Examining the validity and utility of two secondary sources of food environment data against street audits in England

Emma L Wilkins[a]*, Duncan Radley[a], Michelle A Morris[b] and Claire Griffiths[a]

[a]*Carnegie, Leeds Beckett University, Headingley Campus, Headingley, Leeds LS6 3QS, UK*

[b] *Leeds Institute for Data Analytics, School of Medicine, University of Leeds, Leeds LS2 9JT, UK.*

**Corresponding author*

e.wilkins@leedsbeckett.ac.uk (Emma Wilkins); duncan.radley@leedsbeckett.ac.uk (Duncan Radley); m.morris@leeds.ac.uk (Michelle Morris); c.griffiths@leedsbeckett.ac.uk (Claire Griffiths)

Abstract

Background: Secondary data containing the locations of food outlets is increasingly used in nutrition and obesity research and policy. However, evidence evaluating these data is limited. This study validates two sources of secondary food environment data: Ordnance Survey Points of Interest data (POI) and food hygiene data from the Food Standards Agency (FSA), against street audits in England and appraises the utility of these data.

Methods: Audits were conducted across 52 Lower Super Output Areas in England. All streets within each Lower Super Output Area were covered to identify the name and street address of all food outlets therein. Audit-identified outlets were matched to outlets in the POI and FSA data to identify true positives (TP: outlets in both the audits and the POI/FSA data), false positives (FP: outlets in the POI/FSA data only) and false negatives (FN: outlets in the audits only). Agreement was assessed using positive predictive values (PPV: TP/(TP+FP)) and sensitivities (TP/(TP+FN)). Variations in sensitivities and PPVs across environment and outlet types were assessed using multi-level logistic regression. Proprietary classifications within the POI data were additionally used to classify outlets, and agreement between audit-derived and POI-derived classifications was assessed.

Results: Street audits identified 1172 outlets, compared to 1100 and 1082 for POI and FSA respectively. PPVs were statistically significantly higher for FSA (0.91, CI: 0.89-0.93) than for POI (0.86, CI: 0.84-0.88). However, sensitivity values were not different between the two datasets. Sensitivity and PPVs varied across outlet types for both datasets. Without accounting for this, POI had statistically significantly better PPVs in rural and affluent areas. After accounting for variability across outlet types, FSA had statistically significantly better sensitivity in rural areas and worse sensitivity in rural middle affluence areas (relative to deprived). Audit-derived and POI-derived classifications exhibited substantial agreement ($p < 0.001$; Kappa = 0.66, CI: 0.63 - 0.70).

Conclusions: POI and FSA data have good agreement with street audits; although both datasets had geographic biases which may need to be accounted for in analyses. Use of POI proprietary classifications is an accurate method for classifying outlets, providing time savings compared to manual classification of outlets.

Keywords:

'Retail food environment', validity, 'street audit', foodscape, 'secondary data', 'obesogenic environments', sensitivity, 'positive predictive value', 'administrative data', 'commercial business list'

Examining the validity and utility of two secondary sources of food environment data against street audits in England

Background

Policymakers are increasingly recognising the role of the environment in driving obesity and associated health outcomes [1-3]. The 'retail food environment', characterised by the number, location and accessibility of food outlets within local environments, has been repeatedly targeted as a lever to tackle obesity [4-7]. However, evidence supporting these interventions is mixed, and predominantly null [8].

Research investigating the links between the retail food environment and obesity-related outcomes commonly uses data on food outlet locations to measure food access [9]. Access is measured using numerous spatial metrics such as density or proximity, with the majority of research investigating access to certain *types* of food outlet (e.g. 'fast food outlets' or 'supermarkets') hypothesised to have either a positive or negative effect on diet or weight status. Data on food outlet locations can be obtained through street audits; however, for efficiency reasons, it is more commonly obtained from secondary sources. The validity of these secondary data is an important consideration, repeatedly noted by authors as a limitation of these study designs [10-12]. Poor quality data can lead to uncertainty, bias, and reduced statistical power; potentially helping explain the mixed and predominantly null findings in retail food environment-obesity research. Indeed, a recent study found that the use of different data sources (from InfoUSA, and Dunn and Bradstreet) led to differences in both the strength and number of statistically significant associations between food outlet density and area-level demographics [13].

Recently, there has been increasing interest in the validity of secondary food environment data, which is typically assessed against the 'gold standard' of street audits [14, 15]. The vast majority of research originates from the US, wherein validity has been found to vary between different data sources, and across outlet types and environmental characteristics (e.g. deprivation and urbanicity) [14]. Overall, the percentage of food outlets captured in various US data sources has been found to range from 38% to 98% [16]. However, relatively little evidence exists in relation to the validity of UK-specific data.

Two very commonly used data sources in UK research are Ordnance Survey Points of Interest data (‘POI data’) [10-12, 17-19] and food hygiene data from local authorities [20-27]. Food hygiene data are collected by the Environmental Health department of each authority and comprise locational and business type information for all businesses engaged in ‘food operations’ (i.e. selling, cooking, storing, handling, preparing or distributing food/drink). Food hygiene data are often presented as a valid representation of the UK foodscape [24, 26, 28]; although these data have only been validated in three studies [29-31], which had relatively small sample sizes (ranging from 19 to 617) and limited geographic scope, restricting generalisability to the UK as a whole. In particular, two of these studies [29, 31] validated data within only one local authority (Newcastle and Glasgow respectively), and the third [30] validated data within three local authorities (Northumberland, Sunderland and Durham), but the audits only spanned 6 small sample areas. Given that food hygiene data is collected independently by local authorities, data quality may vary across authorities. Additionally, there is evidence that the validity of food environment data from other countries may vary across urban/rural and socioeconomic contexts [14, 16]. Geographic context is therefore important in establishing the validity of food hygiene data, and further investigation is needed across a broader range of contexts.

Historically, food hygiene data had to be requested separately for each local authority [32]. However, these data are now available centrally for all UK local authorities via the Food Standards Agency (FSA) website [33]. Personal communications with environmental health officers have indicated that there may be some differences between data obtained from the FSA and data obtained directly from local authorities (e.g. in relation to the scope of the data) meaning the validity of data obtained from the FSA website (hereinafter referred to as ‘FSA data’ to distinguish from ‘local authority data’ obtained directly from local authorities) may differ from that obtained directly from local authorities. While the food outlet data on the FSA website is updated daily, it is unclear how regularly local authorities update their own records, which would impact the validity of both the FSA and local authority data. In view of the above, validation of FSA data is needed.

POI data contains locational and classification information on over 4 million points of interest (e.g. businesses and public facilities) across the UK [34]. As well as being prominent in research, it is also used in emerging policy tools, such as the Food Environment Assessment Tool (FEAT) and the Public Health England fast food map [35, 36]. However, it has only been evaluated in one study [28], which was of limited geographic scope, and did not compare the data to the 'gold standard' of street audits. Thus, validation of this important dataset over a broader geographic scope, and against street audits is needed. Validation of both FSA and POI data against the same street audit data will also enable comparison between these two important datasets.

The aim of this study is to validate POI and FSA data against street audits in England. A first objective is to establish the overall agreement between the audits and the POI and FSA data respectively. As the validity of US data sources has been found to vary across outlet types and environmental characteristics a second objective is to determine whether the agreement of the POI or FSA data varies across different environment types (characterised by deprivation and urbanicity) or outlet types. As POI data includes detailed proprietary outlet classifications that have been previously used to define outlet types [10], a third aim is to establish the accuracy of POI-derived outlet classifications relative to audit-derived classifications. Finally, insights into the utility of the data are presented in order to help researchers and policymakers make a fully-informed decision around which (if any) of the two data sources to use.

Methodology

Audit area selection

Audit areas were selected from within four local authorities in England: Leeds (having a range of urban areas with a spread of deprivation levels), Durham (having a range of rural, deprived areas), North Kesteven and Calderdale (both having a range of rural areas of middle/high affluence). There are 327 local authorities in England. Lower Super Output Area (LSOA) boundaries were used to define audit areas. LSOAs are an administrative geography in the UK with a minimum population of 1000 [37]. LSOA boundary data was obtained from the UK Data Service [38].

LSOAs were selected across six environment types: ‘urban deprived’, ‘urban middle affluence’, ‘urban affluent’, ‘rural deprived’, ‘rural middle affluence’ and ‘rural affluent’. Urban/rural designations were applied using Office for National Statistics Rural Urban Classifications at the LSOA level [38] as defined in Table 1. Deprivation designations were applied based on English Index of Multiple Deprivation (IMD) rankings [39]. As the degree of deprivation in England is not evenly distributed across urban and rural areas (e.g. only 0.8% of rural LSOAs, versus 12.0% of urban LSOAs are within the lowest decile of deprivation), LSOAs were stratified by urban/rural designation, and were re-ranked for deprivation relative to all other LSOAs with the same rural/urban classification (supplementary materials). For urban and rural areas separately, the new deprivation rankings were divided into deciles, and environment designations were applied (see Table 1).

Table 1

Definitions of the Six Environment Types

Environment Type	IMD Deciles*	Rural/Urban Classifications
Urban Affluent	urban IMD deciles 8-10	A1, B1, C1, C2
Urban Middle Affluence	urban IMD deciles 4-7	A1, B1, C1, C2
Urban Deprived	urban IMD deciles 1-3	A1, B1, C1, C2
Rural Affluent	rural IMD deciles 8-10	D1, D2, E1, E2
Rural Middle Affluence	rural IMD deciles 4-7	D1, D2, E1, E2
Rural Deprived	rural IMD deciles 1-3	D1, D2, E1, E2

Note. A1: Urban major conurbation; B1: Urban minor conurbation; C1: Urban city and town; C2: Urban city and town in a sparse setting; D1: Rural town and fringe; D2: Rural town and fringe in a sparse setting; E1: Rural village and dispersed; E2: Rural village and dispersed in a sparse setting; IMD: Index of multiple deprivation.

*IMD deciles were calculated separately for urban and rural environments as described in the main text.

LSOAs were selected for auditing based on the ease with which they could be reached by the audit team and the number of expected outlets within each LSOA, as indicated by the POI data; with higher numbers chosen preferentially. LSOAs were selected to ensure at least 100 food outlets were expected within each of the six environment types (e.g. ‘rural deprived’). All LSOAs were eligible for selection. Overall, 52 LSOAs were selected for auditing (supplementary materials).

Street Audits

The boundaries of the selected LSOAs were copied by hand onto printed street maps [40-42] to define audit areas. Some small modifications were made to the LSOA boundaries for practicality reasons (see supplementary materials for details). All streets falling within each audit area were walked and the name, street name, and outlet classification of all food outlets were recorded, forming an 'Audit List' of food outlets. Food outlets within private premises (e.g. members' clubs or workplaces) or outlets not visible from the roadside (e.g. cafes within hospitals or sports centres) were not recorded.

Outlets were designated one of seven outlet types ('Restaurant', 'Pub', 'Cafe', 'Fast Food', 'Supermarket', 'Convenience', and 'Speciality') as defined based on the classification scheme of Lake et al. [29] (supplementary materials). All audits were performed by one of two teams of trained auditors and took place in September and October 2016. To assess inter-rater agreement, four LSOAs were audited independently by both sets of auditors.

Secondary data

The most recent version of POI data available at the time of the street audits was downloaded from Edina Digimap (Leeds: March 2016 version [43]; all other areas: June 2016 version [44]). The FSA data was downloaded from the Food Standards Agency website [33] on 8th December 2016. A flow chart detailing data processing steps in respect of these data is shown in Figure 1.

Figure 1 here

Firstly, food outlets with the proprietary classification codes listed in Table 2 were extracted from each dataset (POI: n = 29,586; FSA: n = 8,976; full classification schemes for each dataset available in supplementary materials). The two datasets were then screened for missing coordinate data and/or address data. Entries missing both coordinate and address data (FSA: n = 99; POI: n = 0) were deleted, and those missing coordinate data only (FSA: n = 82; POI: n = 0) were inspected to establish whether the address fell within an audit area (FSA: n = 3). The remaining food outlets were plotted in ArcMap 10.4 using their associated coordinate data to identify outlets falling within the audit areas. This generated a list of expected outlets (the 'Expected Outlets List') for the POI and FSA data respectively.

Table 2

POI and FSA Classification Codes Used to Extract Food Outlets from the Original Dataset

POI Classification Codes (Classification Name)	FSA Classification Names*
1020013 (cafés, snack bars and tea rooms)	"Pub/Club"
1020025 (internet cafés)	"Restaurant/Café/Canteen"
9470699 (convenience stores and independent supermarkets)	"Retailers – Supermarkets/Hypermarkets"
10540737 (petrol and fuel stations)	"Retailers – Smaller "Retailers"**
1020043 (restaurants)	"Retailers – Other"
1020034 (pubs)	"Takeaway"
1010006 (hotels, motels, country houses and inns)	"Primary Producer"
9470662 (butchers)	"Distributors/Transporters"
9470665 (delicatessens)	"Manufacturers/Packers"
9470666 (fishmongers)	"Hotel /Guest House"
9470668(green and new age goods)	
9470669 (grocers, farm shops and pick your own)	
9470670 (herbs and spices)	
9470672 (organic, health, gourmet and kosher foods)	
7400524 (baking and confectionery)	
9470663 (confectioners)	
9470819 (supermarkets)	
9470667 (frozen foods)	
1020018 (fast food and takeaway outlets)	
1020019 (fast food delivery services)	
1020020 (fish and chip shops)	
9470661 (bakeries)	
4250312 (nightclubs)	
9470705 (markets)	

Notes. *Classification names listed are the official classifications as provided in the local authority Enforcement Monitoring System documentation [52]. These names deviate slightly from the actual classification names applied to the data used in the present study, as detailed in the Supplementary Materials.

**The 'Retailers – Smaller Retailers' classification is listed for completeness. However, for the data included in the present study, no food outlets had been classified within this category, with the 'Retailers – other' category appearing to be applied instead.

Data Matching

In order to assess agreement between the audits and the POI and FSA data, entries within the Expected Outlets List for the POI and FSA data respectively were compared to the Audit List to identify matches. Matches were coded as true positives. All un-matched outlets within the Expected Outlet Lists were coded as false positives and all un-matched outlets within the Audit List were coded as false negatives.

Two separate matching criteria were utilised; referred to herein as 'strict' and 'relaxed' criteria, both mirroring matching criteria that have been employed in previous validation studies [28, 45, 46]. Under the strict matching criteria, matches were established if outlet names and street names were the same or similar. Naming discrepancies were allowed if they were grammatical e.g. 'The Cod Father' and 'The Codfather' or when the names and classifications were substantially similar (e.g. 'Magic Wok' and 'Mr Wong's Magic Wok', both classified as 'Restaurant'). Discrepancies in street name were allowed if an outlet was located at a junction (and could therefore have multiple legitimate street addresses) or if the outlet was on a street having multiple names (e.g. 'Armley Road' merging into 'Canal Street', supplementary materials). The 'strict' criteria are relevant to study designs that utilise store names in analyses e.g. to extract food outlets. However, typically retail food environment research investigates access to certain *types* of food outlets (e.g. 'fast food outlets'), and for much of this research, outlet *names* are inconsequential. Thus, under the 'relaxed' matching criteria, outlet names were allowed to differ, and a match was instead required between outlet classifications and street names. Thus, outlets that had different names e.g. 'Eastern Delight' and 'Double Dragon', but the same outlet classification ('Fast Food'), and were located on the same street were considered a match.

After data matching, the entries were manually screened to identify and subsequently remove duplicates (additional details in supplementary materials). For the POI data, 111 entries (8.9%) were removed as duplicates. For the FSA data, 8 entries (0.6%) were removed as duplicates.

Entries coded as false positives were additionally examined to assign one of the seven outlet classifications defined above, using a combination of the outlet's proprietary classification, outlet name, and Google searching. Outlets falling outside the seven classifications additionally fell outside the scope of the street audits (e.g. childcare centres and workplace canteens), and were classified as 'other' and excluded. For the POI data, 35 entries (2.8%) were determined to be 'other'-type outlets, compared to 158 (12.5%) for the FSA data. It was possible to assign a classification to all false positive entries in the POI data. However, 14 (1.1%) of the outlets in the FSA data were unclassifiable because the businesses could not be identified online. These outlets were also excluded.

Agreement between POI-derived and audit-derived classifications

As mentioned above, the POI data includes very detailed proprietary outlet classifications, which have been used to define outlet types in research. This process was simulated in this study, with 'POI-derived' classifications being defined as shown in Table 3. These classifications were applied to all true positives, to allow comparison with the audit-derived classifications. Agreement between FSA classifications and audit-derived classifications was not assessed because the proprietary classifications in the FSA data lacked sufficient detail for comparison with the audit classifications.

Table 3

POI-derived classification scheme

Classification Name	POI Codes
Restaurant	1020043 (restaurants) 1020034 (pubs – manual Google search to identify those serving food) 1010006 (hotels, motels, country houses and inns)
Pub	1020034 (pubs) 4250312 (nightclubs)
Café	1020013 (cafés, snack bars and tea rooms) 1020025 (internet cafés)
Fast Food	1020018 (fast food and takeaway outlets) 1020019 (fast food delivery services) 1020020 (fish and chip shops) 9470661 (bakeries)
Supermarket	9470699 (convenience stores and independent supermarkets)* 9470819 (supermarkets) 9470667 (frozen foods)
Convenience	9470699 (convenience stores and independent supermarkets)* 10540737 (petrol and fuel stations)
Specialty	9470662 (butchers) 9470665 (delicatessens) 9470666 (fishmongers) 9470668 (green and new age goods) 9470669 (grocers, farm shops and pick your own), 9470670 (herbs and spices) 9470672 (organic, health, gourmet and kosher foods), 7400524 (baking and confectionery) 9470663 (confectioners)

Note. POI: Points of Interest data.

*Outlets with this classification were coded as ‘supermarket’ if they were a small format major national chain supermarket (Tesco Express, Sainsbury's Local, M&S Simply Food, Little Waitrose and Co-operative). Otherwise, the outlets were classified as convenience stores.

Statistical Analyses

All statistical analyses were conducted in R (v 3.2.3). The threshold for statistical significance was set at $p < 0.05$. All results presented are for 'relaxed' matching criteria (requiring a match on outlet classifications and street addresses, but not outlet names as described above), unless expressly stated.

Inter-rater agreement was assessed by comparing counts of outlets identified in the audit areas. Percentage agreement and the Kappa statistic were used to assess agreement between broad outlet classifications.

Traditional measures of agreement for categorical data (e.g. the Kappa statistic) cannot be used to assess agreement with the street audits, because the number of 'true negatives' (i.e. outlets found neither in the audits nor the secondary data) is undefined. Agreement between the secondary datasets (POI and FSA) and the audits was therefore assessed via sensitivity statistics and positive predictive values (PPV); defined as shown in Figure 2. Sensitivity statistics indicate the prevalence of missing outlets within the POI and FSA data, whereas PPV statistics indicate the prevalence of 'erroneous' food outlets within these data. Clopper-Pearson 'exact' 95% confidence intervals (CI) were calculated for sensitivities and PPVs [47].

Figure 2 here

To assess variation in agreement across environment and outlet types, PPVs and sensitivities were modelled using separate respective random intercepts multi-level logit models to account for the multi-level sampling approach used in this study (outlets nested within LSOAs). PPVs and sensitivities were treated as respective binary outcomes (sensitivity: true positive vs false negative; PPV: true positive vs false positive). Thus, the resultant odds derived from these models can be interpreted as indicating the odds of an outlet listed in the secondary dataset being a true positive versus a false positive (PPV odds) and the odds of an outlet found in the audits being a true positive versus a false negative (sensitivity odds) (Figure 2).

A series of models were run to estimate the associations between urbanicity, deprivation and outlet type, and PPVs and sensitivities. In Model 1, urbanicity was included as a single fixed effect to determine whether PPVs or sensitivities vary across urban/rural environments. In Model 2, urbanity was replaced with deprivation, to explore variation in PPVs or sensitivities across deprivation levels.

Variability in data quality across environment types may be explained by inherent geographic biases. However, it may also be explained by variation in data quality across outlet types, and differing food outlet composition across environment types (e.g. if fast food outlets have high PPVs/sensitivities then areas with higher concentrations of fast food outlets, such as deprived urban areas, will appear to have higher PPVs/sensitivities). To explore whether differing food outlet composition explains any observed geographic biases, Model 3 included urbanicity, deprivation and outlet type as fixed effects in a single model. An interaction between urbanicity and deprivation was also included to account for the dependency of deprivation on urbanicity.

Agreement between audit-derived and POI-derived classifications was compared using percentage agreement and Cohen's Kappa statistic.

Results

Inter-rater agreement

Across the four LSOAs audited by both audit teams, the first identified 115 outlets and the second identified 109 (88.2% agreement). Percentage agreement for outlet classifications was 88.6%, and Kappa agreement was 0.86 (CI: 0.78 - 0.94), which is considered 'almost perfect' according to Landis and Koch [48].

Overall agreement with audits

Counts of outlets

Overall, 1172 outlets were identified in the street audits, compared to 1100 and 1082 in the POI and FSA data respectively (Table 4). Both datasets under-represented the total count of food outlets across most environment and outlet types compared to the street audits. As exceptions to this, the count of outlets in middle deprived areas was equal in the audits and POI data. Additionally, pubs were over-represented in both the POI and FSA datasets (9.5% and 4.8% respectively), and supermarkets were over-represented by the POI dataset (8.6%). Counts of outlets across each local authority and LSOA are reported in supplementary materials. Counts of outlets identified in the audits ranged from 1 to 176 at the LSOA level, and from 73 to 795 at the local authority level.

Table 4
Counts of outlets and corresponding positive predictive values and sensitivities

Environment/	Audits	POI			FSA		
Outlet Type	Count	Count	PPV	Sens	Count	PPV	Sens
Total	1172	1100	0.86	0.81	1082	0.91	0.84
Urban	742	729	0.83	0.82	680	0.91	0.83
Deprived	249	244	0.83	0.81	225	0.91	0.82
Middle	342	344	0.81	0.81	319	0.90	0.84
Affluent	151	141	0.91	0.85	136	0.92	0.83
Rural	430	371	0.91	0.78	402	0.92	0.86
Deprived	173	161	0.86	0.80	172	0.91	0.91
Middle	135	114	0.93	0.79	122	0.91	0.82
Affluent	122	96	0.97	0.76	108	0.95	0.84
Restaurant	306	288	0.91	0.86	283	0.95	0.88
Pub	63	69	0.65	0.71	66	0.73	0.76
Café	194	152	0.87	0.68	175	0.89	0.80
Fast Food	299	299	0.87	0.87	280	0.96	0.90
Supermarket	81	88	0.82	0.89	76	0.97	0.91
Convenience	115	103	0.83	0.75	111	0.80	0.77
Specialist	114	101	0.86	0.76	91	0.92	0.74

Note. Sens: sensitivity. PPV: positive predictive value. POI: Points of Interest. FSA: Food Standards Agency.

PPV and Sensitivities

Overall, the PPV was statistically significantly higher for FSA data (0.91, 95% confidence interval (CI): 0.89-0.93) than for POI data (0.86, CI: 0.84-0.88, $p < 0.05$, Figure 3). There was no statistically significant difference in sensitivity values between the two datasets (POI: 0.81, CI: 0.78-0.83; FSA: 0.84, CI: 0.82-0.86). Both the FSA and POI data had 'good' agreement with street audits according to the classification system of Paquet et al. [49].

Figure 3 here

When strict matching criteria were applied (i.e. requiring a match based on outlet name), PPV and sensitivity values were lower than under the relaxed matching criteria (POI: PPV: 0.79, CI: 0.77 – 0.82; sensitivity: 0.74, CI: 0.72 – 0.77; FSA: PPV: 0.87, CI: 0.85 – 0.89; sensitivity: 0.81, CI: 0.78 – 0.83).

Variation by environment and outlet type

POI data

For the POI data, PPV odds varied statistically significantly across deprivation and urbanicity. In rural areas, the odds of an outlet listed in the POI data being present in reality (a 'true outlet') were 2.07 (1.18 – 4.02) times higher than in urban areas (Table 5). The odds were also 2.63 (1.34 - 5.43) higher in affluent areas compared to deprived areas. However, after controlling for variability in validity across outlet types, neither deprivation nor urbanicity bias remained. PPV odds varied significantly across outlet types, and were statistically significantly lower for pubs, supermarkets and convenience stores relative to restaurants.

Sensitivity odds did not vary across deprivation or urbanicity, even after controlling for variability in food outlet composition across areas (Table 6). However, sensitivity odds varied significantly across outlet types and were statistically significantly lower for pubs, cafes, convenience stores and speciality outlets relative to restaurants.

Findings were similar for the strict matching criteria, except that, for PPV odds there was a very small, but statistically significant urban/rural bias, with the odds of an outlet listed within the POI dataset being a 'true outlet' 1.69 (1.00 2.92) times higher in rural than in urban areas (supplementary materials). The PPV and sensitivity odds were also less variable, with supermarkets no longer statistically significantly different from restaurants for PPV odds and pubs and speciality stores no longer statistically significantly different for sensitivity odds.

Table 5

Odds of true positive relative to false positive (PPV odds) for POI data

Environment/	Model 1			Model 2			Model 3		
Outlet Type	OR	95% CI		OR	95% CI		OR	95% CI	
Urban	REF						REF		
Rural	**2.07**[1]	**1.18**	**4.02**				1.31	0.69	2.61
Deprived				REF			REF		
Middle				1.08	0.58	2.05	0.78	0.39	1.41
Affluent				**2.63**[3]	**1.34**	**5.43**	1.80	0.85	3.81
Restaurant							REF		
Pub							**0.19**[3]	**0.09**	**0.37**
Café							0.67	0.36	1.28
Fast Food							0.66	0.37	1.16
Supermarket							**0.42**[1]	**0.21**	**0.88**
Convenience							**0.39**[2]	**0.19**	**0.80**
Speciality							0.56	0.27	1.21
Rural*Middle							2.69	0.89	8.40
Rural*Affluent							2.71	0.68	13.81

Note. OR: Odds ratio. CI: Confidence interval. REF: Reference category. All models are multi-level models accounting for nesting of outlets within LSOAs. [1]p<0.05, [2]p<0.01, [3]p<0.001

Table 6

Odds of true positive relative to false negative (sensitivity odds) for POI data

Environment/	Model 1			Model 2			Model 3		
Outlet Type	OR	95% CI		OR	95% CI		OR	95% CI	
Urban	REF						REF		
Rural	0.80	0.59	1.08				0.97	0.58	1.61
Deprived				REF			REF		
Middle				1.00	0.71	1.39	1.07	0.67	1.70
Affluent				1.02	0.70	1.51	1.31	0.75	2.34
Restaurant							REF		
Pub							**0.42**[2]	**0.22**	**0.81**
Café							**0.36**[3]	**0.23**	**0.56**
Fast Food							1.16	0.72	1.89
Supermarket							1.37	0.65	3.15
Convenience							**0.52**[1]	**0.30**	**0.90**
Speciality							**0.55**[1]	**0.32**	**0.98**
Rural*Middle							0.91	0.43	1.89
Rural*Affluent							0.60	0.27	1.33

Note. OR: Odds ratio. CI: Confidence interval. REF: Reference category. All models are multi-level models accounting for nesting of outlets within LSOAs. [1]$p<0.05$, [2]$p<0.01$, [3]$p<0.001$

FSA data

For the FSA data, there was no variability in PPV odds across urbanicity or deprivation, even after controlling for variability in food outlet composition across environment types (Table 7). There were, however, statistically significant variations in PPV odds across outlet types, with the odds of an outlet listed in the POI data being a ‘true outlet’ markedly lower for pubs, cafes, and convenience stores relative to restaurants.

In relation to sensitivity odds, Models 1 and 2 found no association with deprivation or urbanicity (Table 8). However, controlling for variability in sensitivity values across outlet types revealed a statistically significant urban/rural bias. Moreover, there was a significant interaction between deprivation and urbanicity, which after stratification of the data based on urbanicity revealed a statistically significant deprivation bias in rural areas. More particularly, the odds of a ‘true outlet’ being listed in the FSA data were 2.23 (CI: 1.21-4.28) times higher in rural than in urban areas, and among rural areas, the odds were 0.49 (CI: 0.24-0.97) times lower in middle affluence than in deprived areas. There was statically significant variation in sensitivity odds across outlet types, with ‘true’ pubs, cafes, convenience stores and speciality stores having lower odds of being listed in the FSA data than restaurants. However, after stratification of the data based on urbanity, this outlet-type variability was only evident in urban areas. All findings for the FSA data were substantively the same for the strict matching criteria (supplementary materials).

Table 7

Odds of true positive relative to false positive (PPV odds) for FSA data

Environment/	Model 1			Model 2			Model 3		
Outlet Type	OR	95% CI		OR	95% CI		OR	95% CI	
Urban	REF						REF		
Rural	1.22	0.79	1.94				1.40	0.64	3.19
Deprived				REF			REF		
Middle				0.95	0.59	1.51	0.99	0.45	2.10
Affluent				1.42	0.78	2.69	1.23	0.53	2.94
Restaurant							REF		
Pub							**0.13**[3]	**0.06**	**0.28**
Café							**0.43**[1]	**0.20**	**0.88**
Fast Food							1.15	0.51	2.67
Supermarket							1.97	0.52	12.93
Convenience							**0.20**[3]	**0.09**	**0.42**
Speciality							0.62	0.24	1.76
Rural*Middle							0.90	0.27	3.00
Rural*Affluent							1.45	0.37	6.16

Note. OR: Odds ratio. CI: Confidence interval. REF: Reference category. All models are multi-level models accounting for nesting of outlets within LSOAs. [1]p<0.05, [2]p<0.01, [3]p<0.001 [a]Reference category for

Table 8

Odds of true positive relative to false negative (sensitivity odds) for FSA data

Environment/ Outlet Type	Model 1			Model 2			Model 3			Model 3 (urban only)			Model 3 (rural only)		
	OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI	
Urban	REF						REF								
Rural	1.27	0.86	1.88				**2.23[1]**	**1.21**	**4.28**						
Deprived				REF			REF			REF			REF		
Middle				0.87	0.53	1.38	1.09	0.66	1.87	1.03	0.53	1.96	**0.49[1]**	**0.24**	**0.97**
Affluent				0.85	0.52	1.36	1.01	0.57	1.77	0.94	0.49	1.77	0.60	0.29	1.26
Restaurant							REF			REF			REF		
Pub							**0.41[2]**	**0.21**	**0.82**	**0.24[3]**	**0.10**	**0.56**	1.26	0.37	5.85
Café							**0.56[1]**	**0.34**	**0.92**	**0.47[1]**	**0.25**	**0.89**	0.76	0.33	1.75
Fast Food							1.17	0.69	1.98	0.92	0.47	1.79	1.75	0.73	4.40
Supermarket							1.45	0.64	3.73	0.97	0.38	2.85	4.13	0.77	76.60
Convenience							**0.46[2]**	**0.26**	**0.83**	**0.36[2]**	**0.17**	**0.77**	0.68	0.27	1.75
Speciality							**0.38[3]**	**0.21**	**0.67**	**0.27[3]**	**0.13**	**0.56**	0.68	0.27	1.76
Rural*Middle							**0.43[1]**	**0.17**	**0.98**						
Rural*Affluent							0.53	0.21	1.30						

Note. OR: Odds ratio. CI: Confidence interval. REF: Reference category. All models are multi-level models accounting for nesting of outlets within LSOAs. [1]p<0.05, [2]p<0.01, [3]p<0.001.

Agreement between POI-derived and audit-derived classifications

POI-derived classifications agreed with audit-derived classifications 72.2% of the time (n = 871) (supplementary materials), exhibiting ‘substantial’ agreement ($p < 0.001$; Kappa = 0.66, CI: 0.63 - 0.70) [48].

Discussion

Secondary data on the food environment is commonly used in research and is also emergently used in policy tools [35, 36]. This study sought to validate two easily-accessible sources of UK-specific food environment data (POI and FSA) against the ‘gold-standard’ of street audits. Our key finding was that POI and FSA data both have 'good' agreement with street audits according to the classification system of Paquet et al. [49], providing policymakers with confidence in using research and tools based on these data.

The overall PPV was *statistically* significantly higher for the FSA data than the POI data for PPV (no difference for sensitivity). However, the magnitude of this difference is relatively small and may not substantively impact the validity of findings based on these data. Indeed, Hobbs et al. [50] compared the strength and direction of associations between food access and weight status when using POI and local authority data, and obtained similar findings for both datasets (12/12 versus 11/12 of the tested associations were null for the respective data sources).

This study used both ‘strict’ and ‘relaxed’ matching criteria, with the former requiring outlet names and street addresses to agree, and the latter being more lenient in allowing outlet names to differ, provided outlet classifications agreed. For the FSA data, agreement statistics were similar under the two matching criteria (albeit slightly lower under the strict matching). For the POI data, however, there was a more marked difference between the agreement statistics under the two matching criteria. This may indicate that the POI data are less up-to-date with changes in store *names* (but not *function*) than the FSA data. For most research, relaxed matching criteria provide the most appropriate indication of the validity of the data, because typically only the classification of an outlet is of importance, and the outlet name is not considered when deriving food access measures.

This is the first study to assess the validity of food hygiene data from the FSA. However, several studies have validated food hygiene data obtained directly from local authorities [29-31]. These found similar PPVs and sensitivities to those found in this study, with PPVs ranging from 0.79 – 0.92 and sensitivity values ranging from 0.60 – 0.95 [29-31]. This suggests that any differences in data management between the FSA and independent local authorities do not give rise to any substantive differences in data quality.

This is also the first study to assess the validity of POI data against the 'gold standard' of street audits. However Burgoine and Harrison [28] instead evaluated POI data against local authority data, finding a PPV of 0.75 and sensitivity of 0.60. Both values are lower than those found in the present study. It is likely that this discrepancy is due to the use of local authority data as the comparator to the POI data, rather than street audits as used in our study.

Several studies have investigated potential geographical biases in POI and local authority data [28, 30, 31]. However, these have either used small sample sizes, or have not compared the secondary data to the 'gold standard' of street audits, limiting the strength of their findings. Understanding geographic biases in data is important so that steps can be taken to avoid confounding; especially within the context of retail food environment research, which seeks to capture differences in the retail food environment across areas. This study found POI data to have statistically significantly higher PPV odds in rural and affluent areas (which can be interpreted as meaning that the likelihood of an outlet listed in the POI data being a 'true outlet' - i.e. one that exists in reality – is higher in rural than in urban areas). However, these geographic biases were entirely explained by differences in food outlet composition across these environment types. After accounting for variability in PPVs across outlet types, there was no evidence of a geographic bias. Thus, when POI data is used to study specific outlet types (e.g. fast food outlets only), geographic bias is unlikely. However, for food access metrics that consider multiple food outlet types together (e.g. fast food outlets divided by total food outlets) then geographic bias may exist.

Contrary to the present findings, Burgoine and Harrison [28] found no evidence of urban/rural bias in PPVs when comparing POI to local authority data, but did find statistically significantly lower sensitivities and percentage agreement in rural areas. However, as mentioned, Burgoine and Harrison used local authority data as reference data, and inaccuracies in the local authority data may have given rise to these different findings. Additionally, as the study area was limited to the relatively affluent and predominantly rural area of Cambridgeshire, there may have been insufficient variation in environmental characteristics to reliably detect geographic bias across the UK as a whole.

For the FSA data, there was no overall geographic bias in the PPV or sensitivity odds, which is in agreement with previous literature [30, 31]. However, after accounting for variability in agreement across outlet types, sensitivity odds were statistically significantly higher in rural than in urban areas (which can be interpreted as meaning that the likelihood of a ‘true outlet’ being listed in the FSA data is higher in rural than urban areas). Among rural areas, sensitivity odds were also lower in middle than deprived areas. This means, if FSA data is used to study specific food outlet types (as is often the case), the count of outlets may be under-estimated in urban areas relative to rural areas, and in middle affluence rural areas relative to deprived rural areas.

Many food environment studies investigate access to certain outlet types; most commonly supermarkets, convenience stores and fast food outlets [8]. Our study found that both POI and FSA data exhibited variation in both PPV and sensitivity odds across outlet types. Notably, PPV and sensitivity odds for convenience stores were low for both datasets. Low accuracy for convenience stores has also been noted in other international datasets [14], suggesting convenience store provision may be inherently difficult to capture. That said, PPVs and sensitivity values were still ‘good’ according to the classifications of Paquet et al. [49] for both datasets.

After stratifying by urbanicity, statistically significant variation in sensitivity values across the FSA data disappeared in rural environments. This is likely to be caused by smaller sample sizes within rural environments and an associated lack of power to detect significant variation across outlet types, rather than representing that sensitivity values are stable in rural environments but not in urban environments.

POI data includes approximately 24 different classification codes for food outlets, providing relatively detailed information on outlet function. The proprietary codes within the POI data have previously been used to define outlet types in research [10]. However, the accuracy with which outlets can be classified using these proprietary codes was unknown. Our study found that POI-derived classifications substantially agreed with audit-derived classifications, suggesting that use of proprietary classifications to automatically assign outlets to broad outlet classifications is a viable method for classifying outlets. This method is considerably more time-efficient than manually classifying each outlet e.g. based on Google searching, as has been carried out in other research [24, 25].

It should be noted that the reliance on outlet classifications to characterise the retail food environment is simplistic, and does not take into account food provision within individual outlets nor other factors that may influence purchasing decisions, such as pricing and preferences. However, capturing detailed features of the retail food environment such as these typically requires within-store audits, which are not practical for large-scale studies. Thus, while use of outlet classifications may not be the 'best' method for capturing the availability of foods within local environments, it presents a practical compromise for large-scale research.

Although FSA and POI data have been shown to be similarly valid, in our view the POI data has better utility. Firstly, POI data has more detailed proprietary outlet classifications than FSA data. It has been shown in our study that use of POI classifications to automatically assign outlets to broad outlet classifications is a viable method for classifying outlets. Conversely, for the majority of research, FSA classifications do not provide sufficient detail to characterise the retail food environment, and thus outlets must be classified via some other means e.g. use of business directories or Google searching, which is labour-intensive.

Secondly, the percentage of outlets that had to be removed from the FSA data was higher than for the POI data (14.3% vs 11.7%). Additionally, the majority of these (95.6%) were excluded as 'other'-type (e.g. childcare centres and workplace canteens) or unclassifiable outlets, which are not usually of interest in food access studies. Conversely only a relatively small percentage (24.0%) of outlets excluded from the POI data were 'other'-type outlets, with the remainder being duplicates. Screening for ‘other’-type outlets is thus very important for the FSA data, but less-so for the POI data. This screening process is very labour intensive, requiring all outlets to be manually classified using e.g. Google searching. Removal of duplicates from a dataset, on the other hand, is relatively simple and can be partially automated. Thus, data cleaning may be considerably more labour intensive for the FSA data.

Finally, POI data are more geographically accurate; with addresses geocoded to the address level (i.e. the precise building) [34], whereas FSA data are geocoded to the postcode level, which include multiple addresses (an average of 15 and a maximum of 100) [51]. This is illustrated in the fact that only one food outlet was missed from the POI Expected Outlets List due to a geocoding inaccuracy; whereas 16 were missed from the FSA Expected Outlets List. While it is possible to geocode the FSA data with better spatial accuracy using address look-ups, this requires additional time. Also, address information within the FSA data was sometimes missing or incomplete, meaning these addresses could not be geocoded to the address level.

Overall both datasets required considerable data cleaning. The total time taken to carry out this process was not recorded. Nevertheless, it was substantially less than the resource requirements of the street audits, which took 37 full working days and cost £555 in travel and accommodation costs, supporting the use of secondary data as an efficient means to characterise the retail food environment.

Strengths of this research included the relatively large sample sizes allowing variability in the validity of the data across outlet and environment types to be examined, and the use of ‘strict’ and ‘relaxed’ matching criteria which are applicable to different use cases that do and do not require accurate listings of outlet names. Further, in addition to data validity this study considered the utility of the data (i.e. in terms of the amount of data cleaning required, and the level of detail and accuracy of proprietary classifications); a factor that is influential in data selection.

Due to time restrictions, only four local authorities were covered in the audits. While this is an improvement over prior literature, our findings may still not be generalisable to all local authorities nationally. Additionally, as the FSA data are collected by independent local authorities, there may be variability in data quality across authorities. It is also possible (albeit less likely) that the quality of POI data varies across local authorities. To account for this, we considered including local authority as a fixed effect in our models. However, there was a high degree of correlation between local authority and urbanicity (due to the local authorities being predominantly either urban or rural, r = 0.84), which can lead to unstable parameter estimates [52]. We therefore chose to exclude local authority from our final model. We cannot rule out that the observed variations in data quality across urban and rural environments could also be explained by variations across local authorities.

Time and financial restrictions also meant that it was not possible to cover many ‘dispersed’ rural areas, with the majority of rural LSOAs (96.7%) being classified as ‘rural town and fringe’. Thus, results might not be generalisable to more dispersed rural environments.

Temporal mismatch between the street audits and date of acquisition of the POI and FSA data may have reduced agreement between these data and the street audits. However, the temporal mismatch was no more than 2 months, and the foodscape is unlikely to have changed substantially in this time. Additionally, temporal mismatch of this magnitude and more between exposure and outcome data is common in food access research [19, 23, 53, 54], so the present findings remain applicable to such research. It was not possible to obtain POI and FSA data from the same timeframe, and thus comparisons between the validity of the POI and FSA data may have been affected by temporal mismatch between these datasets.

Finally, the present study excluded food outlets whose primary function was not food retail from the audits e.g. department stores and entertainment venues. This was firstly because it was often not possible to establish from the roadside whether such outlets sold food, and secondly because such establishments are generally not considered in retail food environment research. However, Lucan et al. [46] found that 23.9% of outlets selling food in New York were businesses not primarily engaged with food retail. Thus, such establishments may make up an important component of the retail food environment. These establishments appear to be listed in both the FSA and POI data, although the completeness of these listings is unknown and extraction of such outlets, particularly for the POI data, will be challenging. One technique may be to extract major chain outlets not primarily engaged in food retailing but known to retail food (e.g. large pharmacies and department stores) based on outlet name. This would not capture all businesses where food retail is secondary to another service, but would present an improvement over existing techniques.

Conclusion

The retail food environment is increasingly targeted as a lever to improve diet and reduce obesity. Food hygiene data (e.g. from local authorities or the FSA) and POI data are both frequently used in research and emergently used in policy tools to characterise the UK food environment. This study found POI and FSA data to have ‘good’ agreement with street audits. Both datasets had variable validity across outlet types and geographic biases, which may need to be accounted for in analyses. Overall policymakers can have confidence in tools and evidence based in these data, although for certain applications (e.g. when policymakers need to know locations of specific food outlets) these data may not be sufficiently valid. Presently local authorities have free access to both FSA data and POI data (via the Food Environment Assessment Tool [36]). While both datasets were similarly valid, in our view the utility of the POI data was better than the FSA data. In particular, use of proprietary classifications in POI data to define outlet classifications was shown to be an accurate method for classifying outlets, which could provide substantial time savings compared to manual classification of outlets. Both datasets required substantial data cleaning, requiring several phases (e.g. removal of duplicates, identification of ‘other’-type outlets). These are important methodological steps that impact the validity of data, and should be reported in research papers.

List of Abbreviations

CI: 95% confidence interval

FSA: Food Standards Agency

IMD: Index of Multiple Deprivation

LSOA: Lower Super Output Area

POI: Points of Interest

PPV: Positive Predictive Value

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The POI data that support the findings of this study are available from Edina Digimap but restrictions apply to the availability of these data, which were used under license for the current study, and so are not made publicly available here. The data have however been accurately referenced such that researchers could obtain this data directly from Edina Digimap under their own user license. The FSA data and audit data that support these findings are available from the corresponding author upon reasonable request. More recent data from the FSA can be obtained from the Food Standards Agency website, http://ratings.food.gov.uk/open-data/en-GB (no DOI available).

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

EW collected and analysed the data. All authors contributed to the writing of the manuscript and read and approved the final draft.

Acknowledgements

Thanks is given to Amy Leadbitter and Christina Telford who helped with data collection during the street audits.

Figure Legends

Figure 1. Flow chart detailing data processing procedure. POI: Points of Interest data; FSA: Food Standards Agency Data.

Figure 2. Venn diagram illustrating the classification of outlets as true positives (TP), false positives (FP) and false negatives (FN). The left-hand oval represents all outlets identified in

the audits, and the right-hand oval represents all outlets identified by the secondary data (POI or FSA). The region of overlap depicts outlets that were identified in both the audits and the dataset. The figure also shows the equations used to calculate sensitivity statistics and positive predictive values (PPV) and their respective odds, where P(X) represents the probability of event X.

Figure 3. Positive Predictive Values (PPV) and sensitivities for FSA and POI data.
* statistically significant difference between datasets ($p < 0.05$). FSA: Food Standards Agency data. POI: Points of Interest data. PPV: positive predictive values.

References

1. Butland B, Jebb S, Kopelman P, McPherson K, Thomas S, Mardell J, Parry V. Foresight. Tackling obesities: future choices. Project report. *Foresight Tackling obesities: future choices Project report.* Edited by Science GOf. London: Stationary Office; 2007: 1-164.
2. Ross A, Chang, M. Planning healthyweight environments – A TCPA reuniting health with planning project. 2014. https://www.tcpa.org.uk/healthyplanning. Accessed 1 September 2016.
3. Local Government Association. Building the foundations - tackling obesity through planning and development. London: Local Government Association; 2016.
4. NHS London Healthy Urban Development Unit. Using the planning system to control hot food takeaways. A good practice Guide. 2013. http://www.healthyurbandevelopment.nhs.uk/wp-content/uploads/2013/12/HUDU-Control-of-Hot-Food-Takeaways-Feb-2013-Final.pdf. Accessed 30/05/16 2016.
5. Local Government Association. Tipping the scales. Case studies on the use of planning powers to limit hot food takeaways. London: Local Government Association; 2016.
6. Public health England. Healthy people, healthy places briefing. Obesity and the environment: regulating the growth of fast food outlets London: Stationary Office; 2014.
7. Greater London Authority. Takeaways Toolkit. 2012. https://www.london.gov.uk/sites/default/files/takeawaystoolkit.pdf. Accessed 24 April 2017.
8. Cobb LK, Appel LJ, Franco M, Jones-Smith JC, Nur A, Anderson CA. The relationship of the local food environment with obesity: A systematic review of methods, study quality, and results. Obesity. 2015;23:1331-44.
9. Lytle LA, Sokol RL. Measures of the food environment: A systematic review of the field, 2007–2015. Health & Place. 2017;44:18-34.
10. Cetateanu A, Jones A. Understanding the relationship between food environments, deprivation and childhood overweight and obesity: evidence from a cross sectional England-wide study. Health & Place. 2014;27:68-76.
11. Jennings A, Welch A, Jones AP, Harrison F, Bentham G, van Sluijs EMF, Griffin SJ, Cassidy A. Local food outlets, weight status, and dietary intake: Associations in children aged 9–10 years. American Journal of Preventive Medicine. 2011;40:405-10.
12. Skidmore P, Welch A, Sluijs Ev, Jones A, Harvey I, Harrison F, Griffin S, Cassidy A. Impact of neighbourhood food environment on food consumption in children aged 9–10 years in the UK SPEEDY (Sport, Physical Activity and Eating behaviour:

Environmental Determinants in Young people) study. Public Health Nutr. 2010;13:1022-30.

13. Mendez DD, Kim KH, Hardaway CR, Fabio A. Neighborhood Racial and Socioeconomic Disparities in the Food and Alcohol Environment: Are There Differences by Commercial Data Sources? Journal Of Racial And Ethnic Health Disparities. 2016;3:108-16.

14. Fleischhacker SE, Evenson KR, Sharkey J, Pitts SBJ, Rodriguez DA. Validity of secondary retail food outlet data: a systematic review. Am J Prev Med. 2013;45:462-73.

15. Lebel A, Daepp MI, Block JP, Walker R, Lalonde B, Kestens Y, Subramanian S. Quantifying the foodscape: A systematic review and meta-analysis of the validity of commercially available business data. PLoS One. 2017;12:1-17.

16. Lebel A, Daepp MI, Block JP, Walker R, Lalonde B, Kestens Y, Subramanian S. Quantifying the foodscape: A systematic review and meta-analysis of the validity of commercially available business data. PLoS One. 2017;12:e0174417.

17. Harrison F, Jones AP, van Sluijs EMF, Cassidy A, Bentham G, Griffin SJ. Environmental correlates of adiposity in 9–10 year old children: Considering home and school neighbourhoods and routes to school. Social Science & Medicine. 2011;72:1411-9.

18. Burgoine T, Lake AA, Stamp E, Alvanides S, Mathers JC, Adamson AJ. Research report: Changing foodscapes 1980–2000, using the ASH30 Study. Appetite. 2009;53:157-65.

19. Fraser LK, Clarke GP, Cade JE, Edwards KL. Fast food and obesity: a spatial analysis in a large United Kingdom population of children aged 13-15. Am J Prev Med. 2012;42:e77-e85.

20. Fraser LK, Edwards KL. The association between the geography of fast food outlets and childhood obesity rates in Leeds, UK. Health & Place. 2010;16:1124-8.

21. Fraser LK, Edwards KL, Tominitz M, Clarke GP, Hill AJ. Food outlet availability, deprivation and obesity in a multi-ethnic sample of pregnant women in Bradford, UK. Soc Sci Med. 2012;75:1048-56.

22. Griffiths C, Frearson A, Taylor A, Radley D, Cooke C. A cross sectional study investigating the association between exposure to food outlets and childhood obesity in Leeds, UK. International Journal of Behavioral Nutrition and Physical Activity. 2014;11:1-10.

23. Macdonald L, Ellaway A, Ball K, Macintyre S. Is proximity to a food retail store associated with diet and BMI in Glasgow, Scotland? BMC Public Health. 2011;11:464-72.

24. Williams J, Scarborough P, Townsend N, Matthews A, Burgoine T, Mumtaz L, Rayner M. Associations between food outlets around schools and BMI among primary students in england: A cross-classified multi-level analysis. PLoS One. 2015;10:1-17.

25. Burgoine T, Forouhi NG, Griffin SJ, Wareham NJ, Monsivais P. Associations between exposure to takeaway food outlets, takeaway food consumption, and body weight in Cambridgeshire, UK: population based, cross sectional study. BMJ. 2014;348:1-10.

26. Burgoine T, Forouhi NG, Griffin SJ, Brage S, Wareham NJ, Monsivais P. Does neighborhood fast-food outlet exposure amplify inequalities in diet and obesity? A cross-sectional study. Am J Clin Nutr. 2016;103:1540-7.

27. Burgoine T, Monsivais P. Characterising food environment exposure at home, at work, and along commuting journeys using data on adults in the UK. The International Journal of Behavioral Nutrition & Physical Activity. 2013;10:85-95.

28. 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;12:1-8.

29. Lake AA, Burgoine T, Greenhalgh F, Stamp E, Tyrrell R. The foodscape: classification and field validation of secondary data sources. Health & Place. 2010;16:666-73.

30. Lake AA, Burgoine T, Stamp E, Grieve R. The foodscape: Classification and field validation of secondary data sources across urban/rural and socio-economic classifications in England. The International Journal of Behavioral Nutrition and Physical Activity. 2012;9:37-48.

31. Cummins S, Macintyre S. Are secondary data sources on the neighbourhood food environment accurate? Case-study in Glasgow, UK. Prev Med. 2009;49:527-8.

32. Burgoine T. Collecting accurate secondary foodscape data. A reflection on the trials and tribulations. Appetite. 2010;55:522-7.

33. Food Standards Agency. Food Hygiene Ratings. 2016. http://ratings.food.gov.uk/. Accessed 8 December 2016.

34. Ordnance Survey. Points of Interest user guide and technical specification v3.4. Ordnance Survey; 2015: 1-59.

35. Public Health England. Obesity and the environment. Density of fast food outlets. 2016. https://www.noo.org.uk/securefiles/170413_1349//FastFoodmap_FINAL.pdf. Accessed 13 April 2017.

36. Public Health England. Strategies for Encouraging Healthier 'Out of Home' Food Provision: A toolkit for local councils working with small food businesses. 2017. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/604912/Encouraging_healthier_out_of_home_food_provision_toolkit_for_local_councils.pdf . Accessed 20 June 2017.

37. Office for National Statistics. A Beginner's Guide to UK Geography. 2016. http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/index.html. Accessed 2 July 2016.

38. Office for National Statistics. English LSOA Rural/Urban Classification, 2011 [data collection]. UK Data Service. SN:5819 UKBORDERS: Digitised Boundary Data, 1840- and Postcode Directories, 1980-. http://discover.ukdataservice.ac.uk/catalogue/?sn=5819&type=Data%20catalogue, Retrieved from http://census.ukdataservice.ac.uk/get-data/boundary-data.aspx.; 2011.

39. Office for National Statistics. Index of Multiple Deprivation [data collection]. UK Government. Retrieved from https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015 2015.

40. Ordnance Survey. Street Atlas West Yorkshire. Fifth edition with interim revisions 2015 edn. London: Philip's; 2015.

41. Ordnance Survey. Street Atlas County Durham and Teeside. Third (A) Colour Edition edn. London: Philip's; 2008.

42. Ordnance Survey. OS Vector Map District Raster Ordnance Survey; 2016.

43. Ordnance Survey. Points of Interest March 2016 version. . 2016. http://edina.ac.uk/digimap. Accessed 1 September 2016.

44. Ordnance Survey. Points of Interest June 2016 version. . 2016. http://edina.ac.uk/digimap. Accessed 6 September 2016.

45. Caspi CE, Friebur R. Modified ground-truthing: an accurate and cost-effective food environment validation method for town and rural areas. International Journal of Behavioral Nutrition & Physical Activity. 2016;13:1-8.

46. Lucan SC, Maroko AR, Bumol J, Torrens L, Varona M, Berke EM. Business list vs ground observation for measuring a food environment: saving time or waste of time (or worse)? Journal of the Academy of Nutrition and Dietetics. 2013;113:1332-9.

47. Brown LD, Cai TT, DasGupta A. Interval estimation for a binomial proportion. Statistical Science. 2001;16:101-17.

48. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-74.

49. Paquet C, Daniel M, Kestens Y, Leger K, Gauvin L. Field validation of listings of food stores and commercial physical activity establishments from secondary data. International Journal of Behavioral Nutrition and Physical Activity. 2008;5:1-7.

50. Hobbs M, Green M, Griffiths C, Jordan H, McKenna J. How different data sources and definitions of neighbourhood influence the association between food outlet availability and body mass index: a cross-sectional study. Perspectives in Public Health. 2016;[Online] Available at: https://www.ncbi.nlm.nih.gov/pubmed/27282516 [Accessed: 1 Sept 2016].

51. Ordnance Survey. Postcode Unit. n.d. http://data.ordnancesurvey.co.uk/ontology/postcode/PostcodeUnit. Accessed 30 October 2017.

52. Greenberg RS, Kleinbawm D. Mathematical modeling strategies for the analysis of epidemiologic research. Annu Rev Public Health. 1985;6:223-45.

53. Bodicoat DH, Carter P, Comber A, Edwardson C, Gray LJ, Hill S, Webb D, Yates T, Davies MJ, Khunti K. Is the number of fast-food outlets in the neighbourhood related to screen-detected type 2 diabetes mellitus and associated risk factors? Public Health Nutr. 2015;18:1698-705.

54. Cerin E, Frank LD, Sallis JF, Saelens BE, Conway TL, Chapman JE, Glanz K. From neighborhood design and food options to residents' weight status. Appetite. 2011;56:693-703.

55. Food Standards Agency. Local Authority Enforcement Monitoring System (LAEMS), April 2016. 2016. https://www.food.gov.uk/sites/default/files/multimedia/pdfs/enforcement/laemsguidance.pdf. Accessed 22 Dec 2016.

56. Food Standards Agency. Evaluation of the Food Hygiene Rating Scheme and Food Hygiene Information Scheme: Process Evaluation Stage 2 – Final Report. 2015. https://www.food.gov.uk/science/research/fs244011/fs244011A. Accessed 21 December 2016.

57. Ordnance Survey. Points of Interest Classification Scheme v3.1. Ordnance Survey; 2013: 1-10.

58. Google Maps. Map of Armley Road and Canal Street, Armley, Leeds, UK. 2017. https://www.google.co.uk/maps/place/Canal+Rd,+Bradford/@53.8281078,-1.7695618,14z/data=!4m5!3m4!1s0x487be155a80f4e2d:0xb339086a11481f9b!8m2!3d53.810019!4d-1.7571163. Accessed 13 April 2017.

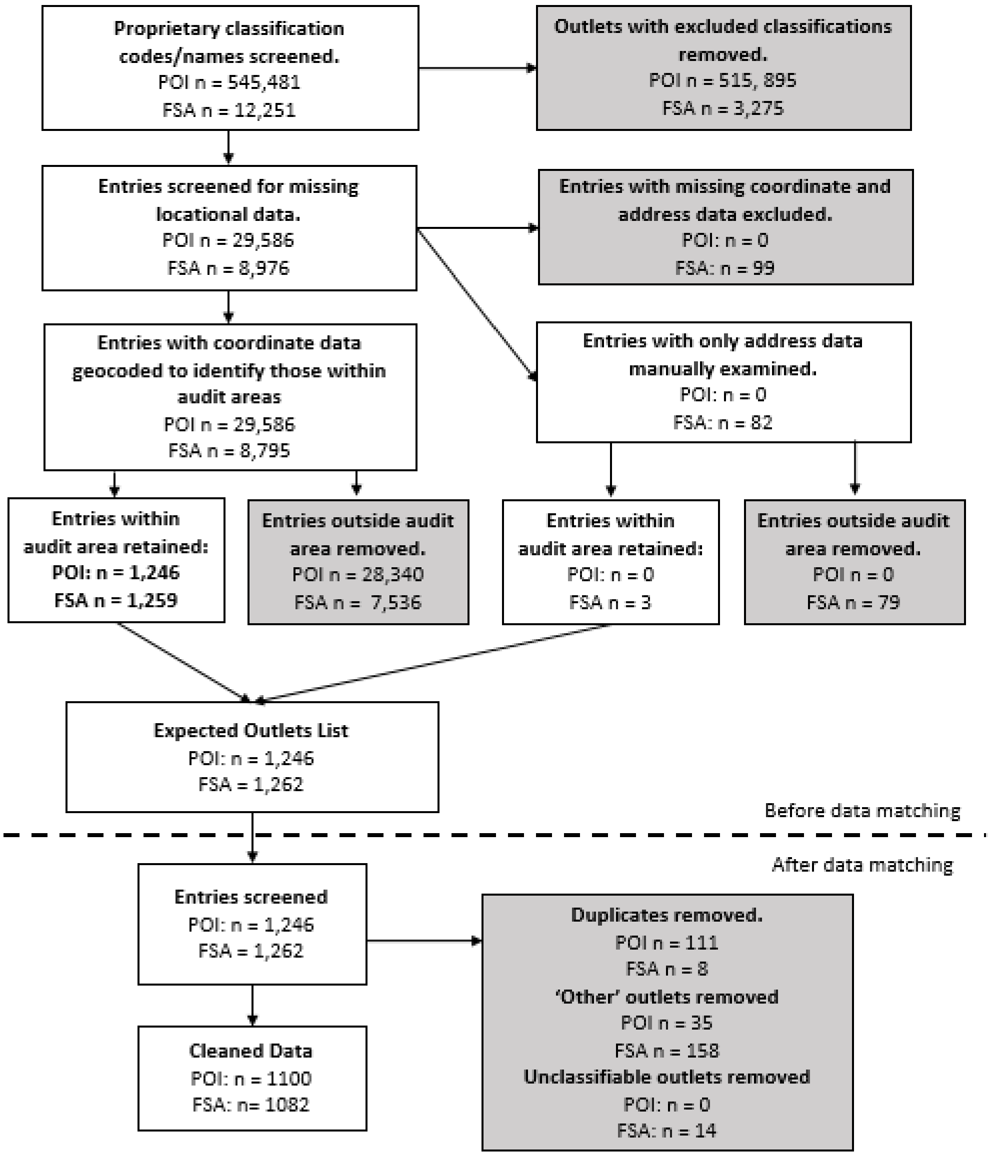
Proprietary classification codes/names screened.
POI n = 545,481
FSA n = 12,251
Outlets with excluded classifications removed.
POI n = 515, 895
FSA n = 3,275
Entries screened for missing locational data.
POI n = 29,586
FSA n = 8,976
Entries with missing coordinate and address data excluded.
POI: n = 0
FSA: n = 99
Entries with coordinate data geocoded to identify those within audit areas
POI n = 29,586
FSA n = 8,795
Entries with only address data manually examined.
POI: n = 0
FSA: n = 82
Entries within audit area retained:
POI: n = 1,246
FSA n = 1,259
Entries outside audit area removed.
POI n = 28,340
FSA n = 7,536
Entries within audit area retained:
POI: n = 0
FSA n = 3
Entries outside audit area removed.
POI n = 0
FSA n = 79
Expected Outlets List
POI: n = 1,246
FSA = 1,262
Before data matching
After data matching
Entries screened
POI: n = 1,246
FSA = 1,262
Duplicates removed.
POI n = 111
FSA n = 8
'Other' outlets removed
POI n = 35
FSA n = 158
Unclassifiable outlets removed
POI: n = 0
FSA: n = 14
Cleaned Data
POI: n = 1100
FSA: n= 1082

Figure 1

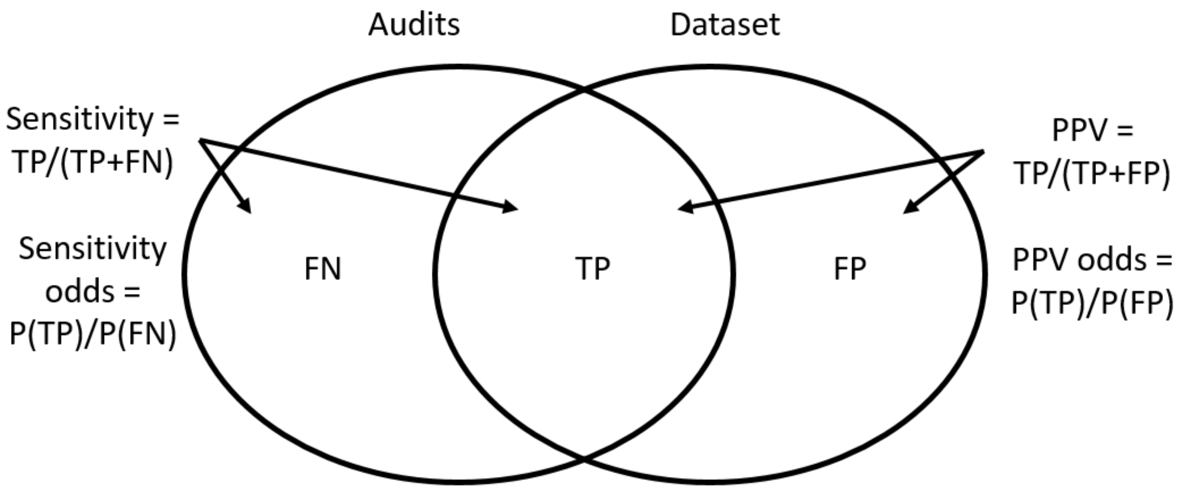
Audits
Dataset
Sensitivity =
TP/(TP+FN)
Sensitivity
odds =
P(TP)/P(FN)
FN
TP
FP
PPV =
TP/(TP+FP)
PPV odds =
P(TP)/P(FP)

Figure 2

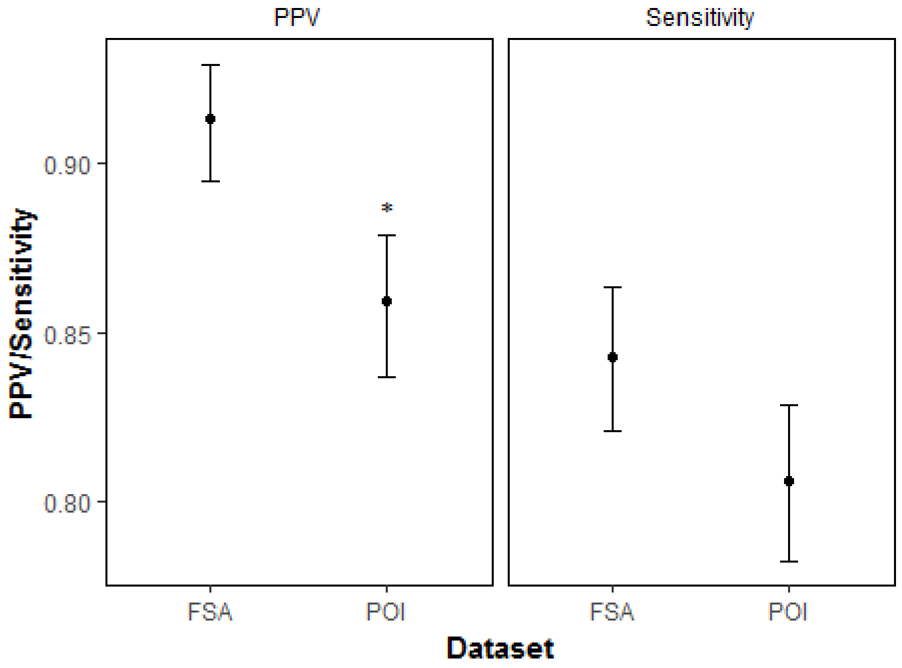
PPV
Sensitivity
PPV/Sensitivity
0.90
0.85
0.80
*
FSA
POI
FSA
POI
Dataset

Figure 3

Supplementary Materials

1 Additional information on POI and FSA data

3.1 Description of FSA Data

In the UK, any business intending to conduct 'food operations' (i.e. selling, cooking, storing, handling, preparing or distributing food/drink) must register the business with their local authority. This data is used for enforcing Food Safety laws and in particular for carrying out inspections of food businesses to assign food hygiene ratings. From 2009 local authorities have been required, under the Food Hygiene Information/Rating Schemes to supply food hygiene ratings data to the Food Standards Agency (FSA) for publication on their website [56]. The schemes require ratings data to be provided for all registered food businesses that supply food/drink directly to consumers. Notably, however, certain FO (e.g. those deemed to be low-risk; such as pharmacies) may not receive a hygiene rating and may be excluded from the FSA data. Thus, while the FSA data should comprise most FO, it may exclude certain 'low-risk' FO that are nevertheless important for RFE assessment.

For each food business, the FSA data includes (inter alia), business name, business address, business classification, and locational coordinates (latitude and longitude according to the WGS84 geographic Coordinate Reference System). There is no publically available information on how the locational coordinates are derived. However, the majority of locational coordinates appear to align with points generated through postcode geocoding, suggesting that the majority of coordinates are geocoded to the postcode level. The business classifications are applied by local authorities according to the 'Local Authority Enforcement Monitoring System' (LAEMS) classification scheme, which comprises 15 classifications (Table 9) [55].

It should be noted that there is some inconsistency in nomenclature between the official FSA classification names (listed in Table 1) and the classification names that were found to have been applied to the data as detailed:

- Primary Producer also named Farmers/Growers
- Caring Establishment also named Hospitals/Childcare/Caring Premises
- Pub/Club also named Pub/Bar/Club
- Hotel/Guesthouse also named Hotel/Bed & Breakfast/Guesthouse
- School/College also named School/College/University
- Take-away also named Takeaway/Sandwich Shop
- Mobile Food Unit also named Mobile Caterer

Local authorities are required to supply their ratings data every 27 days, and this data is then made available to the public via the FSA website. How frequently each local authority updates its own database of food businesses is unclear, and the currency of the FSA data may vary between local authorities. All 392 local authorities across the UK are presently participating in the Information/Rating schemes; however, this is not obligatory for local authorities in England and Scotland.

Table 9

FSA Classification Names and Definitions

Classification Name	**FSA Definition/Examples**
Primary Producer	"Examples: • Fruit and vegetable growers • Pick your own farms • Egg producers • Potato growers • Fish farms • Beekeepers • Vineyards"
Manufacturers/ Packers*	"Examples: • Abattoirs • Brewery • Meat manufacturers • Milk processors & dairy processors • Cheesemakers • Soft drinks, mineral waters • Vegetable drying, freezing, canning • Meat or poultry cutting establishments • Purification centres for shellfish • Fish processors • Butchers shops cooking hams • Fruit & vegetable co-operatives • Egg packers • Contract packers • Food contact material and article manufacturers & suppliers • Bakers with no on-site retail • Bakeries selling through their own shops • Home cake makers selling to other businesses"
Importers/Exporters	"Examples: • Warehouses for import/export purposes • Freight depots, transit sheds, stores"
Distributors/Transporters	"Examples: • Food brokers • Wholesalers • Cash & carries • Cold stores • Haulage companies • Milk distributors"
Restaurant & Caterers - Caring Establishment	"Establishments with catering services for clients/customers who are provided with care, medical treatment, supervision, or assistance. Examples: • Hospitals (include each establishment but not each kitchen) • Nursing/care homes • Childcare facilities/nurseries/childminders"
Restaurant & Caterers - School/College	"Catering services located within a site providing educational instruction and formal qualifications. Examples: • Colleges • Schools (include each establishment but not each kitchen)"
Restaurant & Caterers - Hotel /Guesthouse	"Establishments that provide catering only to customers to whom they are also providing accommodation. (Hotels that provide a

Classification Name	FSA Definition/Examples
	restaurant service to a wider clientele than their guests should be recorded under the 'restaurant/café/canteen' category). Examples: • Hotels • Guest houses • Bed and breakfast”
Restaurant & Caterers - Mobile Food Unit	“A food establishment that comprises a kitchen or catering facility operating from a mobile unit such as a vehicle, trailer, stall, marquee or other non-permanent structure. Examples: • Mobile catering units • Burger vans and other fast food vans/trailers/stalls”
Restaurant & Caterers - Pub/Club	“Commercial establishments that primarily serve alcohol in a public bar. If the establishment has a separate restaurant facility it should be recorded under the pub category. Examples: • Public Houses • Night clubs/clubs with bars”
Restaurant & Caterers - Restaurant/Café/Canteen	“Establishments whose primary business is to cook/prepare food for consumption by customers at a seated area on the premises. Examples: • Restaurants • Cafés • Self-service cater • ‘Fast food' establishments providing seating, e.g. McDonalds, Burger King etc. The drive-thru variants of these chains should also be included in this category.”
Restaurant & Caterers - Take-away	“Establishments that provide convenience food to customers, primarily for consumption off the premises. Establishments must be immobile and housed in a designated building Examples: • Fish & chip shops • Take-away • Sandwich shops • Establishments that prepare and deliver convenience food directly to the customer”
Restaurant & Caterers - Other Catering Premises	“Restaurant/catering establishments that do not fit into one of the other 'restaurants and caterers' categories. Examples: • Home caterers such as cake makers selling directly to consumers • Village halls, community centres etc. used by charitable/community organisations, see www.food.gov.uk/enforcement/enforcework/food-law/guidance-enforcement/community-hall-guidance • Ships' catering spaces”
Retailers – Supermarkets /Hypermarkets	“Supermarkets e.g. Sainsbury, Tesco, Asda, Morrison, Co-op, Marks and Spencer, Waitrose, Aldi, Lidl, Budgens etc. that provide a range of food items from more than one grocery sector and from a range of brands. Also city centre or local variants of larger supermarket groups, e.g. Sainsbury's local, Tesco Metro, Tesco Express etc. Examples: • Supermarkets - the large retail chains • City centre or local variants of larger supermarket groups”

Classification Name	FSA Definition/Examples
Retailers – Smaller Retailers**	“Smaller-scale food businesses such as butchers, bakers, fishmongers, village shops, grocers etc. Independent retailers e.g. Costcutter, One-Stop, Londis, Nisa, Premier etc. Examples: • Grocers • Confectioners • Butchers (retail only) • Fishmongers • Greengrocer/fruiterer • Health food shops • Bakers shops (retail only) • Newsagents • Mobile vans (retailers) • Market stalls (retailers) • Farm shops (if farm not included under producers or other establishments) • Off licences • Garage minimarkets”
Retailers - other	“Retail establishments which do not fit into one of the other retailer categories, e.g. establishments that primarily sell non-food products and a very limited range of food products. Examples: • Shops where the main business is not food, e.g. chemist/pharmacy that sell cough sweets/limited range of other confectionery”

Note. Classification names and definitions are taken directly from Food Standards Agency [55]. * This classification was applied by some local authorities to retail butchers which should have been classified under ‘retailers – smaller retailers’
**This classification had not been applied to any of the data used in the present study; all outlets that should fall within this classification were instead classified as ‘Retailers – other’. Items in bold are those extracted.

3.2 Description of POI Data

POI is a dataset detailing over 4 million geographic features across Great Britain [34]. It is produced by PointX Ltd on behalf of Ordnance Survey (the national mapping agency for Great Britain) for a variety of uses, including the provision of facilities and infrastructure, driver routing and navigation, emergency planning, location-based services and tourism. While access to the data is usually at a cost, it is available for free for research purposes under an educational license, and is often used in research examining the built and natural environment [10-12, 17, 19]

According to the user guide, the POI data is obtained from around 140 data suppliers, which are described as “the most authoritative source or sources for the particular type of feature they supply and for the quality and completeness of the data they supply” [34]. The data suppliers provide updates at different frequencies, ranging from bi-monthly to yearly. Thus the currency of the data can vary between features.

The POI dataset contains coordinate (eastings and northings according to the British National Grid projected coordinate reference system), classification and address information for each feature therein. Feature classifications are shown at Table 10. The classification scheme comprises over 600 classifications descriptive of a feature’s function. The classifications fall within one of nine groups: "accommodation, eating and drinking”, “commercial services”, “attractions”, “sport and entertainment”, “education and health”, “public infrastructure”, “manufacturing and production”, “retail”, and “transport” [57]. Classifications are generally applied to the data by the original data supplier. However, PointX also apply classifications if none is provided by the data supplier. Documentation is provided by PointX detailing common names/brands of businesses falling within each classification to facilitate interpretation of the classifications.

According to the user guide, the coordinate data for each feature is derived by geocoding the feature to an address location (i.e. within a building footprint) wherever possible (79.87% of features were geocoded using this method in the September 2014 release). However when this is not possible, features are either geocoded to an adjacent address, a street segment midpoint or a geographic locality (e.g. village or industrial estate). The latter two methods are only used for a small range of feature types and are not used for food outlets. According to the user guide, 95% of features are geocoded to within 17.51 metres of their true location.

Table 10

List of Points of Interest Classifications and Associated Groupings

Group/Sub-Group	Classification Name
Accommodation, Eating and Drinking	
Accommodation	Camping, caravanning, mobile homes, holiday parks and centres
	Bed and breakfast and backpacker accommodation
	Hostels and refuges for the homeless
	Hotels, motels, country houses and inns
	Self catering
	Timeshare
	Youth accommodation
Eating and Drinking	Banqueting and function rooms
	Cafés, snack bars and tea rooms
	Fast food and takeaway outlets
	Fast food delivery services
	Fish and chip shops
	Internet cafés
	Pubs, bars and inns
	Restaurants
Commercial Services	
Construction Services	Metalworkers including blacksmiths
	Building contractors
	Construction completion services
	Construction plant
	Cutting, drilling and welding services
	Demolition services
	Diving services
	Electrical contractors
	Gardening, landscaping and tree surgery services
	Glaziers
	Painting and decorating services
	Plasterers
	Plumbing and heating services
	Pool and court construction
	Restoration and preservation services
	Road construction services
	Roofing and chimney services
	Fencing and drystone walling services
	Building and component suppliers
Consultancies	Architectural and building related consultants
	Business related consultants
	Computer consultants
	Construction service consultants
	Feng shui consultants, furnishers and shop fitters
	Food consultants
	Image consultants
	Interpretation and translation consultants
	Security consultants
	Telecommunications consultants

Group/Sub-Group	Classification Name
	Traffic management and transport related consultants
Employment and career agencies	Careers offices and armed forces recruitment
	Domestic staff and home help
	Driver agencies
	Employment agencies
	Modelling and theatrical agencies
	Nursing agencies
Engineering services	Aviation engineers
	Chemical engineers
	Civil engineers
	Electrical and electronic engineers
	Hydraulic engineers
	Industrial engineers
	Instrumentation engineers
	Marine engineers and services
	Mechanical engineers
	Pneumatic engineers
	Precision engineers
	Structural engineers
Contract services	Agricultural contractors
	Aircraft charters
	Catering services
	Contract cleaning services
	Display and window dressers
	Drain and sewage clearance
	Linen hire and washroom services
	Office services
	Packers
	Pest and vermin control
IT, advertising, marketing and media services	Advertising services
	Artists, illustrators and calligraphers
	Computer security
	Computer systems services
	Concert/exhibition organisers and services
	Database services
	Desktop publishing services
	Electronic and internet publishers
	Film and video services
	General computer services
	Internet services
	Literary services
	Mailing and other information services
	Marketing services
	Plate makers, print finishers and type setters
	Press and journalism services
	Printing and photocopying services
	Recording studios and record companies
	Telephone, telex and fax services

Group/Sub-Group	Classification Name
	Television and radio services
Legal and financial	Accountants and auditors
	Auctioneers, auction rooms and valuers
	Banks and building societies
	Currency conversion and money transfers
	Cash machines
	Cheque cashing
	Company registration and trademarks
	Copyright and patent
	Credit reference agencies
	Debt collecting agencies
	Financial advice services
	Fundraising services
	Insurers and support activities
	Mortgage and financial lenders
	Pawnbrokers
	Solicitors, advocates and notaries public
	Stocks, shares and unit trusts
	Commodity dealers
	Franchise and holding company services
	Paypoint locations
	Pension and fund management
Personal, consumer and other services	Hotel booking agencies
	Event ticket agents and box office
	Astrologers, clairvoyants and palmists
	Hair and beauty services
	Cleaning services
	Customer service centres
	CV writers
	Detective and investigation agencies
	Funeral and associated services
	Historical research
	Headquarters, administration and central offices
	Introduction and dating agencies
	Lock, key and security services
	Message and greeting services
	Motoring organisations
	Party organisers
	Personalisation
	Photographic services
	Sports services
	Tattooing and piercing services
	Trophies and engraving services
	Vehicle cleaning services
	Weather services
	Wedding services
	Window cleaners
	Musicians, orchestras and composers

Group/Sub-Group	Classification Name
	Sculptors, wood workers and stone masons
	Tailoring and clothing alteration
	Vehicle breakdown and recovery services
	Sewage services
	Spas
	Slimming clubs and services
	Adult services
	Printing on garments
Property and development services	Commercial property letting
	Property sales
	Estate and property management
	Property letting
	Property development services
	Property information services
Recycling services	Recycling, reclamation and disposal
	Rag merchants
	Clearance and salvage dealers
	Scrap metal dealers and breakers yards
	Waste paper merchants
Repair and servicing	Building repairs
	Electrical equipment repair and servicing
	Household repairs and restoration
	Industrial repairs and servicing
	Service industry equipment repairs
	Sports and leisure equipment repair
	Tool repairs
	Vehicle repair, testing and servicing
	Shoe repairs
Research and design	Design services
	Research services
	Testing and analysis services
Transport, storage and delivery	Airlines and airline services
	Animal transportation
	Container and storage
	Courier, delivery and messenger
	Distribution and haulage
	Ferry and cruise companies
	Import and export services
	Railway related services
	Removals and shipping agents
	Taxi services
Hire services	Boat hiring services
	Construction and tool hire
	Leisure equipment hirings
	Renting and leasing of personal and household goods
	Sound, light and vision service and equipment hire
	Vehicle hire and rental
	Clothing hire

Group/Sub-Group	Classification Name
	Bouncy castles and inflatables hire
Attractions	
Botanical and zoological	Aquaria and sea life centres
	Bird reserves, collections and sanctuaries
	Butterfly farms
	Farm based attractions
	Horticultural attractions
	Salmon ladders
	Zoos and animal collections
Historical and cultural	Archaeological sites
	Battlefields
	Historic buildings including castles, forts and abbeys
	Historic and ceremonial structures
	Historical ships
	Museums
	Art galleries
Recreational	Commons
	Country and national parks
	Picnic areas
	Playgrounds
	Municipal parks and gardens
Landscape features	Designated scenic features
	Trigonometric points
Tourism	Laseria, observatories and planetaria
	Model villages
	Railways (heritage, steam and miniature)
	Theme and adventure parks
	Siteseeing, tours, viewing and visitor centres
	Information centres
	Unspecified and other attractions
Bodies of water	Ponds
	Lakes and waters
	Lochs and lochans
	Tarns, pools and meres
	Reservoirs
	Settling, balancing and silt ponds
Sport and Entertainment	
Sport and entertainment support services	Children's activity centres
	Entertainment services
	Firework related services
	Funfair services
	Mobile discos
	Motorsport services
Gambling	Amusement parks and arcades
	Bingo halls
	Bookmakers
	Casinos
	Pools promoters

Group/Sub-Group	Classification Name
Outdoor pursuits	Angling and sports fishing
	Combat, laser and paintball games
	Hot air ballooning
	Parachuting and bungee jumping
	Paragliding and hang gliding
	Watersports
	Riding schools, livery stables and equestrian centres
	Outdoor pursuit organisers and equipment
Sports complex	Athletics facilities
	Bowling facilities
	Climbing facilities
	Golf ranges, courses, clubs and professionals
	Gymnasiums, sports halls and leisure centres
	Ice rinks
	Motorsport venues
	Racecourses and greyhound tracks
	Shooting facilities
	Ski infrastructure and aerial cableways
	Snooker and pool halls
	Sports grounds, stadia and pitches
	Squash courts
	Swimming pools
	Tennis facilities
	Velodromes
Venues, stage and screen	Cinemas
	Discos
	Nightclubs
	Social clubs
	Theatres and concert halls
	Conference and exhibition centres
	Adult venues
Education and Health	
Animal welfare	Animal clipping and grooming
	Dog training
	Horse training
	Kennels and catteries
	Pet cemeteries and crematoria
	Veterinarians and animal hospitals
	Veterinary pharmacies
Education support services	Education authorities
	Education services
	Examination boards
	Playing for success centres
	Secure units
Health practitioners and establishments	Alternative, natural and complementary
	Foot related services
	Dental technicians
	Dieticians and nutritionists

Group/Sub-Group	Classification Name
	Homeopaths
	Midwifery
	Optometrists and opticians
	Physical therapy
	Speech therapists
	Surgeons and cosmetic surgeries
	Chemists and pharmacies
	Clinics and health centres
	Dental and medical laboratories
	Dental surgeries
	Doctors surgeries
	Hospices
	Hospitals
	Mental health centres and practitioners
	Nursing and residential care homes
	Accident and emergency hospitals
	Parenting and childcare services
	Walk-in centre
	Day and care centres
Health support services	Medical equipment rental and leasing
	Ambulance and medical transportation services
	Blood transfusion service
	Counselling and advice services
	Health authorities
	Medical waste disposal services
	Pregnancy related services and help centres
	X-ray services
Primary, secondary and tertiary education	First, primary and infant schools
	Further education establishments
	Independent and preparatory schools
	Broad age range and secondary state schools
	Special schools and colleges
	Higher education establishments
	Unspecified and other schools
	Pupil referral units
Recreational and vocational education	Ballet and dance schools
	Beauty and hairdressing schools
	Diving schools
	Drama schools
	Driving and motorcycle schools
	First aid training
	Flying schools
	Language schools
	Martial arts instruction
	Music teachers and schools
	Nursery schools and pre and after school care
	Sailing schools
	Sports and fitness coaching

Group/Sub-Group	Classification Name
	Training providers and centres
Public Infrastructure	
Central and local government	Armed services
	Coastal safety
	Consular services
	Courts, court services and tribunals
	Driving test centres
	Embassies and consulates
	Fire brigade stations
	Central government
	Local government
	Revenue and customs offices
	Job centres
	Members of parliament and members of european parliament
	Police stations
	Prisons
	Probation offices and police support services
	Registrars offices
	Social service activities
	Tribunals
	Foreign country support activities
Infrastructure and facilities	Electrical features
	Fire safety features
	Gas features
	Meteorological features
	Refuse disposal facilities
	Waste storage, processing and disposal
	Telecommunications companies
	Telecommunications features
	Utility companies and brokers
	Allotments
	Cemeteries and crematoria
	Drinking fountains and water points
	Halls and community centres
	Letter boxes
	Libraries
	Places of worship
	Public telephones
	Public toilets
	Recycling centres
	Wifi hotspots
Organisations	Animal welfare organisations
	Fan clubs and associations
	Sports clubs and associations
	Institutes and professional organisations
	Political parties and related organisations
	Religious organisations
	Youth organisations

Group/Sub-Group	Classification Name
	Community networks and projects
	Charitable organisations
	Conservation organisations
Manufacturing and Production	
Consumer products	Baby, nursery and playground equipment
	Beds and bedding
	Brushes
	Candles
	Canvas goods
	Carpets, flooring, rugs and soft furnishings
	Medals, trophies, ceremonial and religious goods
	China and glassware
	Clothing, components and accessories
	Cookers and stoves - non electrical
	Cosmetics, toiletries and perfumes
	Curtains and blinds
	Cutlery and tableware
	Disability and mobility equipment
	Refrigeration and freezing appliances
	Footwear
	Furniture
	Garden goods
	Giftware
	Hobby, sports and pastime products
	Disposable products
	Jewellery, gems, clocks and watches
	Lampshades and lighting
	Leather products
	Lingerie and hosiery
	Luggage, bags, umbrellas and travel accessories
	Musical instruments
	Photographic and optical equipment
	Saunas and sunbeds
	Tents, marquees and camping equipment
	Tobacco products
	Fireplaces and mantelpieces
	Conservatories
	Bathroom fixtures, fittings and sanitary equipment
Extractive industries	Coal mining
	Oil and gas extraction, refinery and product manufacture
	Ore mining
	Peat extraction
	Sand, gravel and clay extraction and merchants
	Stone quarrying and preparation
	Unspecified quarries or mines
Farming	Animal breeders (not horses)
	Arable farming
	Bee keepers

Group/Sub-Group	Classification Name
	Dairy farming
	Fish and shellfish
	Forestry
	Fruit, flower and vegetable growers
	Hoppers and silos
	Horse breeders and dealers
	Livestock farming
	Mixed or unspecified farming
	Poultry farming, equipment and supplies
	Sheep dips and washes
Foodstuffs	Alcoholic drinks
	Animal feeds, pet foods, hay and straw
	Baking and confectionery
	Dairy products
	Fish, meat and poultry products
	Milling, refining and food additives
	Non alcoholic drinks
	Catering and non specific food products
Industrial features	Business parks and industrial estates
	Chimneys
	Conveyors
	Energy production
	Lighting towers
	Lime kilns
	Oast houses
	Pipelines
	Tanks (generic)
	Travelling cranes and gantries
	Unspecified works or factories
	Water pumping stations
Industrial products	Abrasive products and grinding equipment
	Adhesives and sealants
	Aeroplanes
	Agricultural machinery and goods
	Air and water filtration
	Arms and ammunition
	Bearing, gear and drive elements
	Beekeeping supplies
	Bricks, tiles, clay and ceramic products
	Cable, wire and fibre optics
	Colours, chemicals and water softeners and supplies
	Cleaning equipment and supplies
	Concrete products
	Cooling and refrigeration
	Electrical components
	Electrical motors and generators
	Electrical production and manipulation equipment
	Electronic equipment

Group/Sub-Group	Classification Name
	Electronic media
	Engines
	Fertilisers
	Food and beverage industry machinery
	General construction supplies
	General purpose machinery
	Glass
	Horticultural equipment
	Industrial coatings and finishings
	Tools including machine shops
	Lifting and handling equipment
	Lubricants and lubricating equipment
	Marine equipment including boats and ships
	Measurement and inspection equipment
	Medical equipment, supplies and pharmaceuticals
	Metals manufacturers, fabricators and stockholders
	Moulds, dies and castings
	Office and shop equipment
	Ovens and furnaces
	Packaging
	Paints, varnishes and lacquers
	Pesticides
	Printing related machinery
	Published goods
	Pumps and compressors
	Radar and telecommunications equipment
	Road maintenance equipment
	Ropes, nets and cordage
	Rubber, silicones and plastics
	Seals, tapes, taps and valves
	Signs
	Special purpose machinery and equipment
	Textiles, fabrics, silk and machinery
	Stationery, stamps, tags and labels
	General manufacturing
	Vehicle bodybuilders
	Vehicle components
	Vehicles
	Wood products including charcoal, paper, card and board
	Workwear
	Educational equipment and supplies
	Ice
	Fences, gates and railings
	Access equipment
	Car ports and steel buildings
	Waste collection, processing and disposal equipment
	Glass fibre services
	Shelving, storage, safes and vaults

Group/Sub-Group	Classification Name
Retail	
Clothing and accessories	Clothing
	Footwear
	Jewellery and fashion accessories
	Lingerie and hosiery
	Baby and nursery equipment and children's clothes
Food, drink and multi item retail	Bakeries
	Butchers
	Confectioners
	Delicatessens
	Fishmongers
	Frozen foods
	Green and new age goods
	Grocers, farm shops and pick your own
	Herbs and spices
	Alcoholic drinks including off licences and wholesalers
	Organic, health, gourmet and kosher foods
	Convenience stores and independent supermarkets
	Livestock markets
	Markets
	Cash and carry
	Tea and coffee merchants
	Supermarket chains
Household, office, leisure and garden	Books and maps
	Carpets, rugs, soft furnishings and needlecraft
	China and glassware
	Cosmetics, toiletries, perfumes and hairdressing supplies
	Craft supplies
	Cycles and accessories
	DIY and home improvement
	Furniture
	Garden centres and nurseries
	Garden machinery and furniture
	General household goods
	Hobby, sports and pastime products
	Leather goods, luggage and travel accessories including handbags
	Lighting
	Music and video
	Musical instruments
	Pets, supplies and services
	Camping and caravanning
	Travel agencies
	Department stores
	Discount stores
	Mail order and catalogue stores
	Shopping centres and retail parks
	Surplus goods
	Art and antiques

Group/Sub-Group	Classification Name
	Charity shops
	Florists
	Gifts and cards
	Party goods and novelties
	Secondhand goods
	Computer supplies
	Domestic appliances
	Electrical goods and components
	Photographic and optical equipment
	Stationery and office supplies
	Telephones and telephone cards
	Post offices
	Garages, garden and portable buildings
	Fuel distributors and suppliers
	Adult shops
	Comics bookshops
	Computer shops
	Potteries
Motoring	New vehicles
	Secondhand vehicles
	Vehicle auctions
	Vehicle parts and accessories
Transport	
Air	Aeronautical features
	Airports and landing strips
	Helipads
Road and rail	Bridges
	Cattle grids
	Fords and level crossings
	Motorway service stations
	Parking
	Petrol and fuel stations
	Roadside telephone boxes
	Signalling facilities
	Tunnels
	Viaducts
	Weighbridges
Walking	Finger posts, guide posts and cairns
	Footbridges
	Stepping stones
	Subways
Water	Aqueducts
	Locks
	Moorings and unloading facilities
	Rivers and canal organisations and infrastructure
	Weirs, sluices and dams
	Ferries and ferry terminals
	Bus and coach stations, depots and companies

Group/Sub-Group	Classification Name
Public transport, stations and infrastructure	Railway stations, junctions and halts
	Tram, metro and light railway stations and stops
	Taxi ranks
	Underground network stations
	London underground entrances
Bus transport	Bus stops
	Hail and ride zones

Note. Group names are underlined; sub-group names are not underlined, and are indented. Adapted from Ordnance Survey [57].

2 Additional Methodological Details

2.1 Re-ranking of LSOAs for relative deprivation

As the degree of deprivation in England is not evenly distributed across urban and rural areas (e.g. only 0.8% of rural LSOAs, versus 12.0% of urban LSOAs are within the lowest decile of deprivation), Index of Multiple Deprivation (IMD) rankings were modified to reflect the degree of deprivation of an LSOA relative to other LSOAs with the same rural/urban classification. This was achieved by first stratifying all LSOAs by their Rural/Urban Classification (RUC) codes into urban and rural environments (RUC codes A1, B1, C1, C2 and D1, D2, E1, E2 respectively). The urban and rural LSOAs were then re-ranked separately, based on their England-wide IMD rankings. Figure 1 illustrates the re-ranking process for six fictional LSOAs. LSOAs were divided into deciles of deprivation based on their new urban/rural IMD ranks.

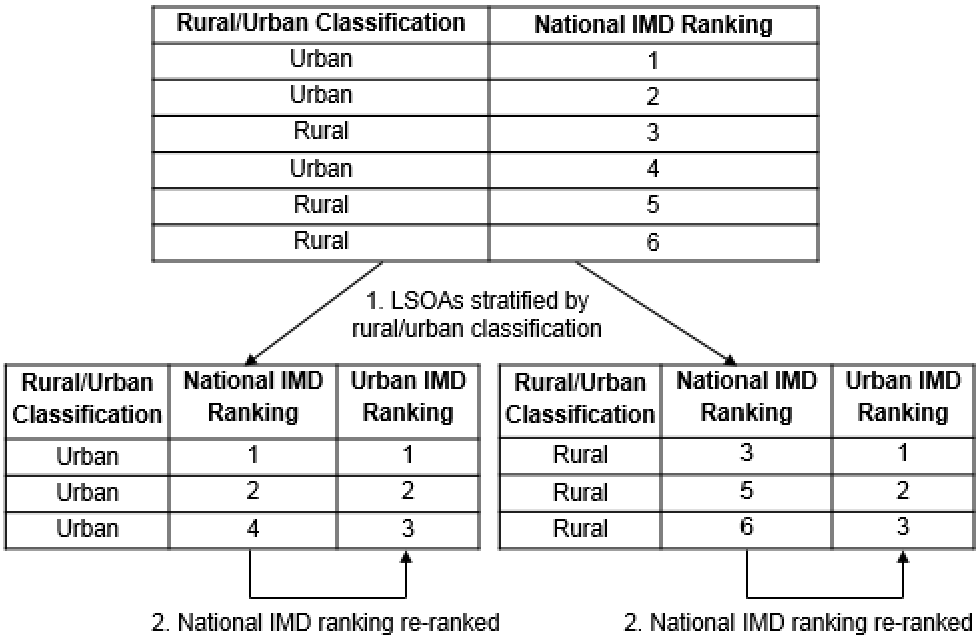

Figure 1. Procedure for re-ranking national IMD rankings to urban and rural IMD rankings. Each row represents one LSOA. Step 1: LSOAs are stratified based on their rural/urban classification. Step 2: For urban and rural LSOAs separately, urban/rural rankings are assigned based on national IMD ranks.

2.2 LSOAs Selected for Auditing

Table 11
Lower Super Output Areas Audited by Area Type, IMD Decile and RUC Classification

Area Type	LSOA Code	Urban/Rural IMD Decile	National IMD decile	RUC Class
Urban Deprived	Leeds 063D	2	2	A1
	Leeds 056E	2	1	A1
	Leeds 056C	2	2	A1
	Leeds 071C	1	1	A1
	Leeds 071B	1	1	A1
	Leeds 048A	1	1	A1
	Leeds 048D	1	1	A1
	Leeds 048C	1	1	A1
	Leeds 053B	1	1	A1
	Leeds 053C	1	1	A1
	Leeds 065A	1	1	A1
Urban Middle	Leeds 009A	7	7	A1
	Leeds 034C	5	5	A1
	Leeds 111A	5	5	A1
	Leeds 111E	6	5	A1
Urban Affluent	Leeds 021C	8	8	A1
	Leeds 027B	10	10	A1
	Leeds 028E	9	9	A1
	Leeds 014B	9	9	A1
	Leeds 014D	9	9	A1
	Leeds 020C	8	8	A1
	Leeds 008A	10	10	A1
Rural Deprived	County Durham 046A	1	3	D1
	County Durham 051D	1	1	D1
	County Durham 066A	2	4	D1
	County Durham 059C	1	2	D1
	County Durham 059D	1	2	D1
	County Durham 038B	1	2	D1
	County Durham 038E	1	2	D1
	North Kesteven 007D	2	5	D1
Rural Middle	County Durham 066C	6	7	D1
	County Durham 033A	6	7	D1
	Calderdale 004E	4	6	D1
	Calderdale 007A	5	7	D1
	North Kesteven 004C	6	7	D1
	Calderdale 027C	5	6	D1
Rural Affluent	Leeds 005B	10	10	D1
	Leeds 005D	10	10	D1
	Leeds 030A	9	9	D1
	Leeds 022C	8	8	D1
	Leeds 007A	10	10	D1
	Leeds 007C	9	9	D1
	Leeds 007F	10	10	D1
	North Kesteven 006B	9	9	D1
	North Kesteven 006D	9	9	E1
	North Kesteven 009C	8	9	D1

Area Type	LSOA Code	Urban/Rural IMD Decile	National IMD decile	RUC Class
	North Kesteven 001A	9	9	D1
	North Kesteven 001B	10	10	D1
	North Kesteven 001C	10	10	D1
	North Kesteven 001D	9	9	D1
	North Kesteven 001E	10	10	D1
	Calderdale 027A	10	10	D1

Note. A1: urban major conurbation; D1: rural town and fringe; E1: Rural village and dispersed

2.3 Modifications to LSOA boundaries

Once selected, the LSOA boundaries were copied by hand onto printed street maps [40-42] to define audit areas. The LSOA boundaries were simplified such that each LSOA only included whole road segments (defined as a segment of road running between junctions or notable geographic features such as the edge of a park). This was so that the auditors would easily be able to determine the extent of an audit area by identifying the junction/geographic feature marking the end of the street segment. In general, a road segment that fell partially within the LSOA was included if more than 50% of the segment fell within the LSOA (assessed visually) and was excluded otherwise. However, some roads had to be excluded for safety reasons e.g. if the road was fast and narrow with no footpath and thus could not be walked safely. Furthermore, occasionally additional streets falling outside the LSOA were included within an audit area if (i) one or more food outlets were indicated to be located on the street in close proximity to the LSOA boundary, (ii) to improve efficiency of the audits e.g. if the audit team would need to cover the street anyway, or would need to back-track if the street was not included. This was done to ensure auditing was as efficient as possible, and maximised the number of food outlets identified relative to the financial and time cost involved.

Figure 2 shows an example of two LSOA boundaries and corresponding audit areas to illustrate how LSOA boundaries were modified (note the street maps used during the audits had a higher level of detail than the street maps shown). While the audit areas were not strictly confined to LSOA boundaries, no audit area boundary deviated so substantially from the LSOA boundaries that the environment type classification (e.g. ‘urban deprived’) was likely to be invalid.

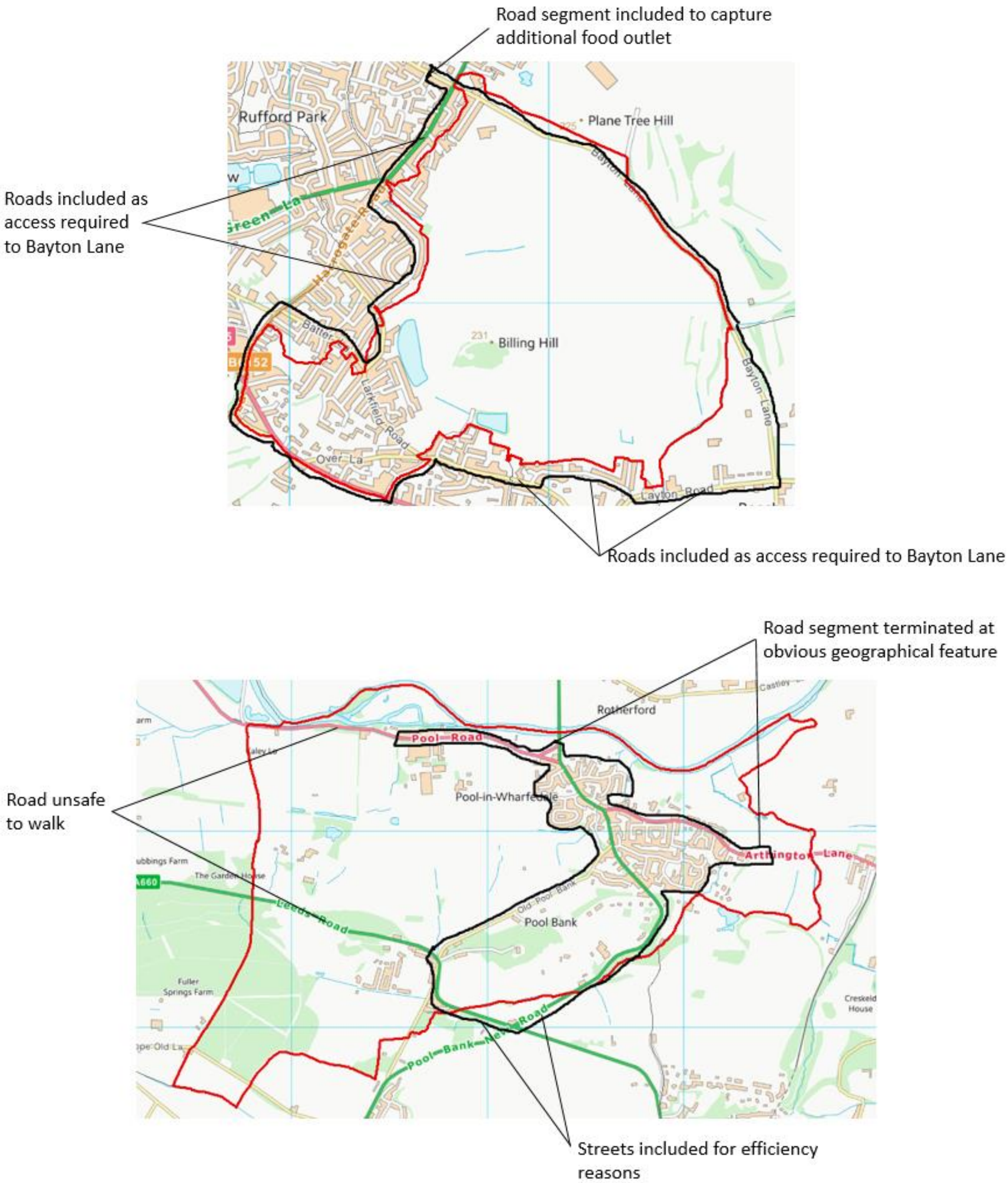

Figure 2. Maps showing LSOA boundaries (red) and audit area boundaries (black). Reasons for inclusion/exclusion of street segments also shown. LSOA boundary data from Office for National Statistics [38]. Base-map from Ordnance Survey [42].

2.4 Cleaning duplicates

Data cleaning was performed after the data matching process. Entries coded as FP (i.e. false positives - entries within the secondary data, but not identified in the audits) were visually examined to identify duplicates. Duplicates were identified as any two (or more) outlets within an Expected Outlets List (either POI or FSA) with substantially similar outlet names and matching addresses, with agreement between the geographical coordinates of the outlets also being checked when names deviated (e.g. 'Coriander Cuisine' and 'Curriander Cuisine'). If a pair of duplicate entries in the Expected Outlets List was determined to match an outlet in the Audit List, then the duplicate entry whose proprietary classification best matched the broad classification of the Audit List entry was coded as a true positive and the other was coded as a false positive. For example, for a pair of duplicate entries respectively classified as 'Cafés, snack bars and tea rooms' and 'Delicatessen' in the POI data, and matched to an outlet classified as 'Café' in the audits, the first entry would be coded as a true positive, and the second was identified as a duplicate and deleted.

2.5 Audit Classification Scheme

1. Restaurant		
1.01	Traditional	Sit down restaurant
		Waiter/waitress takes your order
		Pay for meal after eating
1.02	Buffet	Sit down restaurant
		No waiter service
		May pay at the till after food has been selected from the buffet but before eating
		If 'all you can eat' at a fixed price may pay before or after consumption. Drinks may or may not be included in the price.
1.03	Restaurant with takeaway/delivery option	Primarily a restaurant but has the option to order for takeout
		Waitress/ waiter service or Food is ordered and paid for at the counter and eaten elsewhere
		Usually open after 5pm
		Examples include Chinese restaurants, Indian restaurants, pizza hut
1.04	Fast Casual (e.g. Nandos)	Order and pay for food at counter
		Waitress/ waiter delivers food to table
		Similar to fast food but offers a higher quality of food and atmosphere
		Usually sit down but may have takeaway option
1.05	Pub Sit down restaurant	Sells predominantly alcohol
		Sit down restaurant
		Waiter/waitress takes your order
		Pay for meal after eating
1.06	Pub Fast casual	Sells predominantly alcohol
		Order and pay for food at bar. Waitress/ waiter delivers food to table
		Similar to fast food but offers a higher quality of food and atmosphere
		Sit down only not takeaway

1.07	Pub with takeaway/delivery option	Primarily a pub but has the option to order for takeout
		Waitress/ waiter service or food is ordered and paid for at the counter and eaten elsewhere
1.08	Traditional Hotel	Restaurant with waiter service
		Light bar meals with/without waiter service
		Room service and banqueting rooms
		May have a buffet for selected meals (e.g. breakfast)
2. Pub		
2.01	Pub no food	Only alcoholic and non- alcoholic drinks served.
		May served crisps and nuts behind the bar
		Includes nightclubs
3. Cafe		
3.01	Traditional café	Predominantly coffee and hot beverages sold
		Informal seating area
		May have waiter service or order at the counter
		Pre-made/made to order sandwiches and confectionery available
3.02	Greasy spoon types cafe	Predominately less healthy fried foods
		Informal seating area
		May have waiter service or order at the counter
3.03	Specialist café	Includes milkshake/smoothie bars and ice cream shops
		Similar in style to cafes and coffee shops
		Informal seating area
		Fair trade cafes/coffee shops are included here
3.04	Café with delicatessen/bakery	Predominantly café with delicatessen/bakery counter enabling ready-to-eat foods to be taken away
		Informal seating area
3.05	Sit-in sandwich shop	Small seating area
		Order and pay at the counter
		Made to order sandwiches/salads etc. May sell drinks, branded snacks and homemade cakes
		No waiter service
		Sit down or takeaway
4. Fast Food		
4.01	Takeaway café	Predominantly coffee and hot beverages sold
		No seating - takeaway only
		Pre-made/made to order sandwiches and confectionery available
4.02	Greasy spoon types cafe	Predominately less healthy fried foods
		No seating - takeaway only
4.03	Specialist café	Includes milkshake/smoothie bars and ice cream shops
		Similar in style to cafes and coffee shops
		Takeaway only
		Fair trade cafes/coffee shops are included here
4.04	Traditional sandwich shop	Made to order sandwiches/salads etc.

		May sell drinks, branded snacks and homemade cakes
		No sit in option - takeaway only
4.05	Internet Cafe	
5.01	Baker - Retail	Freshly baked savouries/bread, pre-made sandwiches, baked sweet products and branded products
		Usually a chain e.g. Greggs, Milligan's, Bakers Oven but can be independent
6.01	Traditional takeaway	Hot food ordered and paid for at the till
		Wait whilst food is prepared and cooked
		No sit down option to eat-in but may have a seated waiting area.
		Usually open after 5pm
6.02	Traditional takeaway + delivery/collection	As traditional plus: The option to telephone for delivery and/or collection
6.03	Traditional takeaway + delivery/collection	As traditional plus: Limited seating is available giving the option to eat-in
	With seating	May have the option to telephone for delivery and/or collection
6.04	Instant fast food	Food ordered and paid for at the till
		Available instantly as commonly cooked in bulk in advance and kept hot. Food that can be eaten without cutlery
		Sit down, takeaway and drive-thru facilities
		May be part of a chain or franchise
7. Supermarket		
7.01	Large multiple	Large, departmentalised, self-service food store selling food and household goods
		E.g. Tesco, Asda, Morrisons, Sainsburys, Co-op (large), M&S Simply Food (large), Waitrose
7.02	Discount	E.g. Kwiksave, Netto, Lidl, Aldi, Farmfoods, Fultons Foods, Iceland
7.03	Small multiple	Smaller, self-service food store selling a limited range of food and household goods for greater convenience
		Provides a wider and more consistent supply of fresh produce (e.g. fruits, vegetables, meats, dairy) than traditional convenience stores.
		Not restricted by Sunday trading hours laws.
		Includes small 'local' retailers owned by large multiple companies: Tesco metro/express, Sainsbury's Local, Little Waitrose, Morrison's My Local, Budgens, Co-op (small), M&S Simply Food (small) Also includes large chain convenience retailers e.g. Nisa/Premier/Spar/Best-One/Costcutter/Londis.
8. Convenience		
8.01	Traditional (corner shop)	Sells groceries, newspapers/magazines, snacks, drinks, lottery, tobacco products and sometimes pre-packed sandwiches
		Small and usually independently owned, although includes small Nisa/Premier/Spar
		Usually have extended hours
		Usually found in more residential areas
8.02	Newsagents	Small in size
		Sells primarily newspapers, magazines, snacks, drinks and tobacco products
		In well-trafficked public places
8.03	Petrol Station Shop	Sells groceries, newspapers/magazines, snacks, drinks, lottery, tobacco products and sometimes pre-packed sandwiches
		Usually have extended hours
		May be a small multiple supermarket
8.04	Off-licence	Licensed to sell alcoholic beverages for consumption off the premises

		Also sells groceries, newspapers, magazines, snacks, drinks and tobacco products.
9. Speciality (Purchase to takeaway only, includes permanent market stalls – e.g. a market stall selling fruits/vegetables should be classed as a greengrocer)		
9.01	Organic food stores	
9.02	Health food stores	Health supplements
		No fresh foods
9.03	Fair Trade stores	
9.05	Artisan Food Stores	Stores selling only locally produced goods
9.06	Delicatessen	Grocery type store.
		Sells fresh ready-to-eat foods (made to order sandwiches/salads, cooked meats and cheeses etc.)
9.07	Wine Merchant	E.g. Majestic, Oddbins
9.08	World food (All sizes)	E.g. Oriental, Indian and Continental shops and supermarkets
9.09	Candy/sweet / chocolate shops	Shops that do not fall under the category of convenience or confectioners as sell only bought in sweets
9.10	Butcher	Fresh meat is prepared and sold in store
9.11	Baker	Bread and baked products prepared fresh and sold in store
		Usually independent bakeries
9.12	Fishmonger	Fresh fish is prepared and sold in store
9.13	Greengrocer	Sells fresh fruit and vegetables
9.14	Dry goods only/Weigh house	Dry good only, usually sold by weight

2.6 Example of allowable street naming discrepancy

Figure 3 shows an example of when a street naming discrepancy would be allowed when matching outlets found in the audits to outlets listed in the secondary data. In this example, an outlet listed in the secondary data as being located on Armley Road would be matched to an outlet having the same name/classification that was found in the audits to be located on Canal Street.

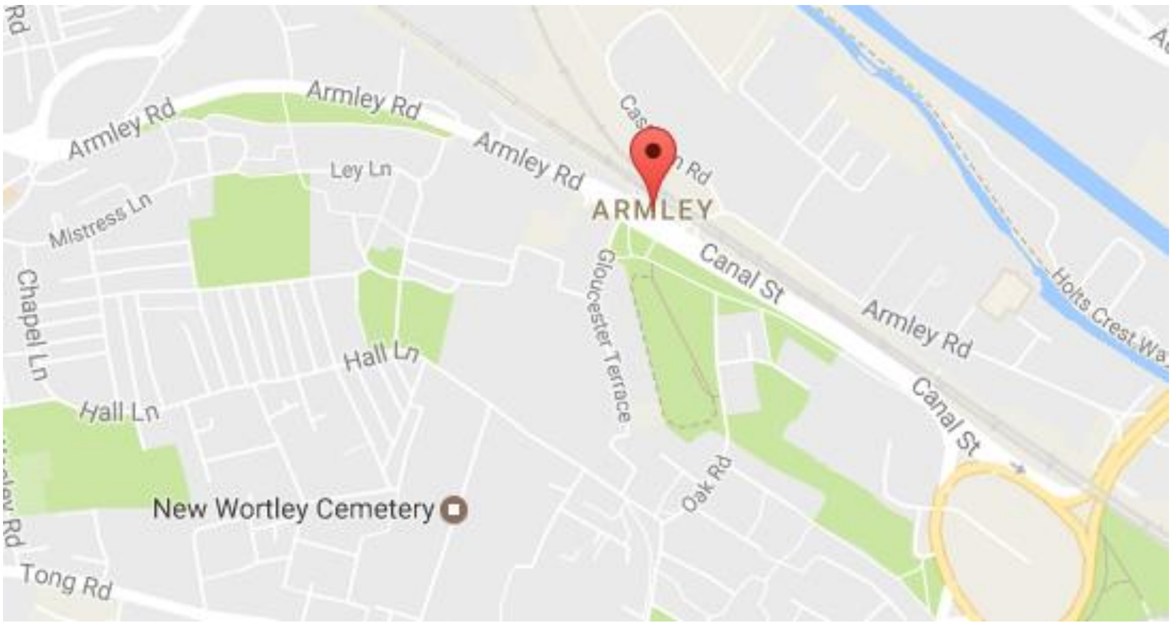

Figure 3. Map showing example of street having multiple names ('Armley Road' and 'Canal Street') [58].

3 Additional Descriptive Statistics

Table 12

Counts of outlets within the audits and secondary datasets and corresponding sensitivity and PPVs for each LSOA based on relaxed matching criteria.

LSOA Name	Audits	POI			FSA		
	Count	Count	Sens	PPV	FSA	Sens	PPV
County Durham	199	187	0.81	0.86	197	0.90	0.91
C. Dur - 033A	22	18	0.82	1.00	22	0.91	0.91
C. Dur - 038B	5	4	0.80	1.00	5	1.00	1.00
C. Dur - 038E	5	5	1.00	1.00	6	1.00	0.83
C. Dur - 046A	48	45	0.79	0.84	49	0.92	0.90
C. Dur - 051D	26	31	0.85	0.71	27	0.96	0.93
C. Dur - 059C	20	16	0.80	1.00	18	0.90	1.00
C. Dur - 059D	18	15	0.67	0.80	16	0.83	0.94
C. Dur - 066A	47	43	0.83	0.91	46	0.87	0.89
C. Dur - 066C	8	10	0.88	0.70	8	0.88	0.88
Calderdale	105	86	0.77	0.94	93	0.81	0.90
Calderdale 004E	63	56	0.84	0.95	55	0.81	0.91
Calderdale 007A	19	14	0.63	0.86	16	0.74	0.88
Calderdale 027A	7	5	0.71	1.00	8	1.00	0.88
Calderdale 027C	16	11	0.69	1.00	14	0.81	0.93
Leeds	795	768	0.81	0.84	726	0.83	0.91
Leeds - 005B	22	14	0.64	1.00	19	0.86	1.00
Leeds - 005D	2	2	1.00	1.00	1	0.50	1.00
Leeds - 007A	5	3	0.60	1.00	3	0.60	1.00
Leeds - 007C	10	9	0.90	1.00	10	1.00	1.00
Leeds - 007F	3	2	0.67	1.00	2	0.67	1.00
Leeds - 008A	13	12	0.92	1.00	14	0.85	0.79
Leeds - 009A	36	30	0.69	0.83	30	0.81	0.97
Leeds - 014B	15	12	0.73	0.92	16	1.00	0.94
Leeds - 014D	11	11	0.64	0.64	8	0.64	0.88
Leeds - 020C	26	28	0.92	0.86	20	0.73	0.95
Leeds - 021C	44	40	0.86	0.95	38	0.82	0.95
Leeds - 022C	3	4	1.00	0.75	3	0.67	0.67
Leeds - 027B	25	23	0.84	0.91	23	0.88	0.91
Leeds - 028E	17	15	0.88	1.00	17	0.94	0.94
Leeds - 030A	8	5	0.63	1.00	8	0.88	0.88
Leeds - 034C	42	46	0.88	0.80	45	0.95	0.89
Leeds - 048A	32	40	0.88	0.70	31	0.88	0.90
Leeds - 048C	7	8	0.86	0.75	8	0.86	0.75
Leeds - 048D	44	44	0.84	0.84	41	0.80	0.85
Leeds - 053B	22	23	0.77	0.74	20	0.82	0.90
Leeds - 053C	24	24	0.83	0.83	19	0.79	1.00
Leeds - 056C	7	7	1.00	1.00	8	0.86	0.75

Leeds - 056E	3	5	1.00	0.60	4	1.00	0.75
Leeds - 063D	31	26	0.77	0.92	31	0.97	0.97
Leeds - 065A	13	11	0.85	1.00	11	0.77	0.91
Leeds - 071B	26	18	0.65	0.94	22	0.85	1.00
Leeds - 071C	40	38	0.80	0.84	30	0.68	0.90
Leeds - 111A	78	85	0.86	0.79	78	0.85	0.85
Leeds - 111E	186	183	0.80	0.81	166	0.82	0.92
North Kesteven	73	59	0.78	0.97	65	0.85	0.95
North Kesteven 001A	3	3	1.00	1.00	3	1.00	1.00
North Kesteven 001B	1	1	1.00	1.00	1	1.00	1.00
North Kesteven 001C	8	7	0.75	0.86	8	1.00	1.00
North Kesteven 001D	6	6	0.83	0.83	5	0.83	1.00
North Kesteven 001E	5	3	0.60	1.00	5	1.00	1.00
North Kesteven 004C	7	5	0.71	1.00	6	0.86	1.00
North Kesteven 006B	12	10	0.83	1.00	8	0.67	1.00
North Kesteven 006D	11	9	0.82	1.00	8	0.64	0.88
North Kesteven 007D	4	2	0.50	1.00	5	1.00	0.80
North Kesteven 009C	16	13	0.81	1.00	16	0.94	0.94

Note. Sens: sensitivity

4 Additional Tables of Results

4.1 Strict Matching Criteria

Table 13

Odds of true positive relative to false positive (PPV odds) for POI data with strict matching criteria applied

Environment/ Outlet Type	Model 1			Model 2			Model 3		
	OR	95% CI		OR	95% CI		OR	95% CI	
Urban	REF						REF		
Rural	**1.86**[2]	**1.26**	**2.88**				**1.69**[1]	**1.00**	**2.92**
Deprived				REF			REF		
Middle				1.26	0.81	2.02	1.16	0.68	1.86
Affluent				**1.90**[2]	**1.17**	**3.14**	1.67	0.94	2.92
Restaurant							REF		
Pub							**0.33**[3]	**0.18**	**0.62**
Café							0.73	0.44	1.24
Fast Food							0.69	0.44	1.09
Supermarket							0.79	0.42	1.55
Convenience							**0.38**[3]	**0.21**	**0.67**
Speciality							0.80	0.43	1.53
Rural*Middle							1.20	0.52	2.90
Rural*Affluent							1.39	0.52	3.86

Note. OR: Odds ratio. CI: Confidence interval. REF: Reference category. All models are multi-level models accounting for nesting of outlets within LSOAs. [1]p<0.05, [2]p<0.01, [3]p<0.001

Table 14

Odds of true positive relative to false negative (sensitivity odds) for POI data with strict matching criteria applied

Environment/ Outlet Type	Model 1			Model 2			Model 3		
	OR	95% CI		OR	95% CI		OR	95% CI	
Urban	REF						REF		
Rural	0.93	0.71	1.23				1.34	0.85	2.13
Deprived				REF			REF		
Middle				1.17	0.86	1.57	1.43	0.95	2.15
Affluent				1.09	0.77	1.54	1.40	0.87	2.28
Restaurant							REF		
Pub							0.61	0.33	1.15
Café							**0.43**[3]	**0.28**	**0.64**
Fast Food							0.99	0.66	1.50
Supermarket							1.98	0.99	4.34
Convenience							**0.47**[2]	**0.29**	**0.76**
Speciality							0.72	0.43	1.21
Rural*Middle							0.61	0.32	1.18
Rural*Affluent							0.58	0.28	1.19

Note. OR: Odds ratio. CI: Confidence interval. REF: Reference category. All models are multi-level models accounting for nesting of outlets within LSOAs. [1]p<0.05, [2]p<0.01, [3]p<0.001

Table 15

Odds of true positive relative to false positive (PPV odds) for FSA data with strict matching criteria applied

Environment/ Outlet Type	Model 1			Model 2			Model 3		
	OR	95% CI		OR	95% CI		OR	95% CI	
Urban	REF						REF		
Rural	1.45	0.99	2.22				1.91	0.97	3.88
Deprived				REF			REF		
Middle				1.18	0.77	1.85	1.29	0.66	2.49
Affluent				1.47	0.90	2.48	1.42	0.70	2.91
Restaurant							REF		
Pub							**0.20**[3]	**0.10**	**0.42**
Café							**0.53**[1]	**0.28**	**1.00**
Fast Food							1.00	0.52	1.91
Supermarket							2.30	0.75	10.11
Convenience							**0.19**[3]	**0.10**	**0.35**
Speciality							0.54	0.25	1.19
Rural*Middle							0.69	0.24	1.98
Rural*Affluent							1.06	0.33	3.49

Note. OR: Odds ratio. CI: Confidence interval. REF: Reference category. All models are multi-level models accounting for nesting of outlets within LSOAs. [1]p<0.05, [2]p<0.01, [3]p<0.001

Table 16

Odds of true positive relative to false negative (sensitivity odds) for FSA data with strict matching criteria applied

Environment/ Outlet Type	Model 1			Model 2			Model 3			Model 3 (urban only)			Model 3 (rural only)		
	OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI		OR	95% CI	
Urban	REF						REF								
Rural	**1.42**[1]	**1.01**	**1.99**				**2.55**[2]	**1.47**	**4.57**						
Deprived				REF			REF			REF			REF		
Middle				1.05	0.69	1.60	1.27	0.81	2.05	1.25	0.78	2.17	**1.25**[1]	**0.27**	**0.98**
Affluent				0.98	0.64	1.50	1.15	0.70	1.92	1.13	0.67	1.91	1.13	0.33	1.29
Restaurant							REF			REF			REF		
Pub							**0.51**[1]	**0.26**	**1.01**	**0.34**[2]	**0.15**	**0.77**	0.34	0.38	5.93
Café							**0.61**[1]	**0.38**	**0.98**	0.58	0.33	1.03	0.58	0.31	1.58
Fast Food							1.08	0.67	1.75	1.00	0.56	1.80	1.00	0.56	2.98
Supermarket							1.68	0.78	4.05	1.29	0.54	3.46	1.29	0.78	77.39
Convenience							**0.36**[3]	**0.21**	**0.60**	**0.33**[2]	**0.17**	**0.64**	**0.33**[1]	**0.18**	**0.99**
Speciality							**0.38**[3]	**0.22**	**0.66**	**0.33**[2]	**0.17**	**0.65**	0.33	0.22	1.30
Rural*Middle							**0.40**[1]	**0.18**	**0.86**						
Rural*Affluent							0.52	0.22	1.20						

Note. OR: Odds ratio. CI: Confidence interval. REF: Reference category. All models are multi-level models accounting for nesting of outlets within LSOAs. [1]$p<0.05$, [2]$p<0.01$, [3]$p<0.001$

4.2 Classification Agreement

Table 17

Percentage Agreement for Broad Classifications Based on Primary and Alternate Classification Schemes

		Points of Interest						
		Rest	Pub	Café	FF	Sup	Conv	Spec
Audit	Rest	**120**	117	3	5	0	0	0
	Pub	3	**42**	0	0	0	0	0
	Café	5	2	**83**	22	0	0	9
	FF	10	1	7	**213**	0	0	4
	Sup	0	0	0	0	**43**	22	6
	Conv	0	0	0	0	1	**69**	2
	Spec	0	0	0	10	0	13	**59**
	%Agree	**46%**	**26%**	**63%**	**78%**	**60%**	**65%**	**57%**

Note. Rest: Restaurant; FF: Fast Food; Conv: Convenience; Spec: Speciality; %Agree: percentage agreement for broad classifications. Numbers in bold indicate the counts of outlets for which the POI-derived and the audit-derived classifications agreed. Numbers in red indicate the counts of outlets for which the POI-derived and the audit-derived classifications disagreed.