P. Adaptive kernel density estimation. Stata J. 2003;3:148–56.
62. Geofabrik downloads. Germany. https://download.geofabrik.de/europe/germany.html. Accessed 10 May 2021.
63. Planet.osm. https://wiki.openstreetmap.org/wiki/Planet.osm. Accessed 10 May 2021.
64. Carr LJ, Dunsiger SI, Marcus BH. Walk score as a global estimate of neighborhood walkability. Am J Prev Med. 2010;39:460–3.

Publisher's Note

Additional File 1: Complete list of chosen variables

Table A1: Variables separated by category

Category	Variables		
Amenity	bar §	bbq *	bicycle_parking §
	bicycle_rental §	biergarten §	boat_rental §
	boat_sharing §	bus_station *	car_rental *
	car_sharing *	cafe §	clinic §
	college §	dentist §	dive_centre §
	doctors §	dojo §	fast_food *
	fire_station §	food_court §	fountain §
	hospital §	hunting_stand §	ice_cream *
	kneipp_water_cure §	language_school §	library §
	marketplace §	nursing_home §	pharmacy §
	police §	prison §	ranger_station §
	restaurant §	school §	taxi *
	university §		
Leisure	beach_resort §	dance §	dog_park §
	fishing §	fitness_centre §	fitness_station §
	garden §	golf_course §	ice_rink §
	miniature_golf §	nature_reserve §	park §
	pitch §	playground §	sports_centre §
	stadium §	summer_camp §	swimming_area §
	swimming_pool §	track §	water_park §
Land use	allotments §	farmland §	farmyard §
	forest §	grass §	greenfield §
	greenhouse_horticulture §	meadow §	orchard §

	plant_nursery §	recreation_ground §	village_green §
	vineyard §		
Natural	fell §	grassland §	heath §
	scrub §	tree §	tree_row §
	water §	wood §	
Shop	bakery §	beverages §	bicycle §
	butcher §	cheese §	coffee §
	convenience *	dairy §	deli §
	farm §	garden_centre §	garden_furniture §
	greengrocer §	fishing §	free_flying §
	hunting §	ice_cream *	medical_supply §
	nutrition_supplements §	outdoor §	pasta §
	pastry *	scuba_diving §	seafood §
	spices §	sports §	supermarket §
	swimming_pool §	tea §	wine §
Vending	bicycle_tube §	bread §	chemist §
	coffee §	first_aid §	fishing_tackle §
	food §	ice_cream *	milk §
	sweets *		
Other variables (category)	aerodrome (aeroway) *	aquarium (tourism) §	bakery (craft) §
	bus_stop (highway) *	camp_site (tourism) §	caterer (craft) §
	halt (railway) *	helipad (aeroway) *	heliport (aeroway) *
	national_park (boundary) §	picnic_site (tourism) §	sport, whole category §
	rest_area (highway) §	station (railway) *	theme_park (tourism) §
	therapist (office) §	tram_stop (railway) *	zoo (tourism) §

* = obesogenic, § = protective

Additional File 2: Methodological details

Points of Interest download

Spatial points of interest (POIs) were downloaded as GeoJSON-Files, a file format in JavaScript Object Notation (JSON) style containing geographic data based on a key value structure, from OSM. This was executed via the online data filtering tool Overpass turbo, which enables access to the Overpass Application Programming Interface (API) of OSM [1]. Queries formulated via specific API query language were executed with this tool [2]. The tool was also used to visualize the spatial data points on an embedded map on the web page.
A rectangular area, a so-called bounding box, around each study region was chosen for the download of POIs. The size of this area was chosen large enough to allow for edge correction of the smaller contained study area in which variation in the SORS parameters was investigated.

Details regarding the coordinate adaptation and the synthetic origin

Within online geocoding services, the actual length of the longitudinal coordinate depends on the latitude to conserve conformal projections [3]. In order to correct this bias for kernel density estimation, we transformed the coordinates to adapt the length of one unit in the latitudinal direction to one unit in the longitudinal direction, i.e., we multiplied the longitudinal coordinate by a factor of 0.64 applicable within regions lying around 48 latitudinal degrees. To simplify this process, we introduced a synthetic origin of coordinates lying to the south-west of both study areas. In order to preserve original map dimensions for graphical visualization of the risk score estimates, we used the corrected risk score estimates and shifted them back to the original map dimension.

Details regarding bus stop clustering

We decided to perform a previous clustering on bus stops, as this variable seemed to be overrepresented compared with the expected impact of a single POI. Therefore, we implemented density-based spatial clustering (DBSCAN) in order to find dense regions of bus stops [4]. For risk score estimation, the obesogenic spatial POIs containing the bus stops and the protective POIs were handled altogether in two separate data layers. In order to reduce the number of bus stops into several dense clusters of bus stops, we integrated the longitude–latitude pairs of all bus stops into a data frame and applied the DBSCAN algorithm. Subsequently, we replaced the bus stop data points of the obesogenic data layer with the centroids of the discovered bus stop clusters. We assumed a cluster to be dense if at least five POIs were discovered within an epsilon neighborhood of 50 meters.

Incremental sampling process for edge correction

We extended the area and simultaneously the number of data points around the study area with increasing amount of edge correction. Therefore, we chose an incremental sampling procedure analogous to the increase in grid points across these edge correction scenarios. This improved comparability across the parameter variation. As a first step, a sample was generated for the smallest area, i.e., the base case edge correction. For each next edge correction scenario, the inner sample points from the previous sample step were maintained, and the remaining sample points were generated from the additional POIs of the gradually extended KDE estimation area. In all, this led to the generation of 1,000 incremental bootstrap samples for the obesogenic POIs and 1,000 incremental samples for the protective POIs.

References

1. Overpass turbo. https://overpass-turbo.eu/. Accessed 11 Aug 2021.

2. Overpass API/Language Guide. https://wiki.openstreetmap.org/wiki/Overpass_API/Language_Guide. Accessed 11 Aug 2021.

3. Genauigkeit von Koordinaten. https://wiki.openstreetmap.org/wiki/DE:Genauigkeit_von_Koordinaten. Accessed 25 Aug 2021.

4. Ester M, Kriegel H-P, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. 1996. p. 226–31.

Additional File 3

12874_2023_1883_MOESM3_ESM.zip

This zip file is attached to the original article on the journal's web page at:
https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-023-01883-y

Content of the zip file:
Data_download_example_Augsburg.txt
Data_preprocessing.txt
Evaluation_Points_code.txt
KDE_overlay_code.txt
Main_code_revised.txt
Ripley_code.txt

Licensing information of Article 2

The article

A spatial obesity risk score for describing the obesogenic environment using kernel density estimation: development and parameter variation
Original authors: Maximilian Präger, Christoph Kurz, Rolf Holle, Werner Maier and Michael Laxy

was published in the scientific journal "BMC Medical Research Methodology" under the terms of the following license:

4 Summary and outlook

Within this dissertation, a spatial obesity risk tool was developed that can assist policy makers and other stakeholders who have an interest in obesity-related topics. The relevant data were identified, and the operationalized data points were validated. It was possible to show the results ex-ante on maps and ex-post after data transformations were applied. Subsequently, corresponding obesity risks could be derived from the pre-processed POIs, as well as shown on composite risk maps.

A further future step in the integration process of the tool into diabetes surveillance is to examine the relationship of the obesity risk score to actual diabetes prevalence. This could be done via a regression modeling approach accounting for all relevant factors and confounders. Furthermore, it would be of interest to check whether gender or age differences exist for the applicability of the model to the respective test populations.

A policy maker could use the tool to identify structurally weak regions to understand the potential for structural prevention, as well as to tailor interventions as needed in place. It has been shown in the literature that those interventions were cost effective, thus underlining the importance of the approach outlined in this dissertation, which can be helpful to counteract rising trends in obesity prevalence (1). Regarding the German context, the tool is useful for reaching the aims of the prevention act that was established in 2015 (2,3). A further advantage could be the use of the methodology as a starting point within regions suffering from a lack of existing disease factor surveillance programs, for example in low-income countries.

The SORS algorithm fits into scalability trends seen in geo-analytics today. Some clustering algorithms for larger datasets were developed recently, such as distributed approaches for k-means clustering or the BIRCH (balanced iterative reducing and clustering using hierarchies) (4). The SORS allows researchers to extend the study area significantly without the need for distributed computing, as model run times are adequate.

Finally, the geographic tool developed in this dissertation can easily be extended to other sectors in health sciences or even beyond. It is possible to adapt the risk score concerning the choice of variables of interest and the way of processing the spatial layers. For example, one could think of provision with basic supplies such as the availability of natural water reservoirs or basic infrastructure for daily life (5). The automation process can be further extended into a general risk tool with a dynamic model that can be adapted based on a previous selection by the user.

In all, the spatial methodology for risk factor surveillance presented in this dissertation is a useful supplement to the established methods such as population-based surveys and is even more important when established surveillance methods are difficult to implement in a given context.

References

1. Sikorski C, Luppa M, Schomerus G, Werner P, König HH, Riedel-Heller SG. Public attitudes towards prevention of obesity. PLoS One. 2012;7(6):e39325.
2. Garlichs D. [The new German prevention act: an effective strategy?]. Dtsch Med Wochenschr. 2015 Oct;140(20):1543–6.
3. Bundesministerium für Gesundheit. Präventionsgesetz [Internet]. [cited 2021 Dec 29]. Available from: https://www.bundesgesundheitsministerium.de/service/begriffe-von-a-z/p/praeventionsgesetz.html
4. Lang A, Schubert E. BETULA: Fast clustering of large data with improved BIRCH CF-Trees. Information Systems. 2022 Sep 1;108:101918.
5. Map features [Internet]. 2021 [cited 2021 Jul 20]. Available from: https://wiki.openstreetmap.org/wiki/Map_features

Acknowledgment

This work was created at the Helmholtz Zentrum Munich, Institute of Health Economics and Health Care Management. Thus, I would like to thank all colleagues for the pleasant discussions and working atmosphere.

Especially, I would like to thank my first supervisor, Prof. Dr. Rolf Holle, for his support during this thesis. Furthermore, I want to thank my second supervisor, Dr. Werner Maier, who was also involved in the acquisition process for the German Diabetes Surveillance Project conducted by the Robert Koch Institute, Berlin (grant by the Federal Ministry of Health, Germany (FKZ: GE20160324)), on which this dissertation is partly based. Many thanks also to colleagues from the Robert Koch Institute, who were our cooperation partners in this project.

Finally, I would like to thank my family, especially my mother Ursula, for valuable support during the creation of my work. You have kept me motivated.

Munich, 2024