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

## Abstract

Large heterogeneity in methodologies continues to inhibit understanding of the impact of the environment on obesity. To estimate the effect of this heterogeneity we assessed the impact of using different definitions of availability and datasets on associations between food outlet availability within the environment and obesity. Previous research has not extended to show any differences in the strength of associations between food outlet availability and obesity across both different buffers and datasets. Our key finding was that there is little change in size of associations between the food outlet availability and obesity across both different food availability measures and datasets. It remains important to consider the impact of this choice in relation to the magnitude of change in associations.

**Key words**

Food environment; obesity; nutrition; availability, buffers.

Obesity is one of the leading burdens of disease in the UK[1](#_ENREF_1) costing an estimated £5.1 billion per year.[2](#_ENREF_2) 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 obesity are inconsistent. This may be due to large heterogeneity in methodologies between studies, examples of which include; the use of a variety of food outlet databases and inconsistencies in neighbourhood definitions.[3](#_ENREF_3) Previous evidence has begun to establish that although Local Authority (LA) food outlet databases may be more accurate than Point of Interest (PoI) databases, PoI is still considered a viable alternative to the LA database.[4](#_ENREF_4) Despite this progress, no research to date has assessed whether inaccuracies between food outlets databases impact on the strength of associations seen between the food outlet availability and obesity.

Inconsistencies in neighbourhood definitions may also explain the mixed associations seen between the food environment and obesity.[5](#_ENREF_5)[6](#_ENREF_6) The neighbourhood definition that best represents actual food outlet usage remains unknown.[6](#_ENREF_6) However, 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) or arbitrary defined radial buffers. Radial buffers may currently represent the best available proxy measure in large epidemiological studies however it is important to note that the choices made when selecting a buffer and dataset are rarely challenged rigorously. This study therefore aims to assess any change in strength of associations between food outlets and body-mass-index by using different databases and 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](#_ENREF_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 database is pre-coded into different categories and classes of commercial services.[8](#_ENREF_8) Rotherham LA provided their ‘current’ environmental health food outlet records for temporal comparison. Food outlets were then categorised 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. Three commonly used definitions of neighbourhood exposure based on previous research 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 LSOA (km2) was computed. LSOA sizes 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 (km2). IMD (Index of Multiple Deprivation) scores were assigned to the lower super-output area (LSOA) of each individual, as determined by their geocoded postcode.

A multi-level modelling (MLM) framework accounts for the hierarchical data structure with people nested within areas. Linear MLMs (β and 95% confidence intervals (CI)) were used to identify how different buffers and datasets are associated with BMI whilst adjusting for both individual- and neighbourhood-level factors. Level 2 residuals (lower super-output area) were distributed normally. Age, gender, ethnicity, rural or urban status (local government classification) and area level socio-economic status (Indices of Multiple Deprivation (IMD)) were included in all analyses as covariates. Differences in the magnitudes of associations were then assessed across different data sources and buffers by assessing the change in (β and 95% CI) (Appendix 1). All statistical analysis will be performed using STATA IC version 14.

Our results show no differences in the strength or direction of associations between food outlet availability when using different databases or neighbourhood definitions. A small difference was noted for supermarkets which differed only within the PoI dataset when comparing radial buffers to LSOA administratively defined neighbourhoods. The LA database contained approximately twice as many food outlet records as the Point of Interest (PoI) dataset; more specifically an additional 8 supermarkets, 23 takeaways and 589 other food retail outlets (Table 1). Exposure to food outlets also varied; within an 800m radial buffer LA data showed that some individuals had no fast-food outlets, whilst the average had 1.48±2.04 and the maximum was 23.00 fast-food outlets. Associations were substantively the same for all associations for instance, across the two different databases, LSOA density (km2) demonstrated no relationship between takeaway availability and BMI (POI (0.013 [-0.056; 0.083]; LA (0.013 [-0.014; 0.040]). It seems that despite some differences in the count of outlets, associations between food outlets and obesity are consistent. The magnitude of change in coefficients was also negligible when comparing administratively and radially defined buffers across all food availability variables but supermarkets (Table 1).

Our findings agree with previous evidence that suggests there is little change in size of associations across different definitions of neighbourhood[9](#_ENREF_9) and different datasets, despite the exception of supermarkets. The PoI database contained eight fewer supermarkets however differences were seen across neighbourhood definitions and databases in strength and direction of associations. This may suggest such database inaccuracies and differences in neighbourhood definitions for supermarkets should not be overlooked. In contrast, other evidence suggests neighbourhood definition may have significant implications on findings.[6](#_ENREF_6)[10](#_ENREF_10) Bodicoat et al. (2015) showed that fast-food outlets were weakly but positively associated with type II diabetes in smaller radial buffers (100m or 250m).[10](#_ENREF_10) 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) 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.[6](#_ENREF_6) Studies often use or only report associations within one neighbourhood definition. Although findings within this study suggest such differences are inconsequential, it is important researchers continue to consider and report the impact of such choices by reporting, where possible, a sensitivity analysis of different definitions of neighbourhood. This is especially true given that accuracy and therefore associations may differ by setting and that there is a dearth of evidence from which to compare this study with.

This study contributes to the research in two ways. Firstly, the association between food outlet availability and obesity was assessed using different definitions of neighbourhood. Secondly, this paper examined the extent to which using different databases 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. Radial buffers have been proposed as an alternative to administrative boundaries to represent an individual’s actual neighbourhood. However, there remains no uniform buffer across 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 be argued that administrative areas continue to inform local level policy best. The choices that researchers make throughout the research process may change the strength and direction associations seen. However, this study suggests that different databases and definitions of neighbourhood are broadly inconsequential in changing statistical inference.[6](#_ENREF_6)

The uncertainty around accurately defining neighbourhood and the use of different datasets remains a complex issue for contemporary environment based research. Results demonstrate little heterogeneity in magnitude of associations using different definitions of neighbourhood. Future papers should continue to explore the accuracy of these datasets and extend analyses to assess associations with health and how these may be modified by factors such as age, gender, ethnicity or deprivation. Furthermore, research may also explore additional definitions of neighbourhood such as proximity, street network buffers, self-defined buffers or GPS defined activity spaces. [5](#_ENREF_5) Finally, results may also differ for the physical activity environment, especially since numerous sources of data are available for the location of parks or green space.

Ultimately, it may be difficult to an accurate and standardised definition of neighbourhood within the environment, particularly given the nature of individual behaviours. It is therefore important to continue to rigorously challenge the choices made at a methodological level, but also to consider the usability of such measures for the development and implementation of local level policy. At present researchers should use the local context and problem being investigated to inform the most appropriate definition of availability and data source used. That is until better evidence emerges suggesting any different. Contrary to expectations, this research demonstrates few differences in the strength of associations between food outlet availability and obesity across both different neighbourhood definitions and databases.

1. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* 2014;384(9945):766-81.

2. 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.

3. 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.

4. 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.

5. 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.

6. 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.

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;121:1-6.

8. Ordnance Survey. *Points of Interest database - user guide and technical specification*. Southampton: Ordnance Survey, 2012.

9. Forsyth A, Hearst M, Oakes JM, Schmitz KH. Design and Destinations: Factors Influencing Walking and Total Physical Activity. *Urban Studies* 2008;45(9):1973-96.

10. 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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Data Source** | **Count (n)** | **LSOA Density (km2)** | **800m Buffer** | **2000m buffer** |
| B | 95% CI | B | 95% CI | B | 95% CI |
| **Local Authority (n=1,489)** | All food outlets (n=1,489) | -0.003 | [-0.010, 0.005] | -0.001 | [-0.018; 0.014] | 0.001 | [-0.003; 0.004] |
| Takeaways (n=257) | -0.001 | [-0.035, 0.033] | 0.013 | [-0.056; 0.083] | 0.013 | [-0.014; 0.040] |
| 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 | [0.09; 0.339] |
|  |  | **Mean(SD),Max**+ | **Mean(SD),Max**+ | **Mean(SD),Max**+ |
| **Local Authority****(n=1,489)** | All food outlets (n=1,489)Takeaway (n=257)Other retail (n=1,172)Supermarkets (n=60) | 12.28(17.55),125.002.21(4.08),20.839.73(13.72),104.610.33(0.81),6.25 | 7.55(8.72),160.001.48(2.04),23.005.80(6.97),135.000.27(0.49),3.00 | 38.46(34.66),244.007.03(5.59),33.0030.22(29.08),204.001.21(1.32),7.00 |
|  |
| **Point of Interest (n=869)** | All food outlets (n=869) | 7.49(14.22),94.08 | 4.86(8.25),114.001.43(2.62),33.003.18(5.89),81.000.25(0.51),4.00 | 24.69(26.80),170.006.78(7.16),44.0016.72(19.79),125.001.18(1.11),5.00 |
| Takeaways (n=234) | 1.96(4.01),29.17 |
| Other retail (n=583) | 5.19(10.45),68.42 |
| Supermarkets (n=52) | 0.35(0.97),6.90 |
| 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. |

**Table 1.** The change in magnitude of association between the environment and BMI by buffer size and data source

**Appendix 1** – Example of random intercept (fixed slopes) multi-level regression model

|  |  |
| --- | --- |
|  | Takeaway 2000m Buffer |
| **Fixed effects** | B [95% CI] | P |
| B (Intercept) | 25.72 [24.85; 26.60] | <0.00 |
| **Availability**  |  |  |
|  Count within 2km radial buffer | -0.002 [-0.023; 0.019] | =0.849 |
| **IMD** | 0.025 [0.015; 0.035] | <0.00 |
| **Gender**  |  |  |
|  Female | -0.354 [-0.631; -0.078] | <0.00 |
| **LSOA Classification**  |  |  |
|  Urban | -0.878 [-1.35; -0.405] | <0.00 |
| **Ethnicity**  |  |  |
|  Non-white | -0.086 [-1.16; 0.990] | 0.876 |
| **Age**  | 0.029 [0.020; 0.037] | <0.00 |
| *Random effects* |  |  |
|  Level-2 variance (u0j) Level-1 variance (u0i) | 2.02e-07 [3.62e-11;0.001]22.88 [21.98; 23.83] |
| Note: all models control for gender, ethnicity, deprivation (Index of Multiple Deprivation (IMD)) and rural/urban classification of the neighbourhood. \* = significant (p<0.05) |