How different data sources and definitions of neighbourhood influence the association between food outlet availability and body mass index: a cross-sectional study

Abstract

Inconsistencies in methodologies continue to inhibit understanding of the impact of the environment on body mass index (BMI). To estimate the effect of these differences we assessed the impact of using different definitions of neighbourhood and datasets on associations between food outlet availability within the environment and BMI. Previous research has not extended to show any differences in the strength of associations between food outlet availability and BMI across both different definitions of neighbourhood and datasets. Descriptive statistics showed differences in the number of food outlets, particularly other food retail outlets between different datasets and definitions of neighbourhood. Despite these differences, our key finding was that across both different definitions of neighbourhood and datasets there was very little difference in size of associations between food outlets and BMI. Researchers should consider and transparently report the impact of methodological choices such as the definition of neighbourhood and acknowledge any differences in associations between the food environment and BMI.

Key words
Food outlets; body mass index; density; buffer; lower-super output area; neighbourhood.

Obesity is one of the leading burdens of disease in the UK costing an estimated £5.1 billion per year.[1] Both research and policy now suggest that the 'obesogenic environment' may be a contributing factor to obesity based on the principle that an increased food outlet availability within an individual's neighbourhood may encourage an overconsumption of energy-dense, nutrient poor foods. Despite this, findings linking food outlet availability and body mass index (BMI) are inconsistent. This may be due to large variations in methodologies between studies, two major issues being; the use of a variety of food outlet datasets and inconsistencies in neighbourhood definitions.[2-4] A single study has begun to establish that although Local Authority (LA) food outlet datasets may be more accurate than Point of Interest (PoI) datasets, yet PoI is still considered a viable alternative.[5] Despite this progress, no research to date has assessed whether differences between different food outlet datasets as well as different definitions of neighbourhood impact on the strength of associations seen between food outlet availability and BMI.

The neighbourhood definition that best represents actual food outlet usage remains unknown.[4] Two definitions of neighbourhood (geocoded around a participant's home) currently dominate the evidence base; administratively defined areas such as a lower-super output area (LSOAs) and arbitrary defined radial buffers[6]. Radial buffers represent a viable alternative to administratively defined neighbourhood areas in large epidemiological studies. However, studies rarely model and measure the environment in the same way and the choices made when selecting a definition of neighbourhood or dataset are rarely challenged rigorously.[6] In order to investigate the impact of differences in choice of data set and definition of neighbourhood, we compared two different datasets of food outlet locations and three different definitions of neighbourhood.

This cross-sectional study uses individual-level data from the Yorkshire Health Study (YHS) which offers a large range of self-reported health-related information such as height and weight on a representative population.[7] Participants within Rotherham LA were exported from the YHS (n=27,809) yielding a final sample of n=4,723 participants who resided within 134 of 166 LSOAs (average of 35 individuals per LSOA) in Rotherham LA. Ethical clearance was granted by the ethics committee of the Carnegie Faculty, Leeds Beckett University.

Data on the food environment was obtained from two sources; (i) the UK Ordnance Survey Points of Interest (PoI) dataset and (ii) Rotherham LA. The PoI dataset contains the location of all commercial facilities across England. The PoI dataset is pre-coded into different categories and classes of commercial services.[8] Rotherham LA provided their current environmental health food outlet records for temporal comparison. Food outlets from both datasets were then categorised by the author into three groups; (i) supermarkets, (ii) takeaways and (iii) other food retail (such as petrol stations, convenience stores selling food).

Home addresses were geocoded based on post-code. Based on previous research,[6] three commonly used definitions of neighbourhood exposure were computed in ArcGIS (*version 10.2.2, ESRI Inc., Redlands, CA*) around the geocoded home location; i) an 800m radial buffer ii) a 2000m radial buffer iii) defined by identifying which LSOA an individual resided in. A LSOA is an administratively defined geographical area that typically contains a minimum population of 1000 and a mean of 1500. A count of food outlets per buffer (800m and 2000m) and density per LSOA (km^2) was computed. LSOA sizes (km^2) was obtained from the 2011 Population Census. Food outlets falling within these buffers and LSOAs were then identified, counted and joined within ArcGIS based on a unique identifier in both the environment dataset and YHS dataset to provide a unique count for each individual based on an 800m, 2000m radial buffer and per LSOA (km^2). IMD (Index of Multiple Deprivation) scores were assigned to the lower super-output area (LSOA) of each individual, as determined by their geocoded postcode.

Single-level linear regression (β, 95% confidence intervals (CI)) was used to assess the association between radial buffers and BMI. A multi-level modelling (MLM) framework accounted for the hierarchical data structure when people were nested within administrative areas (LSOA). Linear MLMs were used to identify how LSOAs were associated with BMI. Both models adjusted for both individual- and neighbourhood-level factors. Age, gender, ethnicity, rural or urban status (local government classification) and area level socio-economic status (IMD) were included in all analyses as covariates. Similar to census estimates (12.0%), 9.2% of participants resided in rural areas. Differences in the magnitudes of associations were then assessed across different datasets and neighbourhood definitions by assessing the change in (β and 95% CI). All statistical analysis were performed using STATA IC version 14.

INSERT TABLE 1 HERE

Our results show that the LA dataset contained approximately twice as many food outlet records as the Point of Interest (PoI) dataset. However, despite some differences in the count of food outlets, very few differences in the strength or direction of associations between food outlets and BMI were observed when using different datasets or neighbourhood definitions. There was little difference in count for supermarkets and takeaways, with 8 and 23 additional outlets identified within the LA dataset. The main discrepancy was an additional 589 other food retail outlets (Table 1). Furthermore, food outlet count varied at the individual level; for instance within an 800m radial buffer LA data showed that some individuals had no fast-food outlets within their neighbourhood, whilst the average had 1.48±2.04 and the maximum experienced was 23.00. Overall, of 24 associations, only 2 differences were noted both of which involved supermarkets. First, within an 800m buffer supermarkets were significantly associated with BMI in the PoI (β=0.392 (95% CI 0.123; 0.662)) but not LA dataset (β= 0.121 (-0.171; 0.414)). Second, supermarkets were associated with BMI within the PoI dataset when using radial buffers (β=0.214 (95% CI 0.09; 0.339)) but not LSOA (β= 0.027 (-0.114; 0.169)) (Table 1). Despite these differences for supermarkets, all other associations were substantively the same.

Despite some differences by count, our findings agree with previous research that suggests there is little change in size and direction of associations across different definitions of neighbourhood and datasets.[9] Only supermarkets exhibited some differences across neighbourhood definitions and datasets in both strength and direction of associations with BMI. This finding is particularly interesting considering the PoI dataset contained only eight fewer supermarkets and that more supermarkets are associated with an increase in BMI, opposite to the hypothesised direction. This may suggest such differences for supermarkets in particular should not be overlooked. Other evidence supports this and suggests neighbourhood definition may have significant implications on findings.[4 9] Bodicoat et al. (2015) showed that fast-food outlets were weakly but positively associated with type II diabetes in smaller radial buffers but not obesity (100m or 250m).[9] However, within larger neighbourhood definitions (500m, 750m, 1000m) the number of fast-food outlets were associated with type II diabetes, obesity and fasting glucose. James et al. (2014) also showed that for intersection count the strongest effect sizes were seen in the 400m buffers; effects reduced as buffer sizes got larger i.e. to 1600m.[4] Studies often use or only report associations within one neighbourhood definition. Findings within this study suggest such differences may have some consequences for research findings but only for associations between supermarkets and BMI.

This study contributes to the research in two ways. Firstly, the association between food outlets and BMI was assessed using different definitions of neighbourhood. Secondly, this paper examined the extent to which using different datasets may contribute to a lack of inter-study comparability. Given that the most appropriate criterion for defining neighbourhood remains open to debate, understanding any resulting differences in the magnitude of these associations is important yet rarely investigated or reported. Radial buffers have been proposed as an alternative to administrative boundaries to represent an individual's actual neighbourhood.[6] However, there remains no uniform definition between studies. Furthermore, most policy based decisions in the UK are still made according to administratively defined areas such as LSOA. For local level dissemination it could therefore be argued that administrative areas continue to inform local level policy best. However, it is important to remember that we were not able to ground truth to assess the true accuracy of each dataset. In summary, this study suggests that other than for supermarkets, different definitions of neighbourhood are broadly inconsequential in changing statistical inference.[4]

The uncertainty around using different secondary datasets and defining neighbourhood remains a complex issue for contemporary environment based research. One possible explanation for our lack of association of food outlets to BMI may be due to the lack of heterogeneity in area types. Only, 9.2% of individuals resided in rural areas, which is below the UK average. However, since the majority of individuals reside in urban areas in the UK, our results remain important. Future research should explore the accuracy of secondary datasets by ground truthing areas and extending their analyses to assess if inaccuracies do lead to substantive differences in associations between BMI and the environment. An additional complexity worth exploring is the impact of different classifications of food outlets, particularly as the main difference here was seen within other food retail outlets and supermarkets were associated with an increase in BMI. Furthermore, research may also explore additional definitions of neighbourhood such as proximity, street network buffers, self-defined buffers or GPS defined activity spaces by per km^2 and raw count.[3]

In conclusion, although differences in the count of outlets were identified, contrary to expectations, findings demonstrated few differences in the strength and direction of associations between food outlets and BMI across both different neighbourhood definitions and datasets. Ultimately, it may be difficult to achieve an accurate and standardised definition of neighbourhood within environmental research, particularly given the nature of individual behaviours. However, it is important to now rigorously challenge the choices made at a methodological level. It is beyond the scope of this paper to suggest the most appropriate definition of neighbourhood or dataset. However, research should consider and transparently report in a sensitivity analysis the impact of methodological choices such as the definition of neighbourhood on associations between the environment and BMI. Researchers should use the local context and problem being investigated to inform the most appropriate definition of neighbourhood and dataset used. That is until better evidence emerges suggesting any different.

References

1. Scarborough P, Bhatnagar P, Wickramasinghe KK, Allender S, Foster C, Rayner M. The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006-07 NHS costs. *Journal of public health* 2011;33(4):527-35.
2. Cobb LK, Appel LJ, Franco M, Jones-Smith JC, Nur A, Anderson CAM. The relationship of the local food environment with obesity: A systematic review of methods, study quality, and results. *Obesity* 2015:n/a-n/a.
3. Flowerdew R, Manley DJ, Sabel CE. Neighbourhood effects on health: does it matter where you draw the boundaries? *Social Science and Medicine* 2008;66(6):1241-55.
4. James P, Berrigan D, Hart JE, Aaron Hipp J, Hoehner CM, Kerr J, et al. Effects of buffer size and shape on associations between the built environment and energy balance. *Health & place* 2014;27(0):162-70.
5. Burgoine T, Harrison F. Comparing the accuracy of two secondary food environment data sources in the UK across socio-economic and urban/rural divides. *International journal of health geographics* 2013;12:2.
6. Feng J, Glass TA, Curriero FC, Stewart WF, Schwartz BS. The built environment and obesity: a systematic review of the epidemiologic evidence. *Health and Place* 2010;16(2):175-90.
7. Green M, Li J, Relton C, Strong M, Kearns B, Wu M, et al. Cohort Profile: The Yorkshire Health Study. *International Journal of Epidemiology* 2014:doi: 10.1093/ije/dyu121.
8. Ordnance Survey. *Points of Interest database - user guide and technical specification*. Southampton: Ordnance Survey, 2012.
9. Bodicoat DH, Carter P, Comber A, Edwardson C, Gray LJ, Hill S, et al. Is the number of fast-food outlets in the neighbourhood related to screen-detected type 2 diabetes mellitus and associated risk factors? *Public health nutrition* 2015;18(9):1698-705.

Table 1. The change in magnitude of association between the environment and BMI by neighbourhood definition and dataset

Data Source	Count (n)	LSOA Density (km^2)		800m Buffer		2000m buffer	
		β	95% CI	β	95% CI	β	95% CI
Local Authority (n=1,489)	All food outlets (n=1,489)	-0.003	[-0.010, 0.005]	-0.002	[-0.018; 0.014]	0.001	[-0.003; 0.005]
	Takeaways (n=257)	-0.001	[-0.035, 0.033]	0.013	[-0.056; 0.083]	0.013	[-0.014; 0.041]
	Other retail (n=1,172)	-0.004	[-0.014; 0.006]	-0.006	[-0.026; 0.015]	0.001	[-0.004; 0.005]
	Supermarkets (n=60)	-0.048	[-0.223, 0.127]	0.121	[-0.171; 0.414]	0.001	[-0.122; 0.124]
Point of Interest (n=869)	All food outlets (n=869)	-0.006	[-0.016; 0.003]	-0.005	[-0.023; 0.012]	-0.001	[-0.006; 0.005]
	Takeaways (n=234)	-0.010	[-0.045; 0.025]	0.014	[-0.041; 0.068]	-0.002	[-0.023; 0.019]
	Other retail (n=583)	-0.010	[-0.024; 0.003]	-0.016	[-0.040; 0.008]	-0.002	[-0.009; 0.005]
	Supermarkets (n=52)	0.027	[-0.114; 0.169]	*0.392	[0.123; 0.662]	*0.214	[0.090; 0.339]
		Mean(SD),Max+		**Mean(SD),Max+**		**Mean(SD),Max+**	
Local Authority (n=1,489)	All food outlets (n=1,489)	12.28(17.55),125.00		7.55(8.72),160.00		38.46(34.66),244.00	
	Takeaway (n=257)	2.21(4.08),20.83		1.48(2.04),23.00		7.03(5.59),33.00	
	Other retail (n=1,172)	9.73(13.72),104.61		5.80(6.97),135.00		30.22(29.08),204.00	
	Supermarkets (n=60)	0.33(0.81),6.25		0.27(0.49),3.00		1.21(1.32),7.00	
Point of Interest (n=869)	All food outlets (n=869)	7.49(14.22),94.08		4.86(8.25),114.00		24.69(26.80),170.00	
	Takeaways (n=234)	1.96(4.01),29.17		1.43(2.62),33.00		6.78(7.16),44.00	
	Other retail (n=583)	5.19(10.45),68.42		3.18(5.89),81.00		16.72(19.79),125.00	
	Supermarkets (n=52)	0.35(0.97),6.90		0.25(0.51),4.00		1.18(1.11),5.00	

Note: all models control for gender, ethnicity, deprivation and rural/urban classification of the neighbourhood.

* = significant ($p<0.05$)

+ = minimum value was zero for all types of outlets.

4