doi: 10.1111/obr.12495

Research Methods

Applications of geographic information systems (GIS) data and methods in obesity-related research

P. Jia,[1†] X. Cheng,[2†] H. Xue[3] and Y. Wang[3,4]

[1]Department of Earth Observation Science, Faculty of Geo-information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands, [2]Department of Geography, University at Buffalo, The State University of New York, Buffalo, NY, USA, [3]Fisher Institute of Health and Well-being, Systems-oriented Global Childhood Obesity Intervention Program, College of Health, Ball State University, Muncie, IN, USA, and [4]Department of Nutrition and Health Sciences, College of Health, Ball State University, Muncie, IN, USA

Received 13 October 2016; revised 1 December 2016; accepted 2 December 2016

Address for correspondence: Y Wang, Fisher Institute of Health and Well-being College of Health, Ball State University Office: HP 302E, Muncie, IN 47306, USA.
E-mail: ywang26@bsu.edu; youfawang@gmail.com

[†]These authors contributed equally to this work.

Summary

Geographic information systems (GIS) data/methods offer good promise for public health programs including obesity-related research. This study systematically examined their applications and identified gaps and limitations in current obesity-related research. A systematic search of PubMed for studies published before 20 May 2016, utilizing synonyms for GIS in combination with synonyms for obesity as search terms, identified 121 studies that met our inclusion criteria. We found primary applications of GIS data/methods in obesity-related research included (i) visualization of spatial distribution of obesity and obesity-related phenomena, and basic obesogenic environmental features, and (ii) construction of advanced obesogenic environmental indicators. We found high spatial heterogeneity in obesity prevalence/risk and obesogenic environmental factors. Also, study design and characteristics varied considerably across studies because of lack of established guidance and protocols in the field, which may also have contributed to the mixed findings about environmental impacts on obesity. Existing findings regarding built environment are more robust than those regarding food environment. Applications of GIS data/methods in obesity research are still limited, and related research faces many challenges. More and better GIS data and more friendly analysis methods are needed to expand future GIS applications in obesity-related research.

Keywords: Built environment, food environment, obesity, obesogenic environment.

Abbreviations: FF, fast food; GIS, geographic information systems; PA, physical activity.

Introduction

Obesity (including overweight) has become a serious public health threat worldwide and the second leading cause of preventable deaths trailing only tobacco (1). During 1980–2013, the combined overweight and obesity prevalence has increased from 28.8% to 36.9% for men and from 29.8% to 38.0% for women (2). Childhood overweight and obesity has also increased dramatically, especially in developed countries (from 16.9% to 23.8% for boys and from 16.2% to 22.6% for girls) (2). Obesity is associated with elevated risks for many other diseases such as coronary heart disease, stroke, hypertension, type 2 diabetes and some cancers (3). In the USA, the estimated annual costs of obesity were $147 billion in 2008, which included roughly 10% of the nation's total medical expenditure in that year (4) and were predicted to double every decade (5). Obesity-related studies are urgently needed to understand and control the obesity epidemic.

Multifaceted changes in the obesogenic environment have been suggested as the crucial underlying drivers of the growing global obesity epidemic (6). Despite an intuitive postulation that healthful (unhealthful) food and the built environment may prevent (promote) obesity, the conventional wisdom that environment influences obesity risk is not fully supported by existing studies, which have reported mixed results (7–11). This can be partially attributed to a wide range of variations in the studies, ranging from the measurement of obesogenic environments to study design and analyses. All these factors have rendered the influences of obesogenic environments insufficiently understood (12).

The rapid development of new technologies, including methods of information collection and analysis, during recent years is offering more opportunities for the development of geographic information system (GIS) methods and their applications in health-related research.

As traditional methods lack the capability to handle spatial information, the GIS, developed in the early 1960s have been gaining more attention and are increasingly being used in public health research in recent years, especially regarding obesity (13–15). GIS are computer systems aiding to capture, store, check and display data with location information, which offer many opportunities for public health programs thanks to their ability to handle complex spatial information and growing spatial data. However, despite an increase of GIS applications in obesity research over the past 2 decades, a high-level review with focusing on GIS data/methods issues is still lacking for laying out the current landscape of GIS applications in obesity research and guiding future studies.

This study was designed to fill these gaps by aiming to (i) describe the kinds of GIS data and methods used in obesity research, (ii) examine primary application areas of GIS in obesity research and related key findings and (iii) identify gaps and limitations in applications of GIS in current related research and inform future research. With increasing availability of spatial data and technologies, this study could help researchers and other relevant stakeholders with limited GIS background to understand the roles of GIS in obesity research and help add GIS components or expand GIS applications in future efforts. The technical details of GIS methods can be found in textbooks and online resources, hence were not fully covered here.

Methods and materials

Literature search

We conducted a systematic search of PubMed for related studies published before 20 May 2016, using the combination of two parts of terms as the keywords in the title or abstract field: (i) 'obesity', 'overweight', 'adiposity', 'weight status', 'body mass index', 'BMI' or 'energy balance' and (ii) 'Geographic Information System', 'Geographic Information Systems', 'Geographical Information Systems', 'Geographical Information Systems' or 'GIS'. Two of the co-authors reviewed the abstracts and chose the studies on the basis of our inclusion criteria separately. The results were cross-checked by each other and discussed with other co-authors for a final agreement on the inclusion of studies. Figure 1 shows the search and screening process.

Study inclusion criteria

Our study inclusion criteria (i) focused on obesity (including overweight) instead of other health outcomes, (ii) could focus on obesogenic environments (e.g. the food environment and built environment), or obesity-related behaviours (e.g. eating behaviour and physical activity [PA]) if not focusing on obesity, but must be related to obesity rather than on the environment/behaviour *per se*, (iii) had GIS component(s) involved (data and/or methods), (iv) were original research and (v) were published in English. A total of 121 out of 230 retrieved articles met our inclusion criteria.

Data extraction

Following the Preferred Reporting Items for Systematic Reviews and Meta-analyses framework, we extracted key information such as author information, publication year, study aim(s), study design, study area, sample size (and age if available) and key findings; in particular, we reviewed and extracted GIS components (data and methods) from each study (Appendix S1).

On the basis of the relevant research findings, we developed a framework to analyse and illustrate how GIS data and methods were integrated to construct GIS indicators and further serve obesity-related research, including its main stages: (i) GIS data collection and preparation, (ii) GIS data processing and (iii) GIS-based indicator generation (Fig. 2).

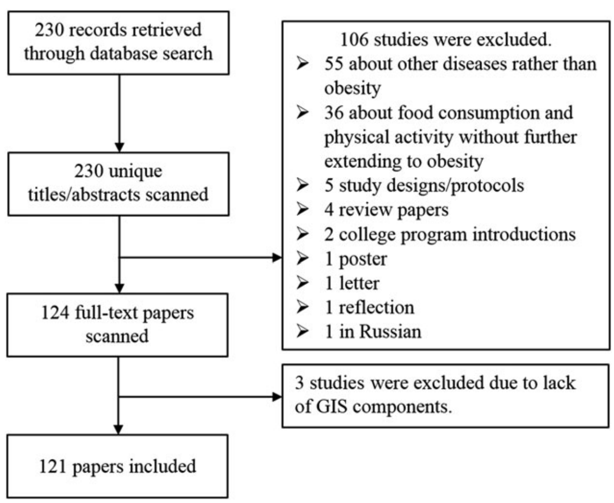

Figure 1 Study exclusions and inclusions. GIS, geographic information system.

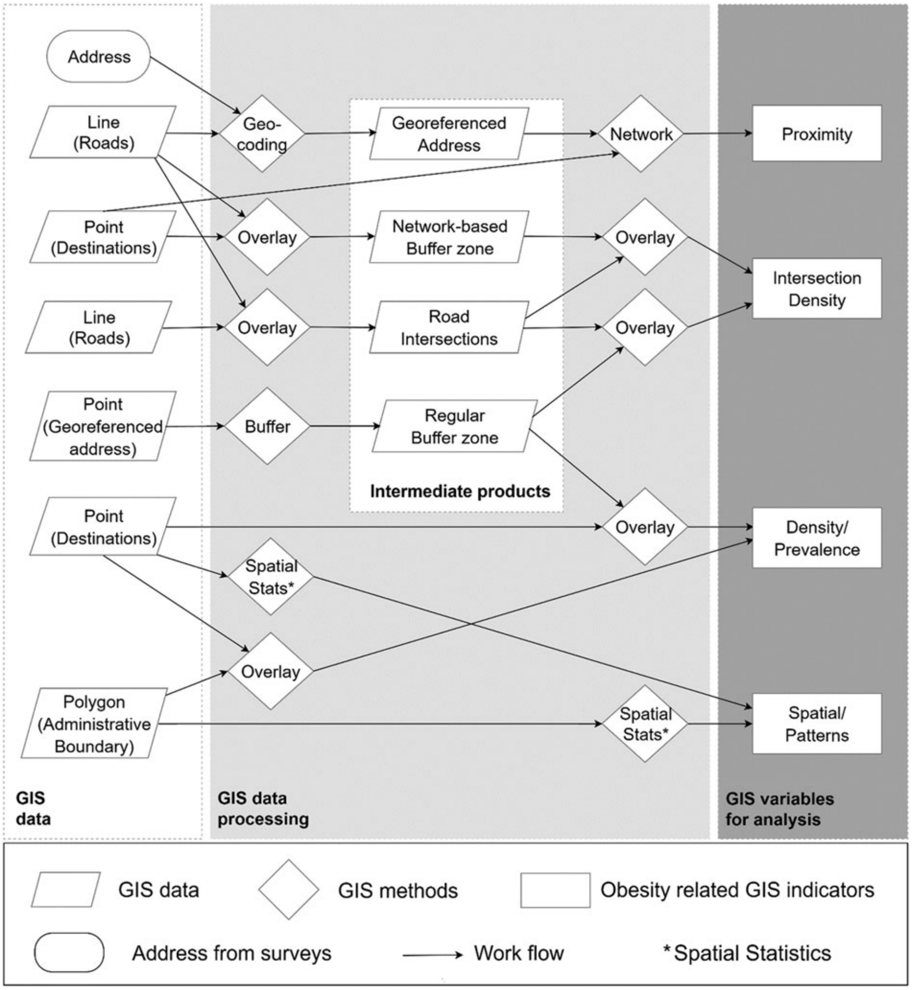

Figure 2 A workflow of geographic information system (GIS) data and analyses in obesity-related research, including three main stages: GIS data (data collection and preparation), GIS methods (data processing) and GIS indicators (GIS-based variable/indicator generation).

Results

We summarized the characteristics of and the key findings from the 121 reviewed studies. Then, we illustrated the main types of GIS data and acquisition methods, and data processing and analysis methods used in obesity research. Further, we exemplified the GIS-based obesogenic variables in obesity research.

Study characteristics

The key characteristics and findings from the 121 studies were shown in Appendix S1. Although the earliest study dated back to 2002, the majority (over three-fourths) of the included studies (96 out of 121, 79.3%) were published during 2010–2016. Most studies were conducted in the USA (81 out of 121, 66.9%), 10 in Canada, 8 in the UK, 7 in Australia, 4 in New Zealand, 2 in China, 1 international study in the USA and France and only 1 in each of Chile, Finland, France, Germany, Iran, Japan, Spain and Sweden. Most of the studies were conducted at a city level (74 of 121, 61.2%), with 13 studies including samples from more than one city (13 out of 121, 10.8%); 17 and 16 studies were at a county and state/province (or equivalent) level, respectively; the remaining 14 studies were at a national level. They were largely cross-sectional (83 of 121, 68.6%), but there were 23 ecological studies (i.e. observational studies at the population/group level, rather than individual level), 14 longitudinal studies and one focus group study. Sample sizes ranged widely from 12 to more than five million participants.

Summary of empirical findings from the studies reviewed

The 121 studies mainly addressed two types of obesity-related research questions: (i) distribution of obesity, related health behaviours and obesogenic environmental factors and (ii) the associations between obesogenic environmental factors and obesity or obesity-related health behaviours.

Spatial distribution of obesity and obesity-related factors

The topics of the reviewed studies focused on the distribution patterns of obesity-related health outcomes ranging from obesity prevalence, eating and PA, to obesogenic environmental factors such as food outlets and PA resources (16–19). Visual interpretation of the thematic maps and spatial statistics was often used to answer these types of questions. Most of the obesity-related health outcomes varied across geography. GIS served as a vital tool to detect and examine these spatial heterogeneities. For example, a national study in Iran demonstrated an uneven distribution of growth disorders (e.g. children's BMI) across provinces (19). A study in North Carolina in the USA showed that the PA resources (i.e. parks, youth services and gyms) and nutrition resources (i.e. convenience stores, fast-food [FF] establishments, restaurants and grocery stores) were not evenly distributed across the region (16). Another local study in Boston in the USA showed spatial heterogeneity in BMI *z*-score of 9th–12th grade students and a wide range of built environment features, including recreational open space, parks, bus stops, subways, and retail, service and cultural/education destinations (20).

Associations between environmental factors and obesity and obesity-related factors

The associations between environmental factors and obesity-related outcomes have been increasingly examined, especially over the past 5 years (9,21,22). The key findings on these associations were summarized in Table 1. For example, the accessibility of supermarkets and grocery stores, considered as sources of healthy food, was found to be negatively associated with BMI or obesity status (10,23,40,48). Accessibility in most of the studies was measured as (i) density of food/built environment features within zip codes or census units and (ii) proximity to food/built environment features in straight-line or real-world distances.

Some studies associated environmental factors with eating and PA behaviours and reported a consistent positive relationship between PA and density of public transportation stops, intersection density and access to open spaces (25,40,55). Some other studies reported a negative association between PA and population density (66) and positive relationships between unhealthy eating behaviour and proximity to convenience stores (67) and FF restaurants (10,67). However, the likelihood of dining in FF restaurants was not found associated with proximity to FF restaurants, whereas the likelihood of dining in non-FF restaurants was found associated with proximity to non-FF restaurants (9).

Because of the heterogeneities in the study samples and scales and differences in GIS data and analysis methods, the reported relationships between some environmental factors and obesity-related outcomes remain mixed. For example, higher land-use mix, residential density, intersection density and accessibility to recreational spaces corresponded to reduced BMI and healthy weight status in some studies (13,14,20,23,25,26,40), but not in others (20,24,26). Studies reported mixed findings regarding the associations between weight status and access to convenience stores and FF outlets: some found positive, insignificant or negative associations, respectively (7,9–11,15,23,24,32,40). Access to full-service restaurants

Table 1 Associations between obesity/overweight (or body mass index) and obesogenic environmental factors identified from the 121 included studies that used geographic information system data and/or methods*

Environment factors \ Direction of association	Negative ($p < 0.05$)	Positive ($p < 0.05$)	Not significant ($p > 0.05$)
1. Median household income	(14,15,23)	–	(24)
2. Percentage of population walking to work	(14,25,26)	–	–
3. Speed limit	–	–	(13,20,27)
4. Density of and/or proximity to			
(1) Convenience store	(15,28)	(7,10,24,28–31)	(11,24,28,30–38)
(2) Fast-food restaurant	(7,39)	(10,11,15,28,32,38–44)	(9,11,15,23,24,28,30,32–35,37,38,42–47)
(3) Grocery store	(10,23,33,48,49)	(7,49)	(31–33,37,45,46)
(4) Restaurant	(15,49)	(24,38)	(9,10,24,31,33,35,38,50)
(5) Supermarket	(10,15,23,34,37,46,48)	(11,35)	(11,24,30,32,35,36,38,45,46,51)
(6) Intersection density	(13,14,25–27,52–54)	(24)	(20,26,33,52–57)
(7) Land-use mix	(13,40,52,55)	–	(13,27,33,44,52,56,57)
(8) Park	(14,24,28,50,55,58)	–	(20,27,28,50,53,59–62)
(9) Recreational space/physical activity facility	(20,23,27,28,42,46,50,54,63,64)	–	(20,23,24,27,28,34,42,44,45,47,50,54,64)
(10) Residential density	(14,20,52)	(14)	(26,27,33,53,55,56)
(11) Side walk	(13)	(20)	(27)
(12) Green space	(25,55)	(25,44)	(25,44,55,60,65)

*Reference numbers are ordered as they appear in the text and reference list.

was found negatively associated with BMI in two studies (15,49), positively associated with BMI in another study (24), and other studies reported no significant association (9,10,24,33).

Some other studies attempted to study the inequality of accessibility of food products and PA facilities across sociodemographic groups, such as whether disadvantaged groups had poorer access to food outlets (7,17,18,68,69). The majority showed that although disadvantaged groups (e.g. minorities and low socioeconomic status) had higher obesity rates, they had better access to general food resources, PA facilities and walkable environments (7,17,68–71).

Geographic information system data

Type of geographic information system data

Broadly speaking, GIS data mean any data that are referenced with geographic location (72). There are generally two classes of GIS data: (i) vector data and (ii) raster data. Both have been used in obesity-related research (73,74). Vector data are data types that represent real-world features in the form of points, lines or polygons with geographic coordinates. For example, a household or restaurant could be simplified as a point on a map; a street could be modeled as a line; a recreational park could be presented as a polygon; and a household or restaurant could also be represented by a polygon if area matters as much as location. Each type of entity (point, line or polygon) could be stored as a separate vector layer, which could be incorporated for advanced analysis.

Raster data are a map of grids or cells with a value assigned to each grid/cell, such as color infrared high-resolution Digital Orthophoto Quarter Quadrangle images (75). The raster format is more often used for continuous variable or products, such as temperature and land use, which are also increasingly involved as natural environmental factors in obesity research (74). Another example of raster maps is density maps with each cell totalling the number of geocoded tweets with obesity-related terms within that cell (76).

Geographic information system data acquisition

Geographic information system data can primarily be obtained from three sources: government data portals, commercial datasets and researcher data collection. The first two are often existing data, whereas the third is new data collected to achieve certain aims.

Government datasets are normally free and open to the public, such as the Topologically Integrated Geographic Encoding and Referencing data regularly released by the US Census Bureau. The Topologically Integrated Geographic Encoding and Referencing data have been serving as a major data source for much local and especially nationwide obesity-related research (77). Many local governments have their own GIS departments that produce more detailed data, which are often more suitable for local-scale studies. For example, the Office of Geographic Information of Massachusetts is offering local recreational open space and road-network data including detailed sidewalk information, which was used for investigating the associations between walkable environments and children's BMI *z*-score (13,20).

Commercial data can be useful for studies such as those focused on the food environment. For example, national food retailer data in the reviewed studies could be obtained from the Reference USA (22), ESRI Business Analyst (24), InfoUSA (40) and Dun and Bradstreet data (23,67,78,79). However, a critical weakness of such commercial datasets is that different data providers may have inconsistent measuring standards or collection methods, so the targeted data from different sources may vary (80). In addition, often such data are costly to obtain and may not be amenable for research purposes.

Some researchers also collect GIS data for their own specific study objectives. For example, common ways to collect such data include using portable Global Positioning System devices and tablet personal computer technology to collect spatial data (81) and using interviews or questionnaires to collect both spatial and non-spatial data (15,21,22,67). These types of data could present more up-to-date and accurate information, but fall short of comparability with other local collections and are also difficult to be collected on a large scale because of constraints of time and labour.

Geographic information system methods

In addition to the inherent mapping and visualization functions of GIS (see section on Spatial Distribution of Obesity and Obesity-related Factors), our review of the 121 studies indicates five main categories of GIS operations being implemented for the data preparation and processing stages (Fig. 2): geocoding, overlay analysis, network analysis, buffer analysis and spatial statistics.

Geocoding

Geocoding refers to the process of converting addresses into longitude and latitude coordinates based on the so-called reference data (e.g. road network and zip code boundary) (72). An address without spatial information is just equivalent to a piece of text message. The aim of geocoding is to place these non-spatial messages into a spatial reference system. For example, in England, the spatial location of 16,956 children from the National Child Measurement Program was identified (32); in the USA, Duncan *et al*. used GIS to successfully geocode the residential addresses of nearly 50,000 children and adolescents in the electronic health records from 14 paediatric practices of Harvard Vanguard Medical Associates (13). In addition, the geocoding function has been involved in two-thirds of the reviewed studies (82 of 121, 67.8%) for identifying the location of individuals and households (Appendix S1), which would have been impossible without GIS.

Buffer analysis

Buffer analysis is used to create a regular (e.g. circular) zone with a certain radius centered on a given address/location to demarcate a catchment or influential area. It has been employed in about half of the reviewed studies (59 of 121, 48.8%), and the buffer radius chosen in these studies ranged from 200 to 8,000 m, depending on subjects and contexts (Appendix S1). Usually, a small radius is assigned when subjects are children or elderly people, and a relatively large radius is assigned for adults because of differences in mobility. For example, Hanibuchi *et al*. chose a 500-m circular buffer to define the neighborhood for elderly people (65 years and older) in Japan (11); Day and Pearce used 400- and 800-m circular buffers to measure the density of food outlets around schools (17). Most studies used a buffer radius ≥1.6 km to define the activity space for adults (9,22,66). Figure 3 illustrates a 1-km circular buffer around an individual.

Network analysis

Geographic information system-based network analysis refers to all spatial analyses conducted on the basis of a real-road network. It offers a way to identify the shortest or any path between addresses (or between centroids for areas (82)) and estimate the travel distance, or expected travel time, if the speed limit or other traffic information is provided. More than one-third (48 of 121, 39.7%) of the reviewed studies used network analysis (Appendix S1). For example, Duncan *et al*. utilized network analysis to find the path and calculate the distance from children's residential addresses to the nearest recreational open space (13). A plus is that this operation can be integrated with buffer analysis, referred to as network buffer analysis, to produce an irregular buffer zone centered on a given location on the basis of the realistic road distances involved. The resultant road-network buffer is in contrast to the straight-line buffer from traditional buffer analysis. Similar to a regular buffer, it covers the same distance (or takes the same time) to travel from any point on the boundary of a network-based irregular buffer to the center location along the shortest path (Fig. 3). For example, Ferguson *et al*. chose a series of road-network buffers (10, 20 and 30 min) to measure the accessibility to PA facilities (83). Sometimes, such a road-based buffer is also referred to as a service area (29), but note that a service area could mean something other than a road-network buffer in different contexts (84,85).

Overlay analysis

Overlay analysis often means intersecting lines or polygons to produce new features or combining multiple feature layers when needed for advanced analysis. It was used in nearly three-fourths (98 of 121, 81.0%) of the included studies. For example, a point layer of FF restaurants could be overlaid with a polygon layer of buffers of individuals' addresses, through which the availability of FF restaurants around each individual could be determined (Fig. 3). Overlay analysis was also used for the features in one layer, e.g. identifying and extracting street intersections from a road network (72).

Spatial statistics

Spatial statistics include all the methods that use topological and geographic properties of entities to analyse their spatial distribution, patterns, processes and relationships. Different from the aforementioned four methods that are often used for data preparation and processing, they are mainly employed to answer questions such as whether a phenomenon or a type of facility is randomly distributed or clustered in a certain way across space or whether two types of facilities attract or keep away from each other.

Spatial statistics were applied in 11 of the 121 studies (9.1%) for identifying spatial patterns of obesity prevalence and obesogenic environment factors (17,18,20,70,73,76,83,86–88), where cluster analysis methods such as *Moran's I* were used for measuring global spatial autocorrelation and *Local Moran's I*, *Getis* G_i^*, *K-function* and *K-Nearest Neighbor* were used to test for the presence of local spatial autocorrelation. For example, Hill *et al*. found that the distribution of food outlets and PA facilities was dispersed in a health disparate area in the Dan River region (situated in south central Virginia and north central North Carolina) (18). Day and Pearce found that food outlets were more clustered within up to 800 m around schools (17).

Many studies assumed a stationary relationship between obesity-related factors and obesity status across geography, where only one global model was built. For example, one included study used an advanced spatial modelling technique (or a local spatial statistical technique), geographically weighted regression, to better model the variable relationships between FF consumption and BMI *z*-score across the four unitary authorities of the former Avon county in the UK (7).

Geographic information system-based obesogenic variables

Usually after the data processing stage, GIS-based variables need be generated and included in traditional statistical analysis. GIS and non-GIS data are integrated and transformed into a variety of study variables, such as indicators of food and built environment factors. The following highlighted two widely used categories of GIS-based variables as examples, i.e. walkability and accessibility.

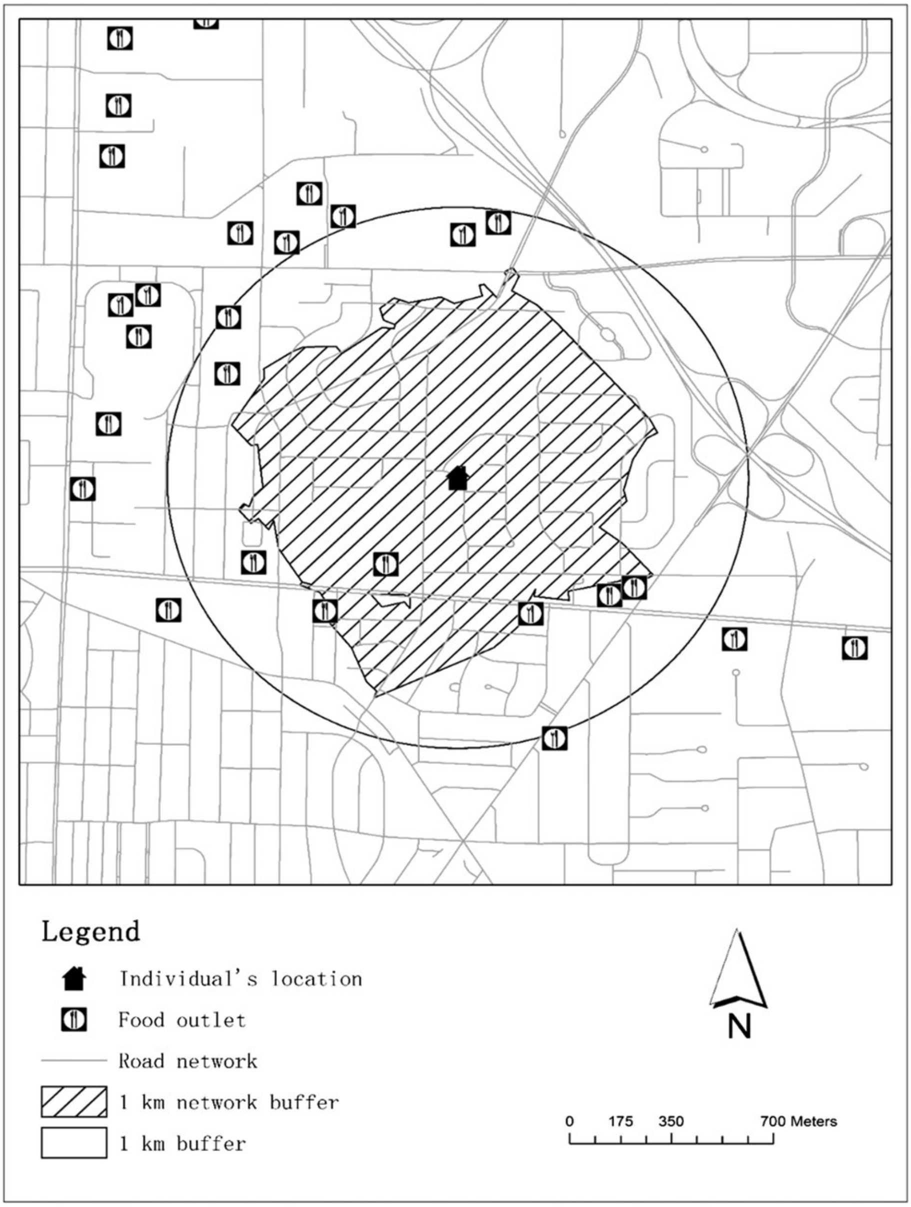

Figure 3 Illustrations of buffer, overlay and network analyses. Note: The circle represents a 1-km straight-line buffer around the individual. The shaded area represents a 1-km road-network buffer around the individual. Ten food outlets are located within the straight-line buffer of the individual, but only two of them (one looks on the boundary) are located within the road-network buffer of the individual.

Walkability of the built environment

Walkability is a measure of the friendliness of an area for walking. A high walkability is more likely to indicate a healthy built environment, protecting residents from obesity by providing a greater chance for outdoor PA. Pikora *et al*. summarized four major aspects of built environment factors that contributed primarily to the walkability of a given neighborhood, including (i) functional, referring to the physical attributes of the streets, (ii) safety, (iii) aesthetics, meaning the pleasure of the environment and (iv) destinations, such as the number and diversity of the facilities (89).

A variety of natural environmental factors are included under four categories, and most need be measured accurately using GIS. For example, street intersection density is one of the most important factors because of its ability to capture the wellness of connections between destinations, which can be calculated by (i) using overlay analysis to produce a point layer of street intersections from a road layer, (ii) using overlay analysis again to combine the point layer and a polygon layer (i.e. census tract) and (iii) using spatial arithmetic operations to count the number of street intersections within polygons and calculate their density. Some other studies reviewed used buffer analysis to produce buffers around point features, replacing the existing polygon layers (20,24–26,51). The ease of measurement by GIS at a large scale (e.g. nationwide or regional) has made walkability the most widely used GIS measure in obesity research (8,20,55).

Accessibility to obesogenic environment features

Accessibility is ubiquitous in obesogenic environmental studies, such as an individual's access to a variety of stores and facilities including supermarkets, restaurants and recreational facilities. In addition to reflecting the availability of the venues to some degree (15,22,51,55,90), accessibility is more inherently a geographic measure of the ease of reaching a venue, which would be impossible to calculate without the support of GIS, especially at a large scale (11,22,23). For example, provided that addresses of individuals and food venues around them are both available, a simple version of individuals' accessibility to food venues can be calculated by (i) using geocoding analysis to locate individuals and food venues on the map and (ii) using buffer analysis to create a buffer (with a certain radius) around each individual for counting the number of food venues within the buffer. In a highly resource-competitive context where each food venue can only serve a small portion of the local residents at a time, the aforementioned second step could be further expanded into two steps: (i) creating a buffer around each food venue and calculating the ratio of that food venue to the individuals within the buffer (1: n, n = number of individuals) and (ii) creating a buffer around each individual and summing up the ratios within the buffer as the accessibility of that person to food venues.

The process described here simply assumes that each food venue is equally accessible to all individuals within its buffer regardless of physical barriers between individuals and food venues in the real world. Network analysis could make this measure more realistic by using a real-world distance as the radius of the buffers in place of a straight-line distance or by designating different weights to each individual around a certain venue based on the real-world travel distances between them (83).

Discussion

Research using various GIS data and methods in the public health field has been growing rapidly over the past decade. The growing global obesity epidemic has stimulated such work. Our study shows that spatial distribution and association analyses were the major applications of GIS in the current literature of obesity research, where the spatial distribution of and potential environmental impacts on the obesity epidemic were examined, respectively. The obesity prevalence, obesity-related behaviours and obesogenic environmental factors are found unevenly distributed across geography. The related mainstream GIS applications include three main stages, namely, data acquisition, data processing and GIS-based variable generation and analysis. Government and commercial GIS data are the main data sources. The GIS methods currently most widely used include geocoding, buffer, network, overlay and spatial statistical analyses.

In addition, our review suggests that (i) the relationships between environmental factors and obesity remain mixed and (ii) GIS should be used more consistently and more in depth in obesity research for measuring individuals' interactions with the obesogenic environment.

Although GIS has been increasingly used in various aspects of obesity research, our review reveals several key research gaps. More efforts are needed before GIS could be fully incorporated into the conceptual framework and implementation of traditional obesity research on related health problems. First, GIS data have limitations. Accurate geocoding has not been widely used in public health studies because of unavailable personal addresses in most studies; ergo, the straight-line and road-network buffers in most studies were established around an area unit (e.g. census unit or zip code) rather than an address point, which created difficulty in measuring the realistic exposures of residents to the surrounding obesogenic environment. Using the zip code system as a unit for such analysis is problematic, because zip codes were initially developed for delivery purposes, and demographic attributes are more likely to be heterogeneous within zip codes than, for instance, within census units.

As an example, thus, far GIS applications in studying food environment and its impact on obesity are restricted in that they measure only the community food environment, e.g. the number, location, type and accessibility of food outlets in individuals' residential communities. There are other critical dimensions of the food environment, e.g. (i) consumer: the food environment that consumers actually interact with instead of what they only have access to (e.g. the food outlets or markets in which the consumers actually buy food), including such factors as availability and quality of healthy food, pricing, promotion, placement of food and food information (e.g. nutrition labelling); (ii) organizational: the food environment not only in the community but also at home, school and the workplace as a whole; and (iii) information: the influences of government and industry policies on public attitudes and the appeal of certain foods and food sources that are created via media reporting and advertising (91). However, limited available data can support such research (92).

We also found limited use of advanced GIS data analysis methods in current research, especially spatial statistics, whose use remains at a descriptive stage, such as seeking spatio-temporal clustering patterns of obesity rates.

Additionally, separate built environment indicators have been simultaneously included in a single analysis model in many existing studies, which may result in an over-control issue (93).

We recommend some future directions for GIS applications in public health research, in particular, regarding obesity. First, change the most basic reporting unit of personal residence from zip code to census unit. This would represent quite an advancement not only in accuracy but also in the measurement of individual's exposure to the obesogenic environment, which would provide a better match with census data. Second, collect more accurate and timely GIS data. Volunteered GIS means citizens voluntarily use Internet-enabled mobile devices to share their real-time location information via mobile apps (e.g. Facebook and Twitter), which can be a potential solution for overcoming the current data limitation (94). Third, enrich GIS data collection to measure multiple dimensions of the food environment in order to assess their impacts on people's behaviours. For example, besides food accessibility, areal-level socioeconomic data from the Census could be used to measure food affordability. Moreover, develop friendly data processing and analysis methods to enable more researchers to process, combine and analyse GIS and other types of data. This can include regression methods that take into account spatial patterns, such as geographically weighted regression, spatial lag regression, spatial error regression and spatial multi-level regression, to investigate spatial inequalities of the associations between environmental factors and health outcomes.

This study has some limitations. Because of the focus on GIS applications in obesity research (rather than empirical results) and a limited number of such related studies indexed in PubMed, this review is not intended to and does not provide a complete list of potential determinants of obesity and conclusive evidence regarding the effects of included determinants on obesity. We used obesity as an example to assess the application of GIS in related research, but did not assess other health outcomes. Future research on other outcomes is warranted. In addition, we could not evaluate and compare the quality of the GIS data and methods used in the reviewed studies, which would deviate from our study aims. Another future study could be further differentiating findings by methods (e.g. *Local Moran's I* vs. *Getis* G_i^*) and even implementations of the same methods (e.g. straight-line vs. road-network buffer analyses), which will establish evidence for appropriateness of each method or implementation in obesity context. Nevertheless, this study would assist readers to (i) comprehend the importance of using GIS in obesity-related research, (ii) understand the versatility of GIS in public health research using obesity as an example and (iii) be aware of the research gaps and challenges in using GIS in obesity context.

Conclusion

Geographic information system provides promising opportunities for studying many public health questions, including to visually illustrate the spatio-temporal distribution and changes of health outcomes and to study the effects of environmental exposures. To expand and advance GIS applications in obesity and public health research, efforts need be made to educate and empower researchers about GIS, what it offers and how to use it. Incorporation of GIS in study design, data collection and analysis is important. Findings from this study will assist future research using GIS data and methods.

Acknowledgements

The study is supported by research grants from the US National Institute of Health (NIH), the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD, 1R01HD064685-01A1 and U54HD070725) and the National Institute of Diabetes and Digestive and Kidney (R01DK81335-01A1). The U54 project is co-funded by the NICHD and the Office of Behavioral and Social Sciences Research. Dr Youfa Wang is the principal investigator of these NIH-funded projects. The content of the paper is solely the responsibility of the authors and does not necessarily represent the official views of the funder. We thank team member Dr Paula Vincent for helping improve the manuscript.

Conflict of interest statement

No conflict of interest was declared.

Authors' contributions

Y. W. directed the whole study process and provided administrative support including funding support. P. J. and X. C. conducted the literature review, extracted the data and drafted the manuscript. Y. W. and H. X. revised the manuscript. All authors have contributed to data collection, analysis, interpretation of the results and manuscript drafting and have read and approved the final manuscript.

Supporting information

Additional Supporting Information may be found in the online version of this article, http://dx.doi.org/10.1111/obr.12495

Appendix S1. Summary of key study characteristics, GIS data and methods used, and key findings of the 121 included studies (ranked by the alphabetical order of the first authors' last name)

References

1. Wang Y, Lobstein T. Worldwide trends in childhood overweight and obesity. *International Journal of Pediatric Obesity*. 2006; **1**: 11–25.
2. Ng M, Fleming T, Robinson M *et al*. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*. 2014; **384**: 766–781.
3. Centers for Disease Control and Prevention. Obesity, halting the epidemic by making health easier, at a glance 2011. Centers for Disease Control and Prevention: CDC webpage 2011.
4. Finkelstein EA, Trogdon JG, Cohen JW, Dietz W. Annual medical spending attributable to obesity: payer-and service-specific estimates. *Health Affairs*. 2009; **28**: w822–ww31.
5. Glicksman A, Ring L, Kleban M, Hoffman C. Is "walkability" a useful concept for gerontology? *Journal of Housing for the Elderly*. 2013; **27**: 241–254.
6. Egger G, Swinburn B. An"ecological" approach to the obesity pandemic. *BMJ: British Medical Journal*. 1997; **315**: 477.
7. Fraser L, Edwards K, Tominitz M, Clarke G, Hill A. Food outlet availability, deprivation and obesity in a multi-ethnic sample of pregnant women in Bradford. *UK. Social Science & Medicine*. 2012; **75**: 1048–1056.
8. Harris D, Aboueissa A, Hartley D. Myocardial infarction and heart failure hospitalization rates in Maine. *USA—variability along the urban–rural continuum. Rural Remote Health*. 2008; **8**: 980.
9. Jeffery RW, Baxter J, McGuire M, Linde J. Are fast food restaurants an environmental risk factor for obesity? *International Journal of Behavioral Nutrition and Physical Activity*. 2006; **3**: 2.
10. Jennings A, Welch A, Jones AP *et al*. Local food outlets, weight status, and dietary intake: associations in children aged 9–10 years. *American Journal of Preventive Medicine*. 2011; **40**: 405–410.
11. Hanibuchi T, Kondo K, Nakaya T *et al*. 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.
12. French SA, Story M, Jeffery RW. Environmental influences on eating and physical activity. *Annual Review of Public Health*. 2001; **22**: 309–335.
13. Duncan DT, Sharifi M, Melly SJ *et al*. Characteristics of walkable built environments and BMI *z*-scores in children: evidence from a large electronic health record database. *Environmental Health Perspectives*. 2014; **122**: 1359–1365.
14. Wen M, Kowaleski-Jones L. The built environment and risk of obesity in the United States: racial–ethnic disparities. *Health & Place*. 2012; **18**: 1314–1322.
15. Li Y, Robinson L, Carter W, Gupta R. Childhood obesity and community food environments in Alabama's Black Belt region. *Child: Care, Health and Development*. 2015; **41**: 668–676.
16. Corsino L, McDuffie JR, Kotch J *et al*. Achieving health for a lifetime: a community engagement assessment focusing on school-age children to decrease obesity in Durham. *North Carolina. North Carolina Medical Journal*. 2013; **74**: 18–26.
17. Day PL, Pearce J. Obesity-promoting food environments and the spatial clustering of food outlets around schools. *American Journal of Preventive Medicine*. 2011; **40**: 113–121.
18. Hill JL, Chau C, Luebbering CR, Kolivras KK, Zoellner J. Does availability of physical activity and food outlets differ by race and income? Findings from an enumeration study in a health disparate region. *International Journal of Behavioral Nutrition and Physical Activity* 2012; **9**: 105.
19. Kelishadi R, Amiri M, Motlagh ME *et al*. Growth disorders among 6-year-old Iranian children. *Iranian Red Crescent Medical Journal* 2014; **16** : e6761.
20. 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.
21. O'Connor T, Cerin E, Lee RE *et al*. Environmental and cultural correlates of physical activity parenting practices among Latino parents with preschool-aged children: Ninos Activos. *BMC public health*. 2014; **14**: 707.
22. Jilcott Pitts SB, Keyserling TC, Johnston LF *et al*. Associations between neighborhood-level factors related to a healthful lifestyle and dietary intake, physical activity, and support for obesity prevention polices among rural adults. *Journal of Community Health*. 2015; **40**: 276–284.
23. Casey R, Chaix B, Weber C *et al*. Spatial accessibility to physical activity facilities and to food outlets and overweight in French youth. *International Journal of Obesity*. 2012; **36**: 914–919.
24. Wall MM, Larson NI, Forsyth A *et al*. Patterns of obesogenic neighborhood features and adolescent weight: a comparison of statistical approaches. *American Journal of Preventive Medicine*. 2012; **42**: e65–e75.
25. Coombes E, Jones AP, Hillsdon M. The relationship of physical activity and overweight to objectively measured green space accessibility and use. *Social Science & Medicine*. 2010; **70**: 816–822.
26. 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**: 237–244.
27. 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 : kwt111.
28. Larson N, Wall M, Story M, Neumark-Sztainer D. Home/family, peer, school, and neighborhood correlates of obesity in adolescents. *Obesity*. 2013; **21**: 1858–1869.
29. Jilcott Pitts 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**: 1610–1617.
30. 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.
31. 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**: 1757–1763.
32. Williams J, Scarborough P, Townsend N *et al*. Associations between food outlets around schools and BMI among primary students in England: a cross-classified multi-level analysis. *PloS One*. 2015; **10** : e0132930.
33. Cerin E, Frank LD, Sallis JF *et al*. From neighborhood design and food options to residents' weight status. *Appetite*. 2011; **56**: 693–703.
34. Boone-Heinonen J, Diez-Roux AV, Goff DC *et al*. The neighborhood energy balance equation: does neighborhood food retail environment + physical activity environment = obesity? *The CARDIA study. PLoS One*. 2013; **8** : e85141.

35. Fiechtner L, Block J, Duncan DT *et al*. Proximity to supermarkets associated with higher body mass index among overweight and obese preschool-age children. *Preventive Medicine*. 2013; **56**: 218–221.
36. 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**: 526–541.
37. Liu GC, Wilson JS, Qi R, Ying J. Green neighborhoods, food retail and childhood overweight: differences by population density. *American Journal of Health Promotion*. 2007; **21**: 317–325.
38. Richardson AS, Meyer KA, Howard AG *et al*. 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.
39. 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.
40. Li F, Harmer PA, Cardinal BJ *et al*. Built environment, adiposity, and physical activity in adults aged 50–75. *American Journal of Preventive Medicine*. 2008; **35**: 38–46.
41. Fraser LK, Edwards KL. The association between the geography of fast food outlets and childhood obesity rates in Leeds, UK. *Health & Place*. 2010; **16**: 1124–1128.
42. Oreskovic NM, Winickoff JP, Kuhlthau KA, Romm D, Perrin JM. Obesity and the built environment among Massachusetts children. *Clinical Pediatrics* 2009.
43. 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**: 315–321.
44. Burgoine T, Jones AP, Brouwer RJN, Neelon SEB. Associations between BMI and home, school and route environmental exposures estimated using GPS and GIS: do we see evidence of selective daily mobility bias in children? *International Journal of Health Geographics* 2015; **14**: 1.
45. Crawford D, Ball K, Cleland V *et al*. Maternal efficacy and sedentary behavior rules predict child obesity resilience. *BMC obesity*. 2015; **2**: 1.
46. Abbott G, Backholer K, Peeters A, Thornton L, Crawford D, Ball K. Explaining educational disparities in adiposity: the role of neighborhood environments. *Obesity*. 2014; **22**: 2413–2419.
47. Burdette HL, Whitaker RC. Neighborhood playgrounds, fast food restaurants, and crime: relationships to overweight in low-income preschool children. *Preventive medicine*. 2004; **38**: 57–63.
48. 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. *Am J Health Promot*. 2014; **28**: e137–e145.
49. Zhang J, Xue H, Cheng X *et al*. Influence of proximities to food establishments on body mass index among children in China. *Asia Pacific Journal of Clinical Nutrition*. 2016; **25**: 134–141.
50. 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.
51. 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*. 2014; **38**: 306–314.
52. Pouliou T, Elliott SJ. Individual and socio-environmental determinants of overweight and obesity in Urban Canada. *Health & Place*. 2010; **16**: 389–398.
53. 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**: 626–642.
54. 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. *International Journal of Behavioral Nutrition and Physical Activity*. 2010; **7**: 1.
55. 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**: 569–578.
56. 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**: 57–60.
57. Gebel K, Bauman AE, Sugiyama T, Owen N. Mismatch between perceived and objectively assessed neighborhood walkability attributes: prospective relationships with walking and weight gain. *Health & Place*. 2011; **17**: 519–524.
58. Mena C, Fuentes E, Ormazábal Y, Palomo-Vélez 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**: 373–383.
59. Lovasi GS, Schwartz-Soicher O, Quinn JW *et al*. Neighborhood safety and green space as predictors of obesity among preschool children from low-income families in New York City. *Preventive Medicine*. 2013; **57**: 189–193.
60. 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**: 434–445.
61. Potestio ML, Patel AB, Powell CD, McNeil DA, Jacobson DR, McLaren L. Is there an association between spatial access to parks/green space and childhood overweight/obesity in Calgary, Canada? *International Journal of Behavioral Nutrition and Physical Activity*. 2009; **6**: 77.
62. 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**: 299–303.
63. 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**: 417–424.
64. Gutiérrez-Zornoza M, Sánchez-López M, García-Hermoso A, González-García A, Chillón P, Martínez-Vizcaíno V. Active commuting to school, weight status, and cardiometabolic risk in children from rural areas the Cuenca study. *Health Education & Behavior*. 2014 : 1090198114549373.
65. Kyttä A, Broberg A, Kahila M. Urban environment and children's active lifestyle: softGIS revealing children's behavioral patterns and meaningful places. *American Journal of Health Promotion*. 2012; **26**: e137–e148.
66. Yang W, Spears K, Zhang F, Lee W, Himler HL. Evaluation of personal and built environment attributes to physical activity: a multilevel analysis on multiple population-based data sources. *Journal of Obesity*. 2012; **2012** : e548910.
67. Leslie T, Pawloski L, Kallman-Price J *et al*. Survey of health status, nutrition and geography of food selection of chronic liver disease patients. *Annals of Hepatology: Official Journal of the Mexican Association of Hepatology*. 2014; **13**: 533–540.

68. Sadler RC, Gilliland JA, Arku G. An application of the edge effect in measuring accessibility to multiple food retailer types in Southwestern Ontario. *Canada. International Journal of Health Geographics*. 2011; **10**: 34.
69. Kestens Y, Daniel M. Social inequalities in food exposure around schools in an urban area. *American Journal of Preventive Medicine*. 2010; **39**: 33–40.
70. Zhu X, Lee C. Walkability and safety around elementary schools: economic and ethnic disparities. *American Journal of Preventive Medicine*. 2008; **34**: 282–290.
71. Weiss CC, Purciel M, Bader M *et al.* Reconsidering access: park facilities and neighborhood disamenities in New York City. *Journal of Urban Health*. 2011; **88**: 297–310.
72. Chang, K-t. Introduction to geographic information systems: McGraw-Hill Higher Education Boston 2006.
73. Duncan DT, Aldstadt J, Whalen J, Melly SJ, Gortmaker SL. Validation of Walk Score® for estimating neighborhood walkability: an analysis of four US metropolitan areas. *International Journal of Environmental Research and Public Health*. 2011; **8**: 4160–4179.
74. Michimi A, Wimberly MC. Natural environments, obesity, and physical activity in nonmetropolitan areas of the United States. *J Rural Health*. 2012; **28**: 398–407.
75. Kim J-H, Lee C, Sohn W. Urban natural environments, obesity, and health-related quality of life among Hispanic children living in inner-city neighborhoods. *International Journal of Environmental Research and Public health*. 2016; **13**: 121.
76. Ghosh D, Guha R. What are we 'tweeting' about obesity? Mapping tweets with topic modeling and geographic information system. *Cartography and Geographic Information Science*. 2013; **40**: 90–102.
77. US Census Bureau. TIGER/Line Shapefiles. U.S. Department of Commerce: U.S. Department of Commerce website 2014.
78. Frank LD, Saelens BE, Chapman J *et al.* Objective assessment of obesogenic environments in youth: geographic information system methods and spatial findings from the Neighborhood Impact on Kids study. *American Journal of Preventive Medicine*. 2012; **42**: e47–e55.
79. Ghirardelli A, Quinn V, Foerster SB. Using geographic information systems and local food store data in California's low-income neighborhoods to inform community initiatives and resources. *American Journal of Public Health*. 2010; **100**: 2156–2162.
80. Boone JE, Gordon-Larsen P, Stewart JD, Popkin BM. Validation of a GIS facilities database: quantification and implications of error. *Annals of Epidemiology*. 2008; **18**: 371–377.
81. Coakley HL, Steeves EA, Jones-Smith JC *et al.* Combining ground-truthing and technology to improve accuracy in establishing children's food purchasing behaviors. *Journal of Hunger and Environmental Nutrition*. 2014; **9**: 418–430.
82. Jia P, Xierali I. Disparities in patterns of health care travel among inpatients diagnosed with congestive heart failure, Florida, 2011. *Preventing Chronic Disease*. 2015; **12** : E150.
83. Ferguson NS, Lamb KE, Wang Y, Ogilvie D, Ellaway A. Access to recreational physical activities by car and bus: an assessment of socio-spatial inequalities in mainland Scotland. *PloS One*. 2013; **8**: e55638.
84. Jia P, Wang F, Xierali I. Delineating hierarchical hospital service areas in Florida. *Geographical Review*. 2017; **107**.
85. Jia P, Xierali I, Wang F. Evaluating and re-demarcating the hospital service areas in Florida. *Applied Geography*. 2015; **60**: 248–253.
86. Huang R, Moudon A, Cook A, Drewnowski A. The spatial clustering of obesity: does the built environment matter? *Journal of Human Nutrition and Dietetics*. 2015; **28**: 604–612.
87. Maroko AR, Maantay JA, Sohler NL, Grady KL, Arno PS. The complexities of measuring access to parks and physical activity sites in New York City: a quantitative and qualitative approach. *International Journal of Health Geographics*. 2009; **8**: 1.
88. Penney T, Rainham D, Dummer T, Kirk S. A spatial analysis of community level overweight and obesity. *Journal of Human Nutrition and Dietetics*. 2014; **27**: 65–74.
89. Pikora T, Giles-Corti B, Bull F, Jamrozik K, Donovan R. Developing a framework for assessment of the environmental determinants of walking and cycling. *Social Science & Medicine*. 2003; **56**: 1693–1703.
90. Cutts BB, Darby KJ, Boone CG, Brewis A. City structure, obesity, and environmental justice: an integrated analysis of physical and social barriers to walkable streets and park access. *Social Science & Medicine*. 2009; **69**: 1314–1322.
91. Glanz K, Sallis JF, Saelens BE, Frank LD. Healthy nutrition environments: concepts and measures. *American Journal of Health Promotion*. 2005; **19**: 330–333.
92. Kerr J, Sallis JF, Owen N *et al.* Advancing science and policy through a coordinated international study of physical activity and built environments: IPEN adult methods. *Journal of Physical Activity and Health*. 2013; **10**: 581–601.
93. Little TD. Longitudinal Structural Equation Modeling: Guilford Press 2013.
94. Goodchild MF. Citizens as sensors: the world of volunteered geography. *GeoJournal*. 2007; **69**: 211–221.

