



LEEDS
BECKETT
UNIVERSITY

Citation:

Wilkins, EL and Radley, D and Morris, M and Hobbs, M and Christensen, A and Lameck Marwa, W and Morrin, A and Griffiths, C (2019) A systematic review employing the Ge-oFERN framework to examine methods, reporting quality and associations between the Retail Food Environment and obesity. *Health and Place*, 57. pp. 186-199. ISSN 1353-8292 DOI: <https://doi.org/10.1016/j.healthplace.2019.02.007>

Link to Leeds Beckett Repository record:

<https://eprints.leedsbeckett.ac.uk/id/eprint/5746/>

Document Version:

Article (Published Version)

Creative Commons: Attribution 4.0

The aim of the Leeds Beckett Repository is to provide open access to our research, as required by funder policies and permitted by publishers and copyright law.

The Leeds Beckett repository holds a wide range of publications, each of which has been checked for copyright and the relevant embargo period has been applied by the Research Services team.

We operate on a standard take-down policy. If you are the author or publisher of an output and you would like it removed from the repository, please [contact us](#) and we will investigate on a case-by-case basis.

Each thesis in the repository has been cleared where necessary by the author for third party copyright. If you would like a thesis to be removed from the repository or believe there is an issue with copyright, please contact us on openaccess@leedsbeckett.ac.uk and we will investigate on a case-by-case basis.



ELSEVIER

Contents lists available at ScienceDirect

Health & Place

journal homepage: www.elsevier.com/locate/healthplace

Review Essay

A systematic review employing the GeoFERN framework to examine methods, reporting quality and associations between the retail food environment and obesity

Emma Wilkins^{a,*}, Duncan Radley^a, Michelle Morris^b, Matthew Hobbs^{a,c}, Alex Christensen^a, Windi Lameck Marwa^a, Adele Morrin^a, Claire Griffiths^a

^a Carnegie School of Sport, Leeds Beckett University, Leeds, UK

^b Leeds Institute for Data Analytics, School of Medicine, University of Leeds, Leeds, UK

^c GeoHealth Laboratory, Geospatial Research Institute, University of Canterbury, Canterbury, New Zealand

ARTICLE INFO

Keywords:

Community nutrition environment
 Foodscape
 GIS
 Food access
 Obesogenic environment
 Methodology

ABSTRACT

This systematic review quantifies methods used to measure the 'retail food environment' (RFE), appraises the quality of methodological reporting, and examines associations with obesity, accounting for differences in methods. Only spatial measures of the RFE, such as food outlet proximity were included. Across the 113 included studies, methods for measuring the RFE were extremely diverse, yet reporting of methods was poor (average reporting quality score: 58.6%). Null associations dominated across all measurement methods, comprising 76.0% of 1937 associations in total. Outcomes varied across measurement methods (e.g. narrow definitions of 'supermarket': 20.7% negative associations vs 1.7% positive; broad definitions of 'supermarket': 9.0% negative associations vs 10.4% positive). Researchers should report methods more clearly, and should articulate findings in the context of the measurement methods employed.

1. Introduction

The idea that the retail food environment (RFE) is a cause of obesity is intuitively appealing. The RFE comprises the spatial availability, accessibility and composition of food outlets within local environments; sometimes referred to as the 'community nutrition environment' (Glanz et al., 2005). Over the past half-century, the RFE has changed drastically. Since the 1970s many western countries have seen a shift in grocery retailing with large superstores establishing in suburban and out of town regions, leading smaller, local high-street grocers to close (White, 2007; Wrigley, 2002; Guy and David, 2004; Walker et al., 2010). This has purportedly led to the development of so-called 'food deserts', where residents lack access to healthy food. Evidence supporting existence of food deserts is particularly strong within the US (Sparks et al., 2011; Dawson et al., 2008; Black et al., 2014). The US and UK have also seen a proliferation of restaurants and fast food outlets, providing low-cost, energy-dense foods (Maguire et al., 2015; Burgoine et al., 2009; Jeffery et al., 2006). These changes have coincided with increases in obesity rates, which have been rising globally since the 1970–80s (World Obesity Federation, 2017). In the UK (Public Health England, 2014; Local Government Association, 2016; Greater

London Authority, 2012), and internationally (Nykiforuk et al., 2018; Diller and Graff, 2011) policymakers have been investigating ways to intervene to create healthier RFEs, for example through banning the opening of fast food outlets around schools (Waltham Forest Spatial Planning Unit, 2009).

Despite considerable research activity, evidence supporting a link between the RFE and obesity is conflicting at best. The largest systematic review to date on the association between the RFE and weight status (Cobb et al., 2015) in the US and Canada found for example that while there were 31 statistically significant positive associations between fast food outlets and increased weight status, 99 associations were null, and 4 were negative (showing increased number or proximity of fast food outlets was associated with decreased weight status). Similarly, supermarkets (often considered as a proxy for healthy food availability) were found to be statistically significantly positively associated with lower weight status in only 24 of 143 associations, with 7 associations in the unexpected direction (showing increased number or proximity of supermarket access was associated with higher weight status). Other systematic reviews have reported predominantly null findings in relation to RFE-obesity associations (Feng et al., 2010; Williams et al., 2014; Gamba et al., 2015; Casey et al., 2014). Although

* Corresponding author. Leeds Beckett University, Headingley Campus, Headingley, Leeds, LS6 3QS, UK.

E-mail address: e.wilkins@leedsbeckett.ac (E. Wilkins).

<https://doi.org/10.1016/j.healthplace.2019.02.007>

Received 8 October 2018; Received in revised form 24 January 2019; Accepted 26 February 2019

1353-8292/© 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

these reviews focus only on p-values, and do not account for the magnitude of associations, they do highlight numerous conflicting results and tend to suggest little or no consistent link between the RFE and obesity.

A common challenge in understanding RFE-obesity associations - repeatedly noted by authors - is the diversity of methods used to measure the RFE (Cobb et al., 2015; Feng et al., 2010; Williams et al., 2014; Gamba et al., 2015; Casey et al., 2014). The majority of literature uses spatial measures, such as the density or proximity of outlets, to characterise the RFE (Caspi et al., 2012; Lytle and Sokol, 2017). A recent review identified five dimensions of methodological diversity with regard to spatial RFE measures: (i) source of food environment data, (ii) methods used to extract food outlets from a wider dataset, (iii) methods for classifying outlets, (iv) geocoding methods, and (v) choice of RFE metric or measure (Wilkins et al., 2017). These are summarised in the GeoFERN reporting framework: a reporting checklist developed specifically for RFE research covering the five dimensions. A number of reviews have quantified methods used in RFE literature across aspects of these domains (Cobb et al., 2015; Williams et al., 2014; Gamba et al., 2015; Charreire et al., 2010). However, no study has systematically and comprehensively quantified the degree of methodological diversity across all five domains. Quantification of the methods used is important to (i) identify priority areas for future research into the impact of methods, and (ii) highlight the scale of the problem in order to encourage researchers to move towards more standardised or evidence-based methods where possible. One aim of this review was therefore to:

- (1) **Conduct a systematic review to comprehensively identify the spatial methods used to measure the RFE across the five GeoFERN dimensions and quantify their frequency of use.**

Previous work (Wilkins et al., 2017) has also highlighted that methodological information is often not reported in papers. However, no study has ever quantified the quality of methodological reporting, and thus little is known about the scale of the problem of incomplete reporting. A second aim was therefore to:

- (2) **Quantify the quality of methodological reporting within the spatial RFE literature.**

Given the varied approaches to measuring the RFE, it is perhaps not surprising that evidence is conflicting. Recent research suggests that methods used to measure the RFE may impact RFE-obesity relationships. For example, while many studies operationalise the RFE in terms of singular food outlet types, such as 'fast food outlets' (Cobb et al., 2015; Williams et al., 2014; Gamba et al., 2015), research suggests that relative measures of food outlet mix (e.g. the ratio of fast food outlets to supermarkets) may be more strongly and consistently associated with obesity-related outcomes (Clary et al., 2015, 2016; Polsky et al., 2016; Feng et al., 2018). Other methodological factors such as choice of food environment data (Mendez et al., 2016; Hobbs et al., 2016) and geocoding methods (Thornton et al., 2012) are also beginning to be investigated.

Associations between the RFE and obesity may additionally vary across population groups. Stronger associations between the RFE and obesity-related outcomes have been found within more deprived neighbourhoods (Bernsdorf et al., 2017; Fiechtner et al., 2015; Thomsen et al., 2016). Differential associations have also been observed for people of differing income and education (Burgoine et al., 2016; Reitzel et al., 2014), ethnicity (Wong et al., 2017), age (Dwicaksono et al., 2017) and across urban/rural residences (Bernsdorf et al., 2017).

Existing systematic reviews either do not account for potential divergent effects across measurement methods or population groups (Williams et al., 2014; Casey et al., 2014), or account only for a limited range of factors using simplistic groupings of studies; for example grouping diverse methods together (Cobb et al., 2015; Feng et al., 2010;

Gamba et al., 2015; Caspi et al., 2012). These reviews may therefore miss important associations at the level of population groups or measurement methods. A final aim of this systematic review was therefore to:

- (3) **Examine the evidence for associations between the spatially measured RFE and obesity, accounting for possible divergent associations across measurement methodologies.**

2. Methodology

2.1. Data sources and search strategy

This review capitalises on work carried out by Cobb et al. (2015) by updating and expanding upon their existing systematic review into associations between the RFE and obesity. Papers identified by Cobb et al. (2015) ('the Cobb review') are included in the present review (subject to exclusion criteria), and were re-reviewed to extract new information as outlined below. Additionally, the Cobb search was re-run to identify latterly published literature, including other western countries (signatories to the Organisation for Economic Co-operation and Development convention) in addition to the US and Canada. Non-western countries were excluded from the present review due to differences in food environment and obesity dynamics (Popkin et al., 2012).

Mirroring the Cobb review, searches were performed using Scopus and PubMed for English-language literature published online or in print relating to the association between the RFE and obesity, using search terms alluding to weight status, such as 'overweight', 'obese' and 'body mass index' and to the RFE, such as 'food environment', 'food access' and 'fast food' (Supplement 1). We sought to identify literature published from the 1st January 2014 (to align with the end of the Cobb review) up to the 8th June 2017. To capture literature that was published but not indexed before 1st January 2014, we searched by 'date created' on PubMed, and allowed a 1-year time-lag for Scopus; thus including in our Scopus search all literature published since 1st January 2013.

2.2. Exclusion criteria

All exclusion criteria replicated those of the Cobb review unless otherwise indicated. More particularly, in accordance with the Cobb review, studies in our review were required to examine associations between objective spatial measures of the RFE around the home and individual-level outcomes of either BMI, weight classification (e.g. 'obesity'), BMI change or weight change (referred to collectively as measures of 'obesity'). Replicating Cobb, our review focussed on home environments, which are the most commonly investigated environments (Gamba et al., 2015). Further following Cobb, studies in the present review were excluded if they (i) examined associations with area-level outcomes only (e.g. obesity prevalence), (ii) treated the RFE as a moderator, mediator or covariate only, (iii) used simulated data, (iv) combined RFE measures with other environmental measures (e.g. access to physical activity facilities), such that the effects of the RFE could not be isolated or (v) if they were case studies investigating the influence of one or more specific outlets, such as a newly opened supermarket or a store that a participant has visited, without measuring the wider RFE. In line with the Cobb review, studies in the present review were additionally required to (i) include at least 200 people, (ii) operationalise the RFE using areal units smaller than or equal to US zip code zones, and (iii) examine spatial measures of at least one of: supermarkets, grocery stores, convenience stores, fast food restaurants, full-service restaurants, composite measures including at least one of these outlet types, or food availability within at least one of these outlet types. One objective of the present review was to evaluate the reporting of methods, and thus, as a departure from the Cobb strategy, papers

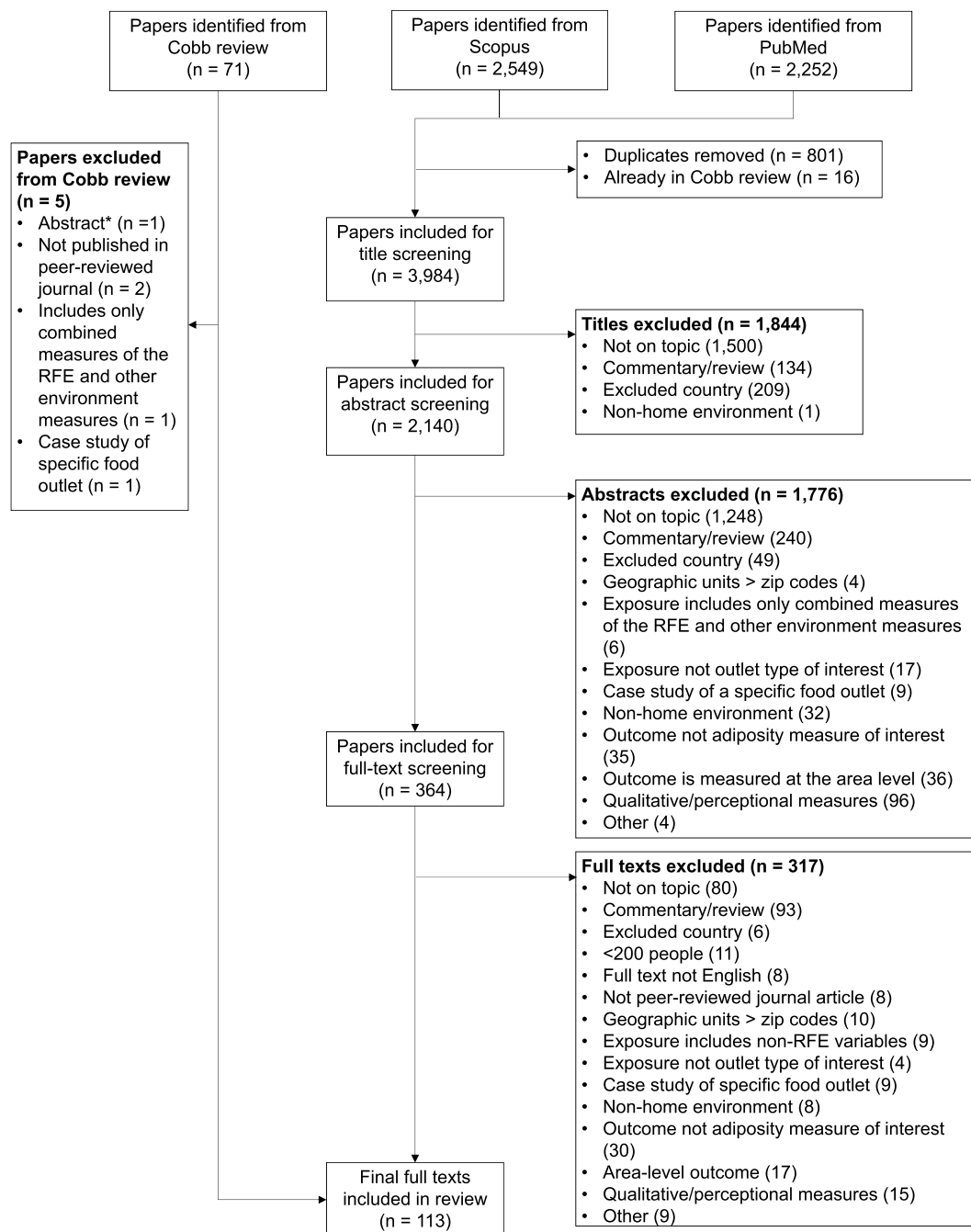


Fig. 1. Flow chart illustrating screening process for this review. RFE = Retail Food Environment. Note, for the papers excluded from the Cobb review, the third and fourth criteria listed above were also applied in the original Cobb review, but appeared to have been incorrectly applied in respect of two papers. *Article was an abstract corresponding to a full-text paper identified in the top-up search and thus was excluded to avoid duplication.

were required to report primary research within peer-reviewed journals.

2.3. Screening

The top-up search returned 4,801 results, which were exported to Endnote for deduplication. Of the remaining results ($n = 3,984$), 1,844 articles were excluded after title screening; 1,776 after abstract screening and 317 after full-text screening. Five studies from the Cobb review were excluded in line with our exclusion criteria (Fig. 1). Overall 113 papers were included in the review.

All articles were screened by the primary reviewer (EW). Double-screening was undertaken for a sample of articles (2015 titles; 1276

abstracts; 70 full-texts) by one of six reviewers (CG, DR, MM, MH, WM or AM). Conservatively, reviewers excluded articles at the title stage only if they very clearly were off-topic or met an exclusion criterion and titles were retained if at least one reviewer determined not to exclude them. Disagreements at the abstract and full-text stage were resolved by a third independent reviewer. After full-text screening, agreement with the primary reviewer's decision was 98.6%, with one paper excluded by the primary reviewer being retained.

2.4. Data extraction

Data were extracted from both the newly identified studies and the original Cobb studies on study design, RFE measurement methods,

study quality, and numbers of null and statistically significant associations, and the directions of statistically significant associations. Data extraction was considerably more extensive than in the Cobb review, totalling 99 data fields (Supplement 2). Methodological information was extracted for each of the reporting items deemed 'essential' in the 'GeoFERN' checklist (Wilkins et al., 2017). Effect sizes were not extracted due to the varied methods and measures used, making collation of these data at the scale of the present review impossible. This approach of counting null and significant associations has been employed by other systematic reviews when faced with similarly methodologically diverse data (Cobb et al., 2015; Williams et al., 2014; Sallis et al., 2000). All data were extracted into Microsoft Excel.

Often papers report associations for multiple statistical models, or repeat analyses for different population groups or exposure measures. Outcomes data (numbers of statistically significant/null results) were extracted for each distinct exposure measure, outcome (e.g. BMI or 'obesity') and population subgroup, including results for the full sample, if reported. This is because these different models represent different research questions. Where multiple models were run for the same exposure-outcome-population grouping (e.g. using different covariates), results were only extracted for the 'main model' (Supplement 3). For most studies, the main model was taken to be the most fully-adjusted model presented in a results table.

Due to the scale of the review, only the aims and objectives, methods and results sections of papers were reviewed, except where explicit reference was made to methodological details provided in supplementary materials or other published papers. Authors were not contacted to obtain missing information due to the high prevalence of missing information.

Data extraction was performed by the primary reviewer. Two second reviewers (MH, AC) independently extracted data from a random 20% sample ($n = 23$), with disagreements being resolved through discussion. Overall, 96.5% of data fields ($n = 2,427$) were in agreement with the first reviewer's initial decision (Supplement 4).

2.5. Quality screening

Studies were appraised for risk of bias using an expanded version of the Cobb review quality checklist, adapted from the Newcastle Ottawa Scale (Wells et al.,). A total of 10 marks were available for features such as validation of food environment data, use of a causal framework, use of multi-level modelling or equivalent methods accounting for clustering within neighbourhoods (where relevant) and controlling for key covariates (age, race, gender and neighbourhood socioeconomic status/racial composition) (Supplement 5). Quality scores were expressed as a percentage of eligible marks, with higher scores indicating lower risk of bias. Additionally, the completeness of methodological reporting was appraised using the GeoFERN reporting checklist (Wilkins et al., 2017). For each paper, one mark was awarded for each 'essential' reporting item on the GeoFERN checklist, with half-marks being awarded when reporting criteria were partially met. An overall GeoFERN reporting score for each paper was calculated as the percentage of eligible marks obtained. After double-screening a 20% sample of papers ($n = 23$), agreement between the final decision and the first reviewer's initial decision was 96.3% for the GeoFERN scores and 97.0% for the study quality tool.

2.6. Data synthesis

We reported the frequency of use of different RFE measurement methods across the five GeoFERN domains (data source, data extraction, food outlet classifications, geocoding methods and RFE metrics) and the prevalence of missing methodological information within each domain. The numbers of null and statistically significant positive/negative associations were reported for 112 studies; one study (Li et al., 2009) was excluded from this aspect of the analyses because it did not

report the main effects of the RFE. For the four main exposures of 'fast food outlets', 'supermarkets/grocery stores', 'convenience stores' and 'restaurants', results were stratified by population groups, and for the two most common exposures ('fast food outlets' and 'supermarkets/grocery stores'), results were stratified by measurement method. We additionally evaluated the numbers of null and statistically significant positive/negative associations for studies within the top decile of quality score only ($\geq 66.7\%$), to determine whether our findings were sensitive to study quality. Data were presented for populations and methods used in 5 or more studies to enable generalised comparisons between methods. Further information on the coding of data is available in Supplement 6.

3. Results

3.1. Study characteristics

There were 113 papers included in this review, published between 2004 and 2017 (Supplement 7), comprising 107 unique datasets. Sixty-six were identified from the original Cobb review, with the remaining 47 newly identified. The median participant sample size was 3,786 (range: 219 to 453,927). Twelve studies derived outcome data from a dataset that was also used in another study (6 unique datasets). Due to the large number of studies included in this review, only summary data are provided in the main text. However, Supplements 8 and 9 respectively provide detailed information on study characteristics and findings at the level of individual papers.

Descriptive statistics of the studies are presented in Table 1. Overall, studies predominantly related to the RFE in the US (82.3%), examined populations of mixed gender (88.5%), who were adults (66.4%), of

Table 1
General study characteristics.

	Full Sample		Cobb Papers		New papers	
	N	%	N	%	N	%
Country						
US	93	82.3	59	89.4	34	72.3
Canada	10	8.8	7	10.6	3	6.4
UK	5	4.4	–	–	5	10.6
Australia	3	2.7	–	–	3	6.4
Germany	1	0.9	–	–	1	2.1
New Zealand	1	0.9	–	–	1	2.1
Gender						
Mixed	100	88.5	59	89.4	41	87.2
Females only	13	11.5	7	10.6	6	12.8
Age						
Mixed	1	1.8	1	1.5	1	2.1
Adults	75	66.4	45	68.2	31	63.8
Children	36	31.9	20	30.3	16	34.0
Race						
Mixed	73	64.5	43	65.2	30	63.8
Mostly ethnic minority	10	8.8	4	6.1	6	12.8
Mostly white	10	8.8	7	10.6	3	6.4
Not reported	20	17.7	12	18.2	8	17.0
SES						
Mixed	71	62.8	43	65.2	28	59.6
High	2	1.8	1	1.5	1	2.1
Low	19	16.8	11	16.7	8	17.0
Not reported	21	18.6	11	16.7	10	21.3
Urban/rural						
Mixed	33	29.2	20	30.3	13	17.7
Urban	30	26.5	1	1.5	12	25.5
Rural	3	2.7	18	27.3	2	4.3
Not reported	47	41.6	27	40.9	20	42.6
Design						
Cross-sectional	87	73.5	56	84.8	31	66.0
Longitudinal	26	23.0	10	15.2	16	34.0

N = number of studies; SES = socioeconomic status; New papers = papers published after the Cobb review.

mixed races (64.5%) and mixed socioeconomic status (SES) (62.8%). Of those studies reporting environmental context, the vast majority were either mixed urbanity or entirely urban (95.5%). Nearly three quarters of the studies were cross-sectional. Contrasting the newly-identified papers to the older papers from the Cobb review, there was a higher proportion of longitudinal studies (34.0% vs 15.2%), studies relating to ethnic minority populations (12.8% vs 6.1%) and studies in predominantly urban areas (25.5% vs 1.5%). Despite the wider geographic scope of the updated review, a high proportion of studies originated from the US (72.3%).

3.2. Study quality

Study quality scores ranged from 22.2% (indicating a high risk of bias) to 88.9% (indicating a lower risk of bias), with a mean of 49.9%. There were 18 studies (15.9%) with a score $\geq 66.7\%$, corresponding to the top decile of quality scores. The most common risks of bias were failure to use a causal framework to guide model development (97.3% of studies), failure to control for neighbourhood self-selection (85.8% of studies), and use of secondary food environment data without validation of the data (78.8% of studies). Further data on study quality is presented at Supplement 5, Table S5.

3.3. Methods used in the research

Fig. 2 displays the frequency of use of different methodologies across studies. Restricting analyses to only those identified in the top-up search did not lead to substantively different findings.

3.3.1. Data source

The vast majority of studies (83.2%) obtained RFE data from a single source, with the remainder combining multiple sources. Commercial data (for business marketing or market research purposes) and government data were the most common data sources (Fig. 2). Commercial data were typically from InfoUSA (InfoGroup, Inc.) (36.2%) or Dun & Bradstreet, Inc. (34.1%) (including the National Establishment Time Series dataset, which is derived from Dun & Bradstreet). Government data were typically from local health, hygiene or licensing departments (71.4%).

3.3.2. Data extraction

Once RFE data have been obtained, it is often necessary to extract specific food outlets of interest from a wider dataset. Data were predominantly extracted using information within the RFE data, which included proprietary classifications, store names, or other attributes, such as store size or revenue. Some studies used secondary data sources such as business directories and websites (Fig. 2). The majority of studies (73.8%) used a single method (e.g. proprietary classifications only), with the remainder using a combination of methods (e.g. proprietary classifications and store names).

3.3.3. Food outlet constructs

Studies typically employed 'fast food outlets' constructs (sometimes referred to as 'takeaways' or 'limited service restaurants'), 'supermarkets' and/or 'grocery stores' (hereinafter 'supermarkets/grocery stores'), 'convenience stores' (including 'bodegas'), and 'full-service' or 'sit-down' restaurants' (hereinafter 'restaurants') (Fig. 2). These outlet constructs were either measured in isolation (for example as the density or proximity of 'fast food outlets') or as part of a composite variable, such as the ratio of 'fast food outlets' to 'restaurants'. Forty studies (35.4%) used other food outlet constructs, such as 'food stores' or 'total restaurants', which encompassed, but did not directly define the four main outlet types. Supermarkets and grocery stores were grouped under one category because studies defined these constructs inconsistently. For example, some studies would use the term 'grocery store' to refer to both large chain supermarkets as well as smaller local grocery stores,

whereas other studies would treat large chain 'supermarkets' and smaller 'grocery stores' as distinct constructs.

Constructs were defined using four main methods: (i) use of proprietary classifications within the RFE data, (ii) use of other attributes within the RFE data, such as store name or size, (iii) a combination of proprietary classifications and other attributes within the RFE data, and (iii) telephone or in-person audits (Fig. 2). Other methods included use of supplementary information, such as internet searching.

The scope of commonly employed food outlet constructs also varied (Fig. 2). For example, 35.2% of studies defined 'supermarkets' narrowly to include only large chain supermarkets, 40.7% employed a moderate scope including large/mid-sized grocery stores, and 24.1% included small grocery stores (see Supplement 6 for further details). While several studies cited use of standardised classification schemes (NAICS classification scheme: 23.9% of studies; Standard Industry Classification (SIC) scheme: 13.3% of studies), these were not necessarily employed in the same way. For example, while some studies used the NAICS code 722513 for 'limited service restaurants' to define 'fast food outlets' (Zhao et al., 2014; Gibson, 2011; An and Sturm, 2012; Lopez, 2007); others additionally included cafeterias (NAICS code 722212) and mobile food services (NAICS code 722330) (Chen and Wang, 2016) or pizza restaurants (NAICS code 722211) (Shier et al., 2012).

3.3.4. Geocoding

Geocoding is the process of converting address information into coordinates or other geographic identifiers through matching of address information to spatially coded reference data. Home addresses were most commonly geocoded to geographic identifiers at the level of census tracts, postcode zones or street segments (Fig. 2), with the latter method typically using building numbers to estimate how far along the street an address is located. Less commonly, addresses were geocoded to the building level, zip code level, census blocks or land/tax parcels. Similar methods were used for food outlets (data not presented due to small number of studies ($n = 15$) reporting this information).

3.3.5. RFE metrics

The metrics used to measure the RFE predominantly included: (i) buffer metrics assessing the RFE within a given distance of the home, (ii) areal metrics assessing the RFE within a predefined areal unit such as a census tract or zip code zone, (iii) proximity metrics, which measure the distance between the home and one or more food outlets, and (iv) gravity metrics, which effectively combine proximity and buffer metrics by measuring the count or density of food outlets within a buffer, with outlets that are more proximal to the home being weighted higher (Fig. 2).

Within these broad types of metric, specific measures were highly diverse. This was particularly true for areal and buffer metrics, as enumerated in Table 2. Areal and buffer measures were used 242 times across the 113 papers. Of the metrics that had a clearly defined unit of measurement, the most common was the count of outlets per unit area, which included counts of outlets within Euclidian (straight-line) buffers (31.2% of measures), followed by raw, non-standardised counts of outlets (19.4% of measures) and presence/absence of an outlet type (14.8% of measures).

The geographic scope of area-based measures also varied. For areal metrics (58 measures), studies most commonly used census tracts to define the scope of the RFE (53.6% of areal measures). Buffer metrics were typically delineated in terms of Euclidian distances (83 measures, 52.2% of buffer metrics) or network distances (50 measures, 31.4% of buffer measures), with 27 measures (17.0%) of undefined delineation. The scope of buffers varied, but were generally between 400 and 1600 m for both network and Euclidian buffers (59.7% of buffer measures). Nearly half of all studies that employed buffer metrics (46.8% of 62 studies) measured the same RFE metric using more than one buffer size. Seven studies included 2 buffer sizes, 10 investigated 3 buffer sizes, and 12 investigated 4 or more buffer sizes.

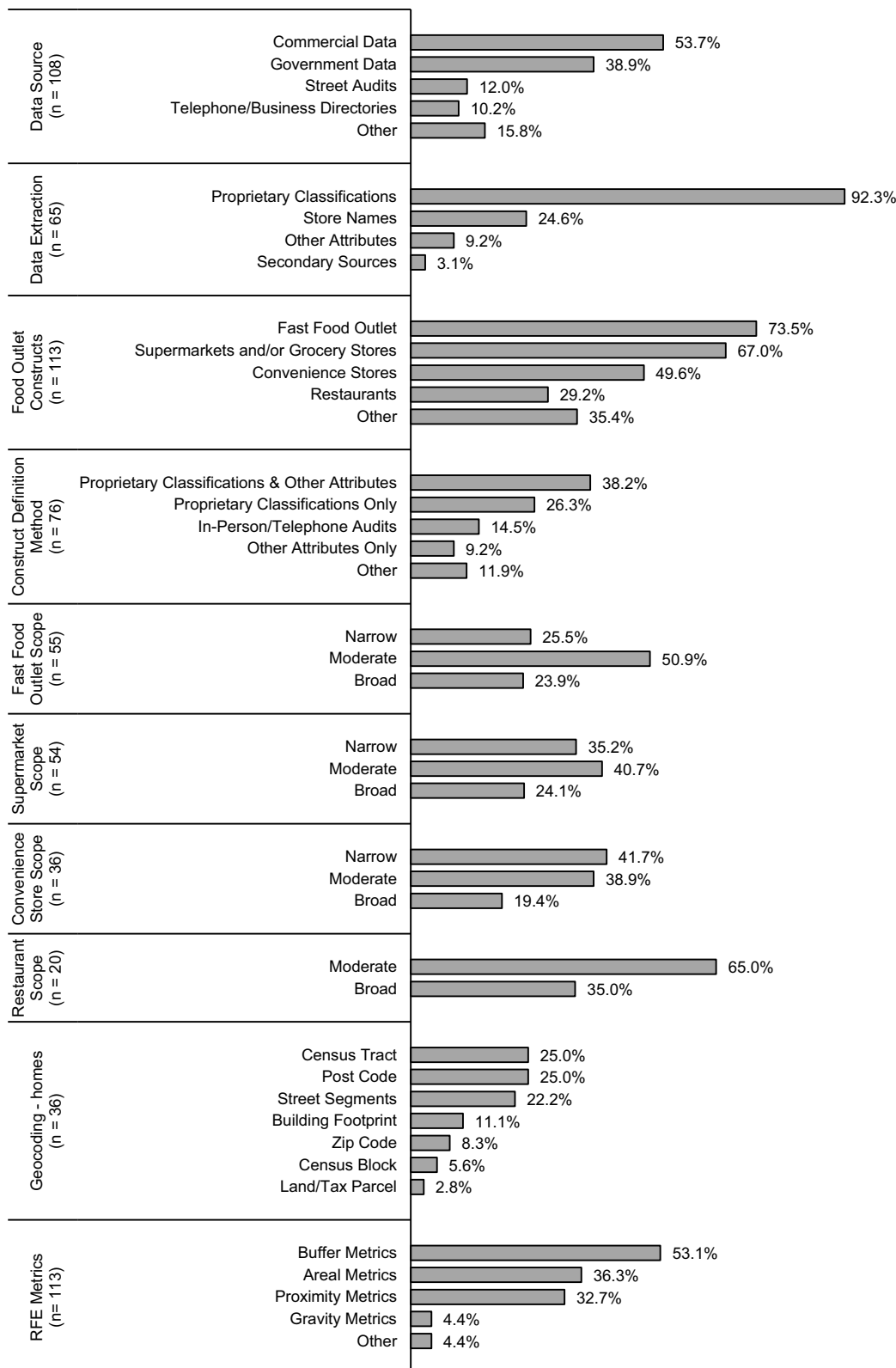


Fig. 2. Methods used within studies. n = number of studies for which methodological aspect is applicable and for which sufficient methodological information was provided. RFE = Retail Food Environment. Some studies used more than one method within a given methodological aspect, and thus percentages shown do not always add up to 100%. ‘Other’ data sources included internet searching, data from national mapping agencies, and satellite imagery. ‘Other’ food outlet constructs included various composite measures such as supermarkets and greengrocers combined, or fast food outlets and convenience stores combined. ‘Other attributes’ used for construct definition was limited to information contained within the RFE dataset and included outlet name, size, number of employees or tills. ‘Other’ methods for applying outlet constructs included use of supplementary information e.g. websites, and interviews with local residents. ‘Other’ RFE metrics included e.g. a binary measure of whether the neighbourhood centroid was closer to a supermarket or ethnic market and measures of relative store ‘attractiveness’ (accounting for distance and store size).

Proximity measures were used 36 times across the 113 studies. These metrics were also variable, but to a lesser extent than for area-based metrics, with the vast majority (88.9%) measuring the distance to the nearest outlet of a given type e.g. ‘supermarket’. Alternative proximity measures included the average distance to the nearest ‘N’ outlets of a given type (5.6% of measures), and the relative proximity of two or more outlet types, such as the distance to the nearest healthy

outlet minus the distance to the nearest unhealthy outlet (5.6% of measures). Proximity was most commonly measured as the network distance (93.5% of measures), with Euclidian distance and travel time being used an equal number of times (16.1% of measures respectively).

Gravity metrics were also varied. Of the five studies that used gravity measures (4.4%), four of these used a fixed circular bandwidth, which ranged from 1 km to 6 miles and one used an adaptive

Table 2
Counts of area-based measures of the RFE.

Geographic Scope	Unit of measurement									TOTAL
	Count/area	Count/population	Non-standardised count	Presence vs absence	Relative	Audit score	Variety	Other	Unclear	
Areal (N = 41)										
Census block group	3		1					1		5
Census tract	4	5	8	4	3	1	2	3		30
Zip code	1	3	1					2		7
Researcher defined	4	3	1	1	2		1			12
Other - administrative	1							1	1	2
Buffer – Euclidian (N = 23)										
< =200 m	3 ^a	1	n/a	1						5
> 200–400 m	5 ^a	1	n/a	1						7
> 400–800 m	14 ^a	2	n/a	2	1					19
> 800–1600 m	15 ^a	2	n/a	4	3	1	1	1	3	30
> 1600–2400 m	5 ^a	1	n/a							6
> 2400–3200 m	5 ^a	1	n/a							6
> 3200 m	5 ^a	1	n/a	1				1	1	9
Buffer – Network (N = 27)										
< =200 m										0
> 200–400 m	1	1	2	3						5
> 400–800 m	3		8	3	2	3		1		12
> 800–1600 m	2	1	4	5	4	4	1	2		19
> 1600–2400 m			3			2				2
> 2400–3200 m	1		3	2	1			1		5
> 3200 m			5	3		3		1		7
Buffer – Undefined (N = 8)										
> 200–400 m	1		1	1				1		4
> 400–800 m			1							1
> 800–1600 m			2	1	1			2		5
> 1600–2400 m	1		4	2	1			2		10
> 2400–3200 m			1		1					2
> 3200 m			2	1	1			2		5
TOTAL	74	22	46	35	20	14	5	21	5	

Note. Many studies employed multiple measures of the RFE, and thus the total number of measures (242) exceeds the number of studies (113). N = number of studies employing each broad method. Non-standardised count = measures of the raw counts of outlets that are not standardised e.g. to a given area or population. Relative = measures of the availability of one outlet type relative to one or more other outlet types. Audit score = measures derived from within-store audits e.g. the total shelf space devoted to fruits and vegetables within a buffer. Variety = measures of the number of different outlet types. Other = other measures of the RFE, including counts of outlets relative to the length of roads within a buffer, weighted counts of outlets and counts per area per population. Buffer – Undefined = studies that described using a buffer measure of the RFE, but did not describe whether this was a network or Euclidian buffer.

^a Measures of the raw count of outlets within Euclidian buffers were classified as count/area, because Euclidian buffers of a given radius have a fixed area.

bandwidth, but it was unclear what the adaptive radius was based on. Two studies used a quartic decay function (defining how quickly the weighting of food outlets falls off with increasing distance), one used a quadratic, one used a Gaussian decay function and one did not report the decay function used.

3.4. Quality of methodological reporting

Overall, the mean GeoFERN reporting quality score was 58.6% (range: 25.0%–97.2%). Table 3 shows the completeness of methodological reporting across the five GeoFERN domains. Methodological reporting was worst for the geocoding domain, with only 3 papers (2.7%) providing full information on the geocoding methods used and an average score of 41.2%. It was commonly unclear whether geocoding was used and/or how this was performed, with this information being omitted in relation to the geocoding of food outlets and homes in 76.1% and 58.4% of studies respectively.

3.5. RFE-obesity associations

Overall, there were 1,937 reported associations between the RFE and obesity. Null associations predominated, making up 76.0% of all associations. Table 4 enumerates the associations between the most common measures of the RFE (fast food outlets, convenience stores, supermarkets/grocery stores, and restaurants) and obesity, including sub-groups of age, gender, ethnicity, and urban/rural status. Throughout, ‘positive associations’ refer to statistically significant

associations indicating increased access/exposure to food outlets is associated with *increased* obesity, and ‘negative associations’ refer to statistically significant associations indicating increased access/exposure to food outlets is associated with *decreased* obesity.

The distribution of associations varied across population groups. For example, there was a stronger tendency toward more positive associations than negative associations for fast food outlets among low-SES children (39.3% positive, 3.6% negative, 57.1% null) than for the general population (20.8% positive, 4.2% negative, 75.0% null). Additionally, there was no trend towards positive/negative associations for convenience stores among the general population (10.9% positive, 9.5% negative, 79.6% null). However, after stratifying by age, convenience stores tended to be more consistently associated with higher rather than lower obesity among children (16.5% positive, 2.6% negative, 80.9% null); particularly those of low-SES and non-white ethnicity (e.g. 39.3% positive, 3.6% negative, 57.1% null for low-SES). Restricting to high-quality studies did not substantively change findings.

Table 5 shows the distribution of positive, negative and null associations after stratification by RFE measurement method. Results by geocoding method are not presented, because geocoding methods were rarely reported. The distribution of positive and negative associations differed across RFE measurement methods. In particular, there was very little evidence supporting an association between supermarkets and obesity when considering all definitions of ‘supermarket’ collectively (6.6% positive, 12.9% negative, 80.5% null). However, when considering only narrow definitions of ‘supermarket’ (i.e. only large chain

Table 3
GeoFERN reporting quality by domain.

GeoFERN domains & associated marking criteria	Full details N (%)	Partial details N (%)	No details N (%)	Not applicable
Domain 1: Data Source				
Overall	55 (48.7%)	47 (41.6%)	11 (9.7%)	–
Name of data creator reported?	79 (76.0%)	5 (4.8%)	20 (19.2%)	9 ^a
Dataset name reported? ^b	42 (58.3%)	6 (8.3%)	24 (33.3%)	4 ^a
Publication date reported ^b	71 (62.8%)	3 (2.7%)	39 (34.5%)	0
Domain 2: Extraction Methods				
Overall	28 (27.5%)	31 (30.4%)	43 (42.1%)	11 ^d
Extraction Method Reported ^c	54 (52.9%)	11 (10.8%)	37 (36.3%)	11 ^d
Search Terms Reported	53 (52.0%)	17 (16.7%)	32 (31.4%)	11 ^d
Domain 3: Construct Definitions^e				
Overall	18 (15.9)	95 (84.1)	0	–
Construct names listed	112 (99.1)	1 (0.9)	0	–
Construct scope clear	60 (53.1)	26 (23.0)	27 (23.9)	–
Classification method clear	51 (45.1)	25 (22.1)	37 (32.7)	–
Examples provided	37 (32.7)	13 (11.5)	63 (55.8)	–
Domain 4: Geocoding Methods				
Overall	3 (2.7)	103 (91.2)	7 (6.2)	–
Clear whether geocoding used?	61 (54.0)	32 (28.3)	20 (17.7)	–
Address model reported?	15 (13.3)	28 (24.8)	63 (55.8)	7 ^f
Match rate reported	83 (73.5)	–	23 (20.4)	7 ^f
Software reported	48 (42.5)	1 (0.9)	57 (50.4)	7 ^f
Urban/rural reported	22 (19.5)	44 (38.9)	47 (41.6)	–
Domain 5: Rfe Metrics				
Overall	22 (19.5)	91 (80.5)	0	–
Conceptual environment defined	110 (97.3)	–	3 (2.7)	–
Areal (N = 41)				
Type of zone defined?	40 (97.6)	–	1 (2.4)	–
Boundary data reported?	14 (37.8)	–	23 (62.2)	4 ^g
Intensity metric reported?	38 (92.7)	–	3 (7.3)	–
Buffer (N = 60)				
Buffer type defined?	52 (86.7)	–	8 (13.3)	–
Buffer size defined?	60 (100.0)	–	0	–
Intensity metric defined?	59 (98.3)	–	1 (1.7)	–
Proximity (N = 37)				
Proximity type defined?	32 (86.5)	–	5 (13.5)	–
Network (N = 46)				
Types of roads/paths described	43 (93.5)	–	3 (6.5)	–
Network data cited?	7 (15.2)	–	39 (84.8)	–
Gravity (N = 5)				
Radius defined?	4 (80.0)	–	1 (20.0)	–
Decay function defined?	4 (80.0)	–	1 (20.0)	–
Other (N = 5)				
Metric clearly described?	5 (100.0)	–	0	–

Table shows the number (N) of studies that reported full, partial or none of the details listed as essential within the GeoFERN reporting checklist (the ‘marking criteria’) for each GeoFERN domain (Casey et al., 2014) (‘Overall’ score). Also shown is a break-down of the reporting quality (i.e. numbers of studies providing full, partial or no details) for each specific marking criterion. Percentages shown are the percentages of eligible studies for each marking criterion; excluding those for which GeoFERN marking criterion was not applicable.

RFE = Retail Food Environment.

^a Not applicable because food environment data was collected through street audits, or because dataset does not have a name.

^b The ‘data creator’ is the person or entity that created the RFE data e.g. ‘Dun & Bradstreet’. The publication date is either the date on which RFE data were collected, or the date on which the data were published.

^c Extraction methods refer to the methods used to extract food outlets of interest from a wider dataset which may contain food outlets not of interest and/or non-food businesses. Often search terms are used, which might include outlet names or classifications within the dataset.

^d Not applicable because data was collected via street audits, or no data extraction was required.

^e Outlet ‘constructs’ include e.g. ‘fast food outlets’, ‘supermarkets’ and ‘convenience stores’.

^f Not applicable because geocoding was not required.

^g Boundary data was not applicable for 4 studies; wherein areal units were researcher-defined.

Table 4
Numbers of statistically significant positive, negative and null associations between RFE and weight status.

Exposure-population grouping	Positive*		Null		Negative*		Total N	No. of studies
	N	%	N	%	N	%		
Fast Food Outlets								
Full sample	84	20.8%	303	75.0%	17	4.2%	404	74
High quality studies only	12	30.8%	26	66.7%	1	2.6%	39	8
Adults	52	22.1%	173	73.6%	10	4.3%	235	44
Males	3	9.4%	24	75.0%	5	15.6%	32	10
Females	13	25.5%	34	66.7%	4	7.8%	51	16
Low SES	4	22.2%	14	77.8%	0	0.0%	18	6
High SES	1	7.7%	12	92.3%	0	0.0%	13	5
White	2	9.1%	19	86.4%	1	4.5%	22	7
Urban	13	29.5%	31	70.5%	0	0.0%	44	13
Children	32	19.5%	126	76.8%	6	3.7%	164	28
Low SES	12	37.5%	19	59.4%	1	3.1%	32	10
Non-white	0	0.0%	8	88.9%	1	11.1%	9	5
Urban	2	4.7%	39	90.7%	2	4.7%	43	9
Convenience Stores								
Full sample	30	10.9%	218	79.6%	26	9.5%	274	52
High quality studies only	0	0.0%	47	95.9%	2	4.1%	49	7
Adults	11	6.9%	125	78.6%	23	14.5%	159	31
Males	1	3.4%	20	69.0%	8	27.6%	29	8
Females	7	12.7%	43	78.2%	5	9.1%	55	14
Low SES	1	5.0%	19	95.0%	0	0.0%	20	7
White	2	8.3%	19	79.2%	3	12.5%	24	5
Urban	2	5.7%	26	74.3%	7	20.0%	35	11
Rural	0	0.0%	18	100.0%	0	0.0%	18	5
Children	19	16.5%	93	80.9%	3	2.6%	115	21
Females	4	16.0%	21	84.0%	0	0.0%	25	5
Low SES	11	39.3%	16	57.1%	1	3.6%	28	8
Non-white	7	33.3%	12	57.1%	2	9.5%	21	6
Urban	2	10.5%	16	84.2%	1	5.3%	19	5
Supermarkets/Grocery Stores								
Full sample	37	6.6%	454	80.5%	73	12.9%	564	70
High quality studies only	5	4.0%	108	85.7%	13	10.3%	126	12
Adults	34	9.0%	293	77.7%	50	13.3%	377	46
Males	0	0.0%	36	83.7%	7	16.3%	43	9
Females	17	12.6%	102	75.6%	16	11.9%	135	20
Low SES	5	7.4%	60	88.2%	3	4.4%	68	12
White	7	12.3%	47	82.5%	3	5.3%	57	5
Urban	6	6.3%	74	77.9%	15	15.8%	95	16
Rural	7	15.9%	37	84.1%	0	0.0%	44	5
Children	3	1.6%	157	86.3%	22	12.1%	182	23
Low SES	0	0.0%	17	89.5%	2	10.5%	19	8
Urban	0	0.0%	24	92.3%	2	7.7%	26	8
RESTAURANTS								
Full sample	11	6.3%	131	75.3%	32	18.4%	174	29
High quality studies only	4	9.8%	34	82.9%	3	7.3%	41	5
Adults	1	1.0%	71	70.3%	29	28.7%	101	18
Males	0	0.0%	18	62.1%	11	37.9%	29	8
Females	0	0.0%	24	72.7%	9	27.3%	33	11
Urban	0	0.0%	16	76.2%	5	23.8%	21	5
Children	10	13.7%	60	82.2%	3	4.1%	73	11

N = number of associations. SES = socioeconomic status.

* 'Positive associations' refer to statistically significant ($p < 0.05$) associations indicating increased access/exposure to food outlets is associated with increased obesity, and 'negative associations' refer to statistically significant associations indicating increased access/exposure to food outlets is associated with decreased obesity.

outlets), there was a tendency for more negative than positive associations (1.7% positive, 20.7% negative, 77.6% null). Additionally, there was a tendency for more positive than negative associations for narrowly defined measures of fast food outlets (major chain outlets only) compared to broader definitions (e.g. 26.1% of associations vs 19.4%).

The distribution of associations additionally varied across RFE

metrics. For example, there was a tendency for more positive associations for measures of count/area, count/population and proximity of fast food outlets than for measures of presence/absence (e.g. proximity of fast food outlets: 28.6% positive, 2.6% negative, 68.8% null; presence/absence of fast food outlets: 0% positive, 3.8% negative, 96.2% null) and raw, non-standardised count (18.9% positive, 13.5% negative, 67.6% null). Measures of relative unhealthiness (such as the ratio of fast food outlets to total outlets) also tended notably towards more positive than negative associations (21.3% positive, 0% negative, 78.7% null). Additionally, there was a stronger tendency towards positive associations for fast food outlets among children for buffers ≤ 400 m (25% positive, 0% negative, 75% null) than for larger buffers (2.0% positive, 6.1% negative, 91.8% null). Finally, use of commercial data tended to be associated with a stronger patterning of associations in the expected directions for both fast food outlets and supermarkets.

4. Discussion

4.1. Methodological diversity and reporting

Existing systematic reviews into the RFE and obesity have repeatedly noted the diversity of methods used to measure the RFE (Cobb et al., 2015; Williams et al., 2014), often pointing to this diversity as limiting or even precluding conclusions that can be drawn from the evidence base (Feng et al., 2010; Gamba et al., 2015; Casey et al., 2014). However, no review has ever comprehensively quantified the diversity of methods across all aspects of methodological diversity, and thus the scale of this problem is unknown. This study extends the evidence base by systematically quantifying methods used across the five dimensions of methodological diversity: (i) the source of food environment data, (ii) the methods used to extract food outlets from a wider dataset, (iii) the methods and definitions used to classify outlets, (iv) geocoding methods and (v) RFE metrics, including all important methodological details rated as 'essential' in the GeoFERN framework (Wilkins et al., 2017). Understanding the methods used in the RFE literature will support emerging research into the comparability of different methods, by highlighting priority areas for further research. This review also quantifies for the first time the prevalence of missing methodological information relating to measurement of the RFE. Methodological information is critical to the interpretation of RFE-obesity studies, particularly given the mixed methods employed, and thus awareness of the extent of the issue will help motivate improved reporting moving forward.

A key finding of this review was that the degree of methodological diversity was extremely high. This finding is in agreement with the earlier Cobb review, who also found considerable methodological diversity in the literature. However, our review provides further information on the methods used across all dimensions of methodological diversity and across a wider selection of countries. In particular, our review provides new information on the methods used to extract food outlets from secondary datasets, apply food outlet classifications, and geocode food outlet and participant addresses. It also quantifies for the first time the variability in food outlet classification scopes and elucidates the true scale of diversity of areal measures of the RFE, which differ both in relation to their scope and unit of measurement. This diversity makes the collation and interpretation of research very challenging, as little is known about the comparability of different methods.

A second key finding was that RFE measurement methods are not well reported in the literature. Indeed, we found that not one single study provided all details rated as 'essential' within the GeoFERN reporting framework, and 33 studies (29.2%) provided less than half of these details. Overall, the high prevalence of missing methodological information, combined with the diversity of methods, severely limits the inferences that can currently be drawn from the evidence base. While policymakers should be praised for taking action against potentially obesogenic RFEs, inadequate methodological reporting

Table 5
Numbers of statistically significant positive, negative and null associations between RFE and weight status, stratified by measurement method.

Exposure-method grouping	Positive*		Null		Negative*		Total	No. of studies
	N	%	N	%	N	%		
Data Source								
Fast Food Outlets								
Government	13	16.0%	65	80.2%	3	3.7%	81	19
Commercial	59	24.5%	170	70.5%	12	5.0%	241	36
Supermarket/Grocery								
Government	16	8.8%	148	81.3%	18	9.9%	182	20
Commercial	5	1.8%	229	81.5%	47	16.7%	281	28
Extraction Methods								
Fast Food Outlets								
Proprietary classifications	54	26.3%	142	69.3%	9	4.4%	205	30
Combination of methods	8	16.3%	39	79.6%	2	4.1%	49	13
Supermarket/Grocery								
Proprietary classifications	16	4.8%	272	82.2%	43	13.0%	331	24
Combination of methods	5	6.3%	56	70.9%	18	22.8%	79	14
Outlet Classifications								
Fast Food Outlets								
All	84	20.8%	303	75.0%	17	4.2%	404	74
Narrow	24	26.1%	66	71.7%	2	2.2%	92	12
Moderate	29	22.3%	97	74.6%	4	3.1%	130	26
Broad	12	19.4%	48	77.4%	2	3.2%	62	10
Supermarket ^a								
All	37	6.6%	454	80.5%	73	12.9%	564	70
Narrow	2	1.7%	90	77.6%	24	20.7%	116	18
Moderate	2	2.9%	59	86.8%	7	10.3%	68	14
Broad	7	10.4%	54	80.6%	6	9.0%	67	9
Neighbourhood Definition								
Fast Food Outlets - Adults								
Areal measures	10	17.9%	45	80.4%	1	1.8%	56	18
Person-centric measures	42	23.5%	128	71.5%	9	5.0%	179	31
Buffer < 1 km	9	36.0%	16	64.0%	0	0.0%	25	6
Buffer 1–2 km	10	18.5%	43	79.6%	1	1.9%	54	17
Buffer > 2 km	15	34.1%	28	63.6%	1	2.3%	44	11
Fast Food Outlets - Children								
Areal measures	6	18.8%	24	75.0%	2	6.3%	32	7
Person-centric measures	26	19.7%	102	77.3%	4	3.0%	132	23
Buffer ≤ 400 m	10	25.0%	30	75.0%	0	0.0%	40	8
Buffer > 400 m	1	2.0%	45	91.8%	3	6.1%	49	13
Supermarket/Grocery - Adults								
Areal measures	11	12.2%	71	78.9%	8	8.9%	90	19
Person-centric measures	23	8.0%	222	77.4%	42	14.6%	287	33
Buffer < 1 km	3	8.3%	30	83.3%	3	8.3%	36	6
Buffer 1–2 km	8	10.8%	53	71.6%	13	17.6%	74	15
Buffer > 2 km	6	7.3%	62	75.6%	14	17.1%	82	8
Supermarket/Grocery - Children								
Areal measures	2	2.9%	62	88.6%	6	8.6%	70	6
Person-centric measures	1	0.9%	95	84.8%	16	14.3%	112	19
Buffer ≤ 400 m	0	0.0%	17	100.0%	0	0.0%	17	5
Buffer > 400 m	1	1.9%	49	94.2%	2	3.8%	52	9
Metric Type								
Fast Food Outlets								
Count (non-standardised)	7	18.9%	25	67.6%	5	13.5%	37	17
Count/area	31	22.8%	104	76.5%	1	0.7%	136	26
Count/capita	7	20.0%	28	80.0%	0	0.0%	35	11
Presence/absence	0	0.0%	25	96.2%	1	3.8%	26	8
Proximity	22	28.6%	53	68.8%	2	2.6%	77	25
Supermarket/Grocery								
Count (non-standardised)	8	6.6%	102	83.6%	12	9.8%	122	16
Count/area	9	5.7%	131	83.4%	17	10.8%	157	17
Count/capita	4	7.5%	40	75.5%	9	17.0%	53	10
Presence/absence	6	18.8%	23	71.9%	3	9.4%	32	11
Proximity	6	6.3%	75	78.1%	15	15.6%	96	26
Alternative Measures								
Healthy – composite	2	3.0%	48	72.7%	16	24.2%	66	9
Unhealthy – relative	23	21.3%	85	78.7%	0	0.0%	108	12
Healthy – score	9	9.2%	74	75.5%	15	15.3%	98	7
Total outlets	3	7.7%	33	84.6%	3	7.7%	39	8
Total restaurants (including fast food)	2	8.3%	19	79.2%	3	12.5%	24	8
Total food stores	3	8.6%	30	85.7%	2	5.7%	35	5

N = number of associations. Supplement 6 provides details on definitions of 'narrow', 'moderate' and 'broad' scope.

*'Positive associations' refer to statistically significant ($p < 0.05$) associations indicating increased access/exposure to food outlets is associated with increased obesity, and 'negative associations' refer to statistically significant associations indicating increased access/exposure to food outlets is associated with decreased obesity.

^a Excludes grocery stores, unless these were included under the same classification as supermarkets.

undermines these efforts. We suggest that authors and journal editors take greater responsibility for ensuring the complete reporting of RFE measurement methods, for example through use of the GeoFERN framework (Wilkins et al., 2017). A reduction in the diversity of measures used would also be of benefit moving forward. Researchers should give closer scrutiny to the methods used to ensure, where possible, that the best or most accurate methods are being used, such as use of validated secondary RFE data and accurate geocoding methods.

It is hoped that the findings from this review motivate further research into the comparability of methods within each of the five dimensions of diversity. Some research has been done relating to the choice of food outlet data (Mendez et al., 2016; Hobbs et al., 2016; Powell et al., 2011; Liese et al., 2010; Lake et al., 2010; Burgoine and Harrison, 2013) and RFE measures (Clary et al., 2015, 2016; Polsky et al., 2016; Feng et al., 2018; Shier et al., 2012; Mason et al., 2013). However, the other dimensions remain largely unexplored. Understanding the impacts of different methodological approaches will not only aid collation and interpretation of existing research, but may highlight best practice methods and help standardise measures used in future research.

One particular priority area for future research is in relation to the definition of food outlet constructs. Considerable diversity was observed across food outlet definitions. For example, fast food outlets were often defined narrowly as comprising only chain fast food outlets, and in other cases were defined broadly to include not only traditional non-chain fast food outlets, but also outlets such as coffee and sandwich shops, and desert shops. This diversity exists in spite of the existence and frequent citation of several industry-standard classification schemes (NAICS and SIC). Indeed, even when standardised classification schemes were cited, they were inconsistently applied. To our knowledge, no study has ever explored the impact of using different definitions for a given outlet construct, so it is unclear whether distinctions between different definitions of outlet constructs are important.

One dimension with particularly high diversity was the choice of RFE measure. For example, while areal and buffer metrics were used 242 times across the 113 studies, specific measures were used, at most, 15 times (count per area within 800 m - 1,600 m Euclidian buffers), and commonly no more than once. As mentioned, some research is beginning to investigate the impacts of using different measures – often focussing on the difference between ‘relative’ (e.g. ratio of healthy to unhealthy outlets) and ‘absolute’ (e.g. outlet count) measures (Clary et al., 2015; Feng et al., 2018; Maguire et al., 2017) or buffer sizes (Thornton et al., 2012; Fan et al., 2014; Burgoine et al., 2013). However, given the high degree of diversity among RFE metrics, this remains another key area for further research.

4.2. RFE-obesity associations

Previous reviews of RFE-obesity associations are limited in that they do not account for differences in measurement methods when collating the evidence (Williams et al., 2014; Casey et al., 2014), or only account for methods in relatively simplistic ways (Cobb et al., 2015; Gamba et al., 2015). However, collation of evidence from disparate methods may be misleading and could hide important associations. This review is the first to systematically stratify study findings by detailed methodological characteristics in order to examine how these factors may interact with outcomes. While reporting of methods was generally poor, there was a sufficient number of studies reporting methodological information with each domain to enable comparisons across methods; with the exception of the geocoding domain.

In agreement with existing reviews (Feng et al., 2010; Williams et al., 2014; Gamba et al., 2015; Casey et al., 2014), we found that overall, null associations considerably outnumbered statistically significant associations. This review is the first to demonstrate, however, that null associations remain the dominant outcome across all RFE

measurement methods. The impact on effect sizes was not considered due to the diverse methods, which made collation of effect sizes impossible at the scale of this review. However, the high prevalence of null results does suggest any associations are likely to be small, irrespective of the methods used, given the large sample sizes used in most studies. That said, there was a tendency toward more positive than negative associations for fast food outlets, which persisted across most methods (for 18/22 investigated methodological groupings, positive associations were between 16 and 36% of all associations while negative associations were < 5%). As p-values are a function of sample size, these findings do not imply meaningfulness of an association. Nevertheless, a consistent trend towards more associations in one direction versus another may be suggestive of a ‘true’ association; albeit of unknown magnitude and recognising that these trends might be an artefact of publication bias, or diversity across methods and populations. Additionally, the influence of methods and population characteristics on the distribution of statistically significant associations is of interest in itself, given that p values seem to be the key outcome many authors and policymakers focus on (Sterne and Smith, 2001).

A further key finding was that the distribution of null and statistically significant associations varied across measurement methods. While it is not possible to attribute this variation entirely to methodological factors (due to differences across studies in sample size, context and other methodological factors not accounted for within methodological groupings, or simply by chance), there were some notable differences that warrant further investigation. Researchers should also ensure that findings are interpreted in view of the methods employed; particularly when collating evidence and translating research into policy.

Of particular note, the distributions of positive, negative and null associations were more supportive of RFE-obesity associations for narrower definitions of ‘supermarkets’ compared to broad definitions. This is a novel finding; as mentioned above, no study has investigated the impact of construct definitions on associations with obesity. Theoretically, narrow construct definitions may provide better measures of the RFE as they may capture food outlets with a more consistent type of food provision. These findings reinforce the above-mentioned need for research into the comparability of different construct definitions and for researchers to clearly define food outlet constructs.

We additionally found the distribution of associations varied across different RFE metrics. For example, the tendency towards positive rather than negative associations between fast food outlets and obesity was considerably stronger for proximity measures than for measures of presence/absence (e.g. 28.6% positive, 2.5% negative vs 0.0% positive, 3.8% negative). Of relevance to RFE policy, which often restricts development of new fast food outlets within 400 m of schools (Public Health England, 2014; Local Government Association, 2016), the distribution of associations was more strongly supportive of a link between fast food outlets and obesity among children for buffers \leq 400 m than for larger buffer sizes. These findings are in broad support of other newly emerging research, which has shown different RFE metrics may critically impact the strength and direction of associations observed between the RFE and obesity-related outcomes, both in terms of the type/unit of measurement (Clary et al., 2015, 2016; Polsky et al., 2016; Mason et al., 2013; Bivoltis et al., 2018; Thornton et al., 2009), and the geographic scope (Thornton et al., 2012; Fan et al., 2014; Burgoine et al., 2013).

It is also worth noting that the distribution of associations varied across population groups. ‘Deprivation amplification’ – whereby people of lower-SES are more strongly influenced by their immediate RFE – has been observed in several studies (Burgoine et al., 2016; Vogel et al., 2017), and we found a stronger tendency toward more positive than negative associations for convenience stores and fast food outlets among low-SES groups. In spite of this, many studies do not allow for potential divergent effects across population groups (possibly due to

insufficient sample sizes), potentially hiding important associations and explaining the high prevalence of null results. If policymakers are to intervene in relation to the RFE, it is imperative that we understand interactions between the RFE and population characteristics to ensure that interventions do not lead to widening health inequalities.

4.3. Limitations of existing research

The studies included within this review had several limitations in addition to those noted above. Overall, study quality was relatively poor, suggesting many studies are at risk of bias. Of most concern, given that this evidence is often used to inform RFE interventions, was the absence of causal frameworks from all but three studies. Causal frameworks inform covariate selection to allow more robust causal inference in observational research (Pearl, 2009). In relation to this, many studies did not account for competing aspects of the built environment which may be correlated with RFE measures. For example, places that have a high availability of unhealthy food retailing may also have a high availability of healthy food retailing, and may be more conducive of walking, due to a higher accessibility of general facilities/destinations (Polsky et al., 2016; James et al., 2014). Without accounting for such competing exposures, associations between specific RFE measures and obesity will be biased. Use of a causal framework would identify these necessary covariates. Recent evidence also suggests that mutual adjustment for competing food outlet types (e.g. both 'healthy' and 'unhealthy' outlets) may be critical in detecting statistically significant associations (Clary et al., 2015; Fiechtner et al., 2015; Bodor et al., 2010; Burgoine et al., 2014), although many of these studies often found no appreciable impact on effect sizes. The above notwithstanding, we found no substantive differences in our findings after restricting to papers within the top decile of quality score.

This review also highlights the vast number of studies that have examined the RFE around the home. However, GPS and travel diary studies show that home-centric neighbourhoods do not correspond well with people's actual activity spaces (Christian, 2012; Crawford et al., 2014), raising questions around the appropriateness of home-centric measures. It is also notable that the majority of research – including numerous longitudinal studies – measure the RFE at only a single time-point, limiting ability to make causal inferences. Studies investigating changes in the RFE through the 1970s - 1990s, when the RFE saw the greatest shifts in food retailing (White, 2007; Wrigley, 2002; Guy and David, 2004; Walker et al., 2010) may provide the greatest opportunities for understanding the impact of the RFE on weight status. Further limitations of the RFE-obesity literature include lack of data on food outlet utilisation and the within-store environment (e.g. pricing, food quality) and failure to account for alternative purchasing opportunities, such as online supermarkets, delivery services, and non-traditional food stores, such as clothes shops and pharmacies (Lucan et al., 2018). Many of these limitations appear to be driven by the availability of secondary data (or lack thereof). Nevertheless, use of spatial methods to operationalise the RFE can also be celebrated in that it has enabled investigation of the RFE at a population level; which is important for national and regional-level policymaking.

4.4. Strengths and limitations of review

This review has several strengths, most notably our systematic search strategy, the very large number of studies included in the review, and the breadth and detail of the data extraction, which provide rich information on the methods used and allow detailed analysis of the distribution of RFE-obesity associations, accounting for measurement methods and population groups.

It is worth reiterating that we decided *a-priori* not to extract effect sizes, because the heterogeneity of RFE measures would preclude collation of these data. Following similar approaches to other reviews in this area (Cobb et al., 2015; Williams et al., 2014; Sallis et al., 2000), we

instead counted the distribution of null and statistically significant associations, together with their associated directions. Our findings do not provide any information regarding the strength of associations. Indeed, the p-value is a function of sample size, and the presence/absence of a significant p-value thus does not imply meaningfulness of an association. Nevertheless, by collating the numbers of statistically significant associations across multiple studies, we were able to infer the possible presence of 'true' associations (of unknown size) from the distribution of associations. In the absence of any 'true' association, the numbers of spurious statistically significant positive and negative associations should be approximately equal. The greater the tendency for more statistically significant associations in one direction than the other, the more suggestive the data of a 'true' association. A limitation of this approach is that publication bias may tend to inflate the numbers of associations in the expected direction. That said, positive and negative associations were balanced for supermarkets/grocery stores, convenience stores and restaurants across the general population, suggesting our results may not be substantively impacted by publication bias. Nevertheless, our findings need to be interpreted with caution in this regard.

This review is the first to consider in detail the methods used to measure the RFE when collating the evidence base. However, within methodological groupings, there was still heterogeneity, which may have confounded our results, and it is not possible to attribute variation in the distribution of associations to methodological factors alone. Due to the high prevalence of missing methodological information, we did not contact authors to obtain missing data, and our results are therefore only representative of those studies that reported complete information for a given methodological aspect. We reviewed the aims and objectives, methods and results sections of papers in detail, so may have occasionally missed methodological information that was reported elsewhere. The Cobb review was limited to studies conducted in the US and Canada only. While we expanded the top-up search to cover other western countries, reliance on the Cobb review to identify earlier studies means US and Canadian studies are over-represented, reducing the generalisability of our findings across western countries. Nevertheless, in sensitivity analyses we restricted our analyses to only those studies identified in the top-up search, and found no notable differences as compared to the full sample of studies. This is unsurprising, given the dominance of US studies both in the Cobb review and the top-up search, suggesting our findings are still of reasonable generalisability across western countries. Lastly, we did not consider other differences across studies such as analysis methods or outcome measures, which may have also impacted study findings.

5. Conclusion

Associations between the RFE and obesity are nuanced, and depend upon the methods used to measure the RFE. However, null associations appear to be the predominant outcome across all measurement methods. At present, the reporting of methods is poor, and severely limits inferences that can be drawn from the evidence base, and the translation of evidence into policy. Authors and journal editors should ensure more robust reporting of RFE measurement methods, for example through use of specially developed reporting frameworks. Authors are also responsible for articulating study findings in the context of the methods employed, so that policymakers can correctly interpret RFE-obesity associations. Direct comparisons between studies employing different methods should be avoided, at least until further evidence emerges in relation to the comparability of different methods. Moving forward, researchers should be more critical of the methods used to ensure the best or most accurate methods are used where possible.

Contributions

EW, DR, MM and CG were responsible for study conception and design. EW coordinated the review and acted as the primary reviewer. All other authors contributed as secondary reviewers. EW led the writing of the manuscript. All other authors provided critical feedback to shape the manuscript.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Disclosure

The authors declared no conflict of interest.

Acknowledgements

Michelle Morris' time was funded via ESRC funding ref: ES/N00941X/1 and Wellcome Trust funding ref: 105615/Z/14/Z.

Appendix ASupplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.healthplace.2019.02.007>.

References

- An, R., Sturm, R., 2012. School and residential neighborhood food environment and diet among California youth. *Am. J. Prev. Med.* 42, 129–135.
- Bernsdorf, K.A., Lau, C.J., Andreasen, A.H., Toft, U., Lykke, M., Glümer, C., 2017. Accessibility of fast food outlets is associated with fast food intake. A study in the Capital Region of Denmark. *Health Place* 48, 102–110.
- Bivoltzis, A., Cervigni, E., Trapp, G., Knudman, M., Hooper, P., Ambrosini, G.L., 2018. Food environments and dietary intakes among adults: does the type of spatial exposure measurement matter? A systematic review. *Int. J. Health Geogr.* 17, 1–20.
- Black, C., Moon, G., Baird, J., 2014. Dietary inequalities: what is the evidence for the effect of the neighbourhood food environment? *Health Place* 27, 229–242.
- Bodur, J.N., Rice, J.C., Farley, T.A., Swalm, C.M., Rose, D., 2010. The association between obesity and urban food environments. *J. Urban Health: Bull. N. Y. Acad. Med.* 87, 771–781.
- Burgoine, T., Harrison, F., 2013. Comparing the accuracy of two secondary food environment data sources in the UK across socio-economic and urban/rural divides. *Int. J. Health Geogr.* 12, 1–8.
- Burgoine, T., Lake, A.A., Stamp, E., Alvanides, S., Mathers, J.C., Adamson, A.J., 2009. Research report: changing foodscapes 1980–2000, using the ASH30 Study. *Appetite* 53, 157–165.
- Burgoine, T., Alvanides, S., Lake, A.A., 2013. Creating 'obesogenic realities': do our methodological choices make a difference when measuring the food environment? *Int. J. Health Geogr.* 12, 1–9.
- Burgoine, T., Forouhi, N.G., Griffin, S.J., Wareham, N.J., Monsivais, P., 2014. Associations between exposure to takeaway food outlets, takeaway food consumption, and body weight in Cambridgeshire, UK: population based, cross sectional study. *BMJ* 348, 1–10.
- Burgoine, T., Forouhi, N.G., Griffin, S.J., Brage, S., Wareham, N.J., Monsivais, P., 2016. Does neighborhood fast-food outlet exposure amplify inequalities in diet and obesity? A cross-sectional study. *Am. J. Clin. Nutr.* 103, 1540–1547.
- Casey, R., Oppert, J.-M., Weber, C., Charreire, H., Salze, P., Badariotti, D., et al., 2014. Determinants of childhood obesity: what can we learn from built environment studies? *Food Qual. Prefer.* 31, 164–172.
- Caspi, C.E., Sorensen, G., Subramanian, S., Kawachi, I., 2012. The local food environment and diet: a systematic review. *Health Place* 18, 1172–1187.
- Charreire, H., Casey, R., Salze, P., Simon, C., Chaix, B., Banos, A., et al., 2010. Measuring the food environment using geographical information systems: a methodological review. *Publ. Health Nutr.* 13, 1773–1785.
- Chen, H.-J., Wang, Y., 2016. Changes in the neighborhood food store environment and children's body mass index at peripuberty in the United States. *J. Adolesc. Health* 58, 111–118.
- Christian, W.J., 2012. Using geospatial technologies to explore activity-based retail food environments. *Spatio Temporal Epidemiol.* 3, 287–295.
- Clary, C.M., Ramos, Y., Shareck, M., Kestens, Y., 2015. Should we use absolute or relative measures when assessing foodscape exposure in relation to fruit and vegetable intake? Evidence from a wide-scale Canadian study. *Prev. Med.* 71, 83–87.
- Clary, C., Lewis, D.J., Flint, E., Smith, N.R., Kestens, Y., Cummins, S., 2016. The local food environment and fruit and vegetable intake: a geographically weighted regression approach in the ORiEL study. *Am. J. Epidemiol.* 184, 837–846.
- Cobb, L.K., Appel, L.J., Franco, M., Jones-Smith, J.C., Nur, A., Anderson, C.A.M., 2015. The relationship of the local food environment with obesity: a systematic review of methods, study quality, and results. *Obesity* 23, 1331–1344.
- Crawford, T.W., Jilcott Pitts, S.B., McGuirt, J.T., Keyserling, T.C., Ammerman, A.S., 2014. Conceptualizing and comparing neighborhood and activity space measures for food environment research. *Health Place* 30, 215–225.
- Dawson, J., Marshall, D., Taylor, M., Cummins, S., Sparks, L., Anderson, A.S., 2008. Accessing healthy food: availability and price of a healthy food basket in Scotland. *J. Mark. Manag.* 24, 893–913.
- Diller, P.A., Graff, S., 2011. Regulating food retail for obesity prevention: how far can cities go? *J. Law Med. Ethics* 39, 89–93.
- Dwicaksono, A., Brissette, I., Birkhead, G.S., Bozlak, C.T., Martin, E.G., 2017. Evaluating the contribution of the built environment on obesity among New York State students. *Health Educ. Behav.* 45, 1–12.
- Fan, J.X., Hanson, H.A., Zick, C.D., Brown, B.B., Kowaleski-Jones, L., Smith, K.R., 2014. 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* 4.
- Feng, J., Glass, T.A., Curriero, F.C., Stewart, W.F., Schwartz, B.S., 2010. The built environment and obesity: a systematic review of the epidemiologic evidence. *Health Place* 16, 175–190.
- Feng, X., Astell-Burt, T., Badland, H., Mavoa, S., Giles-Corti, B., 2018. Modest ratios of fast food outlets to supermarkets and green grocers are associated with higher body mass index: longitudinal analysis of a sample of 15,229 Australians aged 45 years and older in the Australian National Liveability Study. *Health Place* 49, 101–110.
- Fiechtner, L., Mona, S., Sequist, T., Block, J., Duncan, D.T., Melly, S.J., et al., 2015. Food environments and childhood weight status: effects of neighborhood median income. *Child. Obes.* 11, 260–268.
- Gamba, R.J., Schuchter, J., Rutt, C., Seto, E.Y., 2015. Measuring the food environment and its effects on obesity in the United States: a systematic review of methods and results. *J. Community Health* 40, 464–475.
- Gibson, D.M., 2011. The neighborhood food environment and adult weight status: estimates from longitudinal data. *Am. J. Public Health* 101, 71–78.
- Glanz, K., Sallis, J.F., Saelens, B.E., Frank, L.D., 2005. Healthy nutrition environments: concepts and measures. *Am. J. Health Promot.* 19, 330–333.
- Greater London Authority, 2012. Takeaways Toolkit. [cited 2017 24 April]. Available from: <https://www.london.gov.uk/sites/default/files/takeawaystoolkit.pdf>.
- Guy, C.M., David, G., 2004. Measuring physical access to 'healthy foods' in areas of social deprivation: a case study in Cardiff. *Int. J. Consum. Stud.* 28, 222–234.
- Hobbs, M., Green, M., Griffiths, C., Jordan, H., McKenna, J., 2016. How different data sources and definitions of neighbourhood influence the association between food outlet availability and body mass index: a cross-sectional study. *Perspect. Publ. Health* 137, 158–161.
- James, P., Berrigan, D., Hart, J.E., Hipp, J.A., Hoehner, C.M., Kerr, J., et al., 2014. Effects of buffer size and shape on associations between the built environment and energy balance. *Health Place* 27, 162–170.
- Jeffery, R.W., Baxter, J., McGuire, M., Linde, J., 2006. Are fast food restaurants an environmental risk factor for obesity? *Int. J. Behav. Nutr. Phys. Act.* 3, 1–6.
- Lake, A.A., Burgoine, T., Greenhalgh, F., Stamp, E., Tyrrell, R., 2010. The foodscape: classification and field validation of secondary data sources. *Health Place* 16, 666–673.
- Li, F., Harmer, P., Cardinal, B.J., Bosworth, M., Johnson-Shelton, D., 2009. Obesity and the built environment: does the density of neighborhood fast-food outlets matter? *Am. J. Health Promot.* 23, 203–209.
- Liese, A.D., Colabianchi, N., Lamichhane, A.P., Barnes, T.L., Hibbert, J.D., Porter, D.E., et al., 2010. Validation of 3 food outlet databases: completeness and geospatial accuracy in rural and urban food environments. *Am. J. Epidemiol.* 172, 1324–1333.
- Local Government Association, 2016. Tipping the Scales. Case Studies on the Use of Planning Powers to Limit Hot Food Takeaways. Local Government Association, London.
- Lopez, R.P., 2007. Neighborhood risk factors for obesity. *Obesity (Silver Spring, Md)* 15, 2111–2119.
- Lucan, S.C., Maroko, A.R., Patel, A.N., Gjonbalaj, I., Abrams, C., Rettig, S., et al., 2018. Change in an urban food environment: storefront sources of food/drink increasing over time and not limited to food stores and restaurants. *J. Acad. Nutr. Diet Elsevier*.
- Lytel, L.A., Sokol, R.L., 2017. Measures of the food environment: a systematic review of the field, 2007–2015. *Health Place* 44, 18–34.
- Maguire, E.R., Burgoine, T., Monsivais, P., 2015. Area deprivation and the food environment over time: a repeated cross-sectional study on takeaway outlet density and supermarket presence in Norfolk, UK, 1990–2008. *Health Place* 33, 142–147.
- Maguire, E.R., Burgoine, T., Penney, T.L., Forouhi, N.G., Monsivais, P., 2017. Does exposure to the food environment differ by socioeconomic position? Comparing area-based and person-centred metrics in the Fenland Study, UK. *Int. J. Health Geogr.* 16, 1–14.
- Mason, K.E., Bentley, R.J., Kavanagh, A.M., 2013. Fruit and vegetable purchasing and the relative density of healthy and unhealthy food stores – evidence from an Australian multilevel study. *J. Epidemiol. Community Health* 67, 231–236.
- Mendez, D.D., Kim, K.H., Hardaway, C.R., Fabio, A., 2016. Neighborhood racial and socioeconomic disparities in the food and alcohol environment: are there differences by commercial data sources? *J. Racial Ethn. Health Disparities* 3, 108–116.
- Nykiforuk, C.L.J., Campbell, E.J., Macridis, S., McKennitt, D., Atkey, K., Raine, K.D., 2018. Adoption and diffusion of zoning bylaws banning fast food drive-through services across Canadian municipalities. *BMC Public Health* 18, 1–12.
- Pearl, J., 2009. Causal inference in statistics: an overview. *Stat. Surv.* 3, 96–146.
- Polsky, J.Y., Moineddin, R., Dunn, J.R., Glazier, R.H., Booth, G.L., 2016. Absolute and relative densities of fast-food versus other restaurants in relation to weight status:

- does restaurant mix matter? *Prev. Med.* 82, 28–34.
- Popkin, B., Adair, L., Ng, S.W., 2012. Global nutrition transition and the pandemic of obesity in developing countries. *Nutr. Rev.* 70, 3–21.
- Powell, L.M., Han, E., Zenk, S.N., Khan, T., Quinn, C.M., Gibbs, K.P., et al., 2011. Field validation of secondary commercial data sources on the retail food outlet environment in the US. *Health Place* 17, 1122–1131.
- Public Health England, 2014. Obesity and the Environment: Regulating the Growth of Fast Food Outlets. [cited 2018 01 May]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/296248/Obesity_and_environment_March2014.pdf.
- Reitzel, L.R., Regan, S.D., Nga, N., Cromley, E.K., Strong, L.L., Wetter, D.W., et al., 2014. Density and proximity of fast food restaurants and body mass index among African Americans. *Am. J. Public Health* 104, 110–116.
- Sallis, J.F., Prochaska, J.J., Taylor, W.C., 2000. A review of correlates of physical activity of children and adolescents. *Med. Sci. Sports Exerc.* 32, 963–975.
- Shier, V., An, R., Sturm, R., 2012. Is there a robust relationship between neighbourhood food environment and childhood obesity in the USA? *Publ. Health* 126, 723–730.
- Sparks, A.L., Bania, N., Leete, L., 2011. Comparative approaches to measuring food access in urban areas the case of Portland, Oregon. *Urban Stud.* 48, 1715–1737.
- Sterne, J.A., Smith, G.D., 2001. Sifting the evidence—what's wrong with significance tests? *Phys. Ther.* 81, 1464–1469.
- Thomsen, M.R., Nayga, R.M., Alviola, P.A., Rouse, H.I., 2016. The effect of food deserts on the body mass index of elementary school children. *Am. J. Agric. Econ.* 98, 1–18.
- Thornton, L.E., Bentley, R.J., Kavanagh, A.M., 2009. Fast food purchasing and access to fast food restaurants: a multilevel analysis of VicLANES. *Int. J. Behav. Nutr. Phys. Act.* 6, 28.
- Thornton, L.E., Pearce, J.R., Macdonald, L., Lamb, K.E., Ellaway, A., 2012. Does the choice of neighbourhood supermarket access measure influence associations with individual-level fruit and vegetable consumption? A case study from Glasgow. *Int. J. Health Geogr.* 11, 1–12.
- Vogel, C., Lewis, D., Ntani, G., Cummins, S., Cooper, C., Moon, G., et al., 2017. The relationship between dietary quality and the local food environment differs according to level of educational attainment: a cross-sectional study. *PLoS One* 12, 1–16.
- Walker, R.E., Keane, C.R., Burke, J.G., 2010. Disparities and access to healthy food in the United States: a review of food deserts literature. *Health Place* 16, 876–884.
- Waltham Forest Spatial Planning Unit, 2009. Hot Food Takeaway Supplementary Planning Document. [cited 2016 27 May]. Available from: <https://branding.walthamforest.gov.uk/documents/spd-hot-food-takeaway-mar10.pdf>.
- Wells, G.A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., et al., The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-Analyses. n.d. [cited 2018 6 Aug]. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.
- White, M., 2007. Food access and obesity. *Obes. Rev.* 8, 99–107.
- Wilkins, E.L., Morris, M.A., Radley, D., Griffiths, C., 2017. Using Geographic Information Systems to measure retail food environments: discussion of methodological considerations and a proposed reporting checklist (Geo-FERN). *Health Place* 44, 110–117.
- Williams, J., Scarborough, P., Matthews, A., Cowburn, G., Foster, C., Roberts, N., et al., 2014. A systematic review of the influence of the retail food environment around schools on obesity-related outcomes. *Obes. Rev.* 15, 359–374.
- Wong, M.S., Chan, K.S., Jones-Smith, J.C., Colantuoni, E., Thorpe, R.J., Bleich, S.N., 2017. The neighborhood environment and obesity: understanding variation by race/ethnicity. *Prev. Med.* 111, 371–377.
- World Obesity Federation, 2017. Trends in Global Obesity. [cited 2018 01 May]. Available from: <https://www.worldobesity.org/data/obesity-data-repository/resources/trends/12/>.
- Wrigley, N., 2002. 'Food deserts' in British cities: policy context and research priorities. *Urban Stud.* 39, 2029–2040.
- Zhao, Z., Kaestner, R., Xu, X., 2014. Spatial mobility and environmental effects on obesity. *Econ. Hum. Biol.* 14, 128–140.