

The Foodscape: classification and field validation of secondary data sources

Amelia A Lake^{a,b}
Thomas Burgoine^c
Fiona Greenhalgh^b
Elaine Stamp^d
Rachel Tyrrell^b

^aApplied Biosciences, School of Applied Sciences, Northumbria University, NE1 8ST

^bHuman Nutrition Research Centre, Institute of Health & Society, Newcastle University, UK, NE2 4HH

^cHuman Nutrition Research Centre & Geography, Institute of Health & Society, Newcastle University, UK, NE2 4HH

^dInstitute of Health & Society, Newcastle University, UK, NE2 4HH

Corresponding author: Amelia.lake@northumbria.ac.uk

Senior Lecturer in Food & Nutrition
Applied Biosciences,
School of Applied Sciences,
Northumbria University,
Ellison Building,
Newcastle upon Tyne
NE1 8ST.
Telephone: +44191 227 4263
Fax: +44191 277 3519

Acknowledgements

The authors would like to acknowledge The Public Health and Environmental Protection Regeneration Directorate at Newcastle City Council, The Chartered Institute of Environmental Health and Customer Services at Yellow Pages. The authors would like to acknowledge their colleagues in the wider Obesogenic Environments Group at Newcastle University, in particular Professor Ashley Adamson.

Funding

Amelia A Lake was funded by NIHR (National Institute for Health Research) Personal Awards. The views expressed in this publication are those of the author and not necessarily those of the NHS, NIHR or Department of Health. Rachel Tyrrell was funded by a Rank Fund Vacation Studentship. Thomas Burgoine was funded by an ESRC quota PhD Studentship.

The Foodscape: classification and field validation of secondary data sources

Abstract

The aims were to; develop a food environment classification tool and to test the acceptability and validity of three secondary sources of food environment data within a defined urban area of Newcastle-Upon-Tyne, using a field validation method.

A 21 point (with 77 sub-categories) classification tool was developed. The fieldwork recorded 617 establishments selling food and/or food products. The sensitivity analysis of the secondary sources against fieldwork for the Newcastle City Council data was good (83.6%), while Yell.com and the Yellow Pages were low (51.2% and 50.9% respectively).

To improve the quality of secondary data, multiple sources should be used in order to achieve a realistic picture of the foodscape.

Keywords:

Food environment, Foodscape, classification, secondary data, obesogenic environments

Introduction

Unhealthy diets are linked with numerous chronic diseases including obesity, which has been described as one of the largest societal challenges we face (Foresight, 2007). In relation to obesity, environmental factors influence both sides of the energy balance equation; energy intake, in terms of the food environment, and energy expenditure, describing physical activity and the environment. In terms of disease prevention and the promotion of health it is important to establish *which* aspects of the food environment are amenable to change. This paper will focus on the food environment, a relatively new field of research (McKinnon et al., 2009).

Environmental exposures such as the availability and accessibility of food interact with individual factors, such as taste, familiarity/habit and health to drive food choice (Contento et al., 2006). The relationship between the food environment and obesity is complex (Wang et al., 2006). While there is little research linking food access with obesity as an outcome measure in any age group (White, 2007), understanding the relationship between what we eat and the environmental context in which these food choices are made is essential to the development of long term obesity prevention strategies (Holsten, 2009).

Food choices are made within our respective food environments. The food environment, or 'foodscape', encompasses any opportunity to obtain food and includes physical, socio-cultural, economic and policy influences at both micro and macro-levels (Townshend and Lake, 2009; Lake and Townshend, 2006). Glanz et al

(2005) have hypothesised the food environment to incorporate four different elements; community (type and location of food outlets); consumer (availability of healthy options, price, promotion and nutritional information); organisational (home, school and workplace); and informational (media and advertising). This paper focuses on the community food environment and the reality of the 'foodscape'.

Collecting data on the food environment relies on the identification of the specific type of outlet and its location (Story et al., 2008). This identification of the type of food outlet can be described as the classification of the food outlet. Research has tended to develop classification systems dependent upon the data sources available to them. Studies have used existing classification systems such as local council lists (Macdonald et al., 2009), government agency lists (Wang et al., 2008) as well as commercial directories (Burgoine et al., 2009) or combinations of these (Wang et al., 2008). While studies that involve visiting the food outlets can base their classification on the availability of certain foods within in these outlets (White et al., 2004). In order to look at the neighbourhood food store environment in the US Wang et al (2008) adapted the North American Industry Classification System and the Food Marketing Institute definitions to create 11 categories of food outlets including grocery stores, bakeries and pizza shops. They used the secondary data sources available to them to classify the food outlets using their 11 point tool. Morland and Evenson (2009) also modified the North America Industry Classification System codes to describe the types of food stores and food service places located in their US based study area. They used 10 categories in total; five categories for food stores and five for food service places (i.e. restaurants, fast food restaurants etc). In the UK, Burgoine et al (2009) based their three point classification system of the longitudinal food environment on the Yellow Pages classification system which was their data source. White et al (2004) developed 19 categories for shops selling food and described the difficulties in distinguishing between shop types. However, as they were visiting the outlets their classifications could be based on the foods available in these outlets. This is a new area of research and few studies have reported on the development of the food outlet classification system which has been used.

Methods to identify the type of outlet and its location can either be 'intermediate' or 'direct' (Booth et al., 2005). Intermediate refers to the use of secondary data sources (e.g. telephone directories, business directories, government listings), for example, in the UK, Burgoine et al (2009) used paper Yellow Pages to establish the historical food environment retrospectively and Maddock (2004) used the US Yellow Pages to identify fast food outlets. The direct method often refers to an audit of a specific area by trained observers (Booth et al., 2005), as used by several UK studies (White et al., 2004; Donkin et al., 2000). This direct method can also be referred to as 'ground truthing' (Sharkey, 2009) or as 'field validation'.

In the US Wang et al (2006) conducted a study that explored the strengths and limitations of historical food environment data from government and commercial sources. The authors urged caution when choosing sources of secondary data, and in current (rather than historical) studies recommend the use of observation

techniques to assess the validity of secondary data sources. Similarly, in Canada, Paquet et al (2008) aimed to establish the validity of sources of data on food stores and physical activity establishments that were obtained from commercial database and Internet searches. However, Paquet et al (2008) did not identify or validate restaurants or takeaways (considered in this current study) and so fell short of assessing representations of the total food environment.

In the UK, most types of food business (including mobile vans, catering businesses run from home) need to be registered with the environmental health service at the relevant local authority at least 28 days before they plan to open (Food Standards Agency, 2008b). After registration environmental health protection officers will visit the business and then carry out routine visits. The regularity of these visits depends on the risk category of the business and can vary from every 6 months, every 12-18 months up to every 3 years or more frequently as the result of a complaint (Food Standards Agency, 2008a). There is a legal requirement and legislation for businesses to notify the local authority of closures (The European Parliament and of The Council of the European Union, 2004). Therefore Public Health and Environmental Protection Records held by local authorities are constantly being updated and should be a useful and fairly accurate source of secondary data when exploring the food environment.

This study had two aims; firstly to develop a culturally relevant and detailed system of classifying the entire food environment (foodscape) for our current and planned research. This classification system would be used to record and describe the foodscape. The aim was to develop a tool that could be used to classify the food environment from both direct field observations and secondary data sources. This tool is urgently required for studies in the UK to have a classification system that encompasses the retail (shops) and service (restaurants) food environment. Secondly to test the acceptability and validity of three secondary sources of food environment data (Public Health and Environmental Protection data obtained from Newcastle City Council and two forms of commercial search directories [Yellow Pages and Yell.com]) within Newcastle-Upon-Tyne, using a field validation method.

Being able to identify and classify the community food environment (type of outlets) is an important first step for most studies exploring the relationship between the food environment, food intake and adiposity, particularly in relation to the topic of obesogenic environments. Rigorous measures of the food environment are needed in order to explore links between the food environment, individual intake and health indicators (Lytle, 2009). The work described in this paper is the first UK study to explore the field validation of secondary data pertaining to the existence of food outlets from commercial records (Yellow Pages) and from a Local Government Source (Public Health and Environmental Protection Records).

Methods

Classification of the food environment

A literature review was conducted using Ovid Medline with the following search terms: business directories, yellow pages, food environment and food availability/access. In addition searches of existing classification systems for example those used in research studies, those used by commercial organisations (such as the yellow pages) and the classification system used by the local authority.

This review identified problems in incorporating the large variety of food outlets present. Previous work tended to focus on specific types of outlet, for example retail outlets that sell food (like convenience stores or supermarkets) or focus on service or out-of-home food outlets (like restaurants or fast food outlets). Few attempt to classify *all opportunities* to obtain food and provide a tool that can capture the entire food environment or foodscape. Of the other classification systems examined, Yell.com (2008) and Newcastle City Council (2008) had the most extensive and detailed classification systems with 71 and 88 categories respectively, allowing detailed scrutiny of the food environment. However the classification systems used within these data sources were too expansive for our purposes and included classification categories not open to the public (for example staff canteens). The literature review and evaluation of commercial and local authority classification systems revealed that the classification methods used to date did not meet the aims of this study and those of our planned research. Given the lack of suitable tools we set out to develop a new food outlet classification system, which could be used by our team and other researchers in the UK to build a picture of the food environment.

While the classification system took its basis from existing classification systems (mainly Newcastle City Council (2008) and Yell.com (2008)) the field validation element of the study (described below) helped to further develop the classification system by allowing further insight into the types of food outlets present and the kinds of food vended within these outlets. During the field validation study, photographs were taken of the outside of a number of businesses to aid with the classification process and to help define categories as well as sub-categories.

After much consideration and discussion a new comprehensive 21 point food outlet classification tool was developed (Table 1). This tool includes 77 sub-categories (not listed, but available on request from the author), which may be used when collecting 'direct' data and the researcher is present in the outlet. For example 'restaurant' is a main heading and the sub-headings 'traditional', 'buffet', 'restaurant with take-away options' and 'fast casual' allow researchers to differentiate types of restaurant.

Mobile food outlets (mobile fast food vans) and market stalls were excluded from the classification system as they are registered at the place where they are normally kept and the places where they usually trade are not necessarily recorded.

Testing the acceptability and validity of three secondary (intermediate) data sources

The quality of secondary data sources on the food environment is often questionable because of human error and the time investment required to ensure high levels of accuracy (Brownson et al., 2009). This quality is however required or else we risk having an oversimplified picture of the 'real world' (Evenson et al., 2009). It is important that secondary data has been tested for its validity in representing the 'actual' food environment as well as being both acceptable and accessible. McKinnon et al (2009) have emphasised the need for more research in the field of food environments to focus on validity.

Two different methods of researching the food environment within a defined geographical area of Newcastle-Upon-Tyne, UK are described below (one direct and three intermediate [secondary]). The study was conducted within Lower Super Output Area (LSOA) boundaries, small area geographies (smaller than electoral wards for example) that are relatively homogeneous in containing approximately 1500 individuals. Two adjacent LSOAs were selected as the foci for this study – Newcastle-Upon-Tyne 024C and 024F. Both areas are technically 'urban' under the Commission for Rural Communities classification framework (2004), and are similar in size (1373.38m² and 1340.71m² respectively, according to the General Land Use Database (GLUD) (Office for National Statistics, 2005). They are however quite different in terms of their socio-economic characteristics, falling in the first and third quintiles of deprivation (relative to the rest of England) respectively. Despite this, their relatively diminutive size, ease of accessibility and robust food outlet density (101.94 and 281.94 food outlets/km² respectively) led to their selection for this research. These areas made up the 'study area'.

Before conducting the fieldwork, the food outlets from each secondary data source were reclassified according to the 21 point classification system. When it was difficult to classify an outlet based on the secondary information alone, we searched online for a specific outlet at the specific address to obtain further details about the kind of food this outlet was selling, as has been done previously (Morland and Evenson, 2009).

Secondary Data sources

Newcastle City Council

The most up-to-date data from the Public Health and Environmental Protection department of Newcastle City Council (2008) were obtained in Summer 2008 for use within the study. The council data included the names, addresses and postcodes of outlets registered as selling food and/or food products in Newcastle, coded according to the Food Standards Agency's classification system (2008a). The value of such data is that it is readily available under the UK Freedom of Information Act (Office of Public Sector Information, 2000).

Commercial directories

Two forms of freely available commercial search directory were used to determine the food establishments present in each LSOA. The directories used were the 2008/2009 Newcastle-Upon-Tyne Yellow Pages (henceforth referred to as the 'Yellow Pages') published in July 2008 and the Yellow Pages online (Yell.com) accessed in August 2008, henceforth referred to as 'Yell.com'.

The search terms used to locate outlets selling food and/or food products were derived from the Yellow Pages index of classifications (n=71). Existing Yellow Pages index categories were used as search terms and are identical for both online and paper directories. The use of these terms ensured that all food outlets were identified and catalogued.

A company can specify to be listed under multiple classification categories, which can result in a gross overestimation of the number of outlets present, as described by Wang et al (2006). To overcome this potential problem here, and to reduce count error, if the same store was listed under multiple classification categories the outlet was only included once.

Yellow Pages

The Yellow Pages were hand-searched to determine the food outlets present in each defined area. The majority of outlet listings included a number or building name, street name and area but did not contain postcodes. As postcodes could not be used, outlets listed under each specific classification category were searched for by street name on the maps of each LSOA. These maps needed to be adequately detailed and were created using individual tiles downloaded from Digimap (an online collection of Ordnance Survey Maps) (EDINA, 2009). These were subsequently pieced together in PowerPoint to create a map with sufficient resolution to identify street level detail so that outlet presence within the specified areas could be clearly determined.

Yell.com

An electronic search of Yell.com was carried out using the same classification terms used for the Yellow Pages search. The search engine gave the option to search by classification, and/or company name and location. The classification categories were entered along with each LSOA postcode prefix in turn in the location search (NE1, NE2, NE4). As postcodes were included in the results generated, all the results found with the specific postcode prefixes were searched for by street name to ensure they were within the LSOA boundaries on each map. The website also gave the option to view the location of food outlets on a map, which was a useful tool in determining the specific location of the outlet.

Field validation (ground truthing)

Maps of the study area were produced using Digimap (EDINA, 2009) and a fieldwork route was established. The maps produced were of sufficient resolution, however some of the smaller streets did not include names. To guide the field validation the

researchers compiled a list of outlets as recorded by council data. Any changes to the council data were recorded. This included any closures, name changes or new establishments present. If the street listed for an outlet during the search could not be found on the map, Google Maps (Google) was used to ascertain the location of the outlet in relation to the boundaries of the LSOA. This method was also used if a street crossed the LSOA boundary, in which case the outlet was searched for by premises number and address. This would ensure the correct positioning on the particular street was found and hence the correct positioning with regards to the LSOA boundary.

Researchers developed route plans from the maps in order to systematically cover the entire study area. In August 2008 field validation was conducted, on foot, by two researchers between normal business opening hours. The researchers recorded, on paper, all food outlets present within the defined study area. Food outlets had to be open to the public to be included, for example, staff canteens were not included. The investigation took a total of three days to complete. Researchers were only required to enter premises if sufficient information pertaining to the name or type of outlet could not be gathered from the street.

Analysis

Positive predictive values (PPV) were calculated for each data source in comparison to fieldwork. This was done by generating 2x2 tables in Stata V9 (StataCorp, 2008) described in Table 2. The PPV was then calculated manually as true positives divided by true positives plus false positives (see equation Table 2). PPV is a measure of precision and relates to the percentage of outlets found in the three data sources that were actually present in the fieldwork. PPV focuses on outlets present in the data sources. Therefore outlets present in the field that are not in the data source are not taken into account.

Sensitivity analysis assesses the proportion of outlets correctly identified by the data source that were actually present in the field (see Table 2). Sensitivity was assessed between data sources with the fieldwork considered the 'gold standard' to which the council data, Yell.com and the Yellow Pages were compared. For large geographic areas it is often unfeasible to collect direct data, therefore knowing how sensitive secondary data sources are is very important. Sensitivity categories were adapted from percentage agreement categories used by Paquet et al (2008) in their earlier field validation work (see Table 3).

The 21 point classification tool provided a description of the types of outlets that were present in the area studied. Sensitivity by outlet type was also assessed in order to determine whether sensitivity was higher according to type of establishment and/or data source.

The secondary data sources were used in concurrence to investigate the best alternative method to fieldwork (which is both labour and time intensive). The

sensitivity of fieldwork and council data, Yell.com and Yellow Pages collectively was assessed.

Results

Fieldwork

The fieldwork recorded a total of 617 establishments selling food and/or food products within the study areas. The frequency of types of outlet are illustrated in Table 4. The most common food outlet types in the study areas were Pub/Bar (n=136, 22.0%), Restaurant (n=112, 18.2%) and Café/Coffee Shop (n=69, 11.2%). The classifications with a frequency of <5 included Supermarket, Health and Leisure, Pizzeria, Discount Stores and Food Production Services (this category included outlets such as wholesalers, suppliers and 'cash and carries').

While the fieldwork observations recorded a total of 617 outlets, the Newcastle City Council data listed 564 outlets, with 393 listed in the Yellow Pages and 397 listed on Yell.com. There were an additional 141 outlets listed in the three data sources (combined) that were not present in the fieldwork (Table 5). The secondary data source that listed the most outlets, which were *not* present in the fieldwork was Yell.com (58.9%) with the council data having the lowest percentage of outlets not found in the fieldwork (34.0%).

The most frequent missing classifications in the fieldwork were pub/bar (20.6%) and restaurants (19.1%) (Table 6). There were no additional department stores, discount stores, pizzerias or health and leisure stores listed in the data sources compared with those found in the fieldwork.

Positive Predictive Value Analysis

The PPV discounts additional outlet listings found in the data sources and only focuses on the outlets present in the data sources and present in the fieldwork (n=617). An ideal PPV would be 100% i.e. all outlets identified in the data source were actually present in the field. The council data had a PPV of 91.5%, while the Yellow Pages and Yell.com had PPV's of 82.4 and 79.1 respectively (Table 7).

Sensitivity analysis

Using the fieldwork as the gold standard, sensitivity was assessed between the three secondary data sources. Sensitivity categories were adapted from percentage agreement categories used by Paquet et al (2008) (Table 3). Sensitivity between the total number of outlets for the fieldwork and council data was high (83.6%) however for Yell.com and the Yellow Pages the results were moderate (50.9% and 52.5% respectively).

In addition to sensitivity analysis for the total number of outlets, sensitivity analyses were conducted separately for each of the 21 food classification categories.

However four classification categories did not have sufficient observations recorded (fast food, pizzeria, food production services, health and leisure and not classified) and the analysis could not be conducted. When looking at the sensitivity categories by classification group this analysis illustrated that the council data had a much higher proportion of results in the 'good/ excellent' categories compared with Yellow Pages and Yell.com (13 categories compared with, 3 and 1 respectively – see Table 3). The classification categories that had moderate – good sensitivity across all three data sources, were restaurants, pub/bar, convenience, medical (e.g. hospitals, pharmacies) sandwich shops and specialist outlets (e.g. organic food stores, health food stores). The only classification category to have good percentage sensitivity consistently across all three data sources were specialist outlets.

Using all three data sources together as an alternative method to field validation, was explored. The sensitivity between Yell.com, Yellow Pages, council data and fieldwork was excellent (92.9%).

The sensitivity analysis was repeated using the council data as the gold standard, Table 8 highlights the low levels of sensitivity between the council data and the commercial search directories. When the two commercial directories were used simultaneously the sensitivity was still categorised as moderate (56.0%) and only a 4.8% increase from using Yell.com alone was observed. There was a significant difference between the council data and both Yell.com and Yellow Pages (both $p < 0.001$) but no significant difference between Yell.com and Yellow Pages ($p = 0.569$).

Discussion

This study developed a detailed food outlet classification system derived from observations in the field as well as from existing classification systems to be used to describe the types of food outlets which are available to individuals and populations. Following the development of the tool this study explored the acceptability and validity of three secondary sources of food outlet data within a defined area of Newcastle-Upon-Tyne, using a field validation method.

Food environments are believed to play a significant role in the obesity epidemic and robust research methods are required to establish which factors or aspects of the food environment are relevant to food choice and to adiposity. In comparison to existing research that links physical activity, the environment and obesity the research that links food to the environment and obesity is relatively undeveloped (Townshend and Lake, 2009). McKinnon et al (2009) recently reviewed 137 articles that included measures of the food environment. When defining the food environment studies may use a combination of methods for example observations of the price of food, availability of foods. McKinnon et al (2009) describe the different types of methods (geographical methods, check lists, market baskets, menu analysis etc) used by researchers depending on their focus (macro or micro food

environments). For all these methods the classification of the food outlets is very often the first (and difficult) step in the research process. Describing food outlets and the food environment is culturally specific and dependent upon the data available. The food outlet classification tool described in this study has 21 main headings and 77 subheadings which enables it to be used at a very detailed level (using direct field observations) as well as to be used for secondary classification. With the high levels of policy interest around the food environment including the recently published Healthy Weights, Healthy Lives document (Cross-Government Obesity Unit, 2008), researchers and practitioners (both from health and planning) are looking for methods to describe and measure the food environment. The classification tool developed in this work will be a useful basis for researchers and practitioners across the UK.

While some studies do record 'direct' observations of the food environment, many use secondary data such as commercial or local government records. This study has shown that the data collected by Environmental Health Departments in local authorities and freely available under the Freedom of Information Act in combination with other freely available commercial listings provide a good source of food environment data. The data obtained during this investigation highlights that the sensitivity of field observations for the Newcastle City Council data was good (83.6%). Field validation (or direct observations) is both labour and time intensive and may not be suitable for all study protocols or for large study areas; the sensitivity figure (83.6%) illustrates that local authority data would be an acceptable alternative to assess the food outlets present in a particular area. The commercial search directories were a poor substitute for fieldwork as only half the outlets present in reality were found during the commercial databases searches. The council data had a positive predictive value (PPV) of 91.5%. An ideal PPV would be 100% therefore the council data is an acceptable alternative to conducting fieldwork.

Using Newcastle City Council data as a 'gold standard', sensitivity results with the commercial data sources were fair-moderate. Table 8 shows a significant difference in sensitivity between the council data and commercial data sources (Yell.com and Yellow Pages both $p < 0.001$). These results indicate that Yell.com and the Yellow Pages could not be used as an alternative to council data whether used individually (51.2% and 50.9% respectively) or collectively (56.0%).

Our analysis illustrated that multiple data sources (local authority in addition to commercial directories) should be used in order to achieve an acceptable and valid alternative to field validation. The sensitivity when using all three data sources together in comparison to fieldwork was excellent (92.9%) and therefore would be a more reliable method to use than council data alone (83.6%).

Sensitivity was calculated by data source (total) and broken down by classification category. This analysis provided an insight into the type of outlets listed in each data source that were present in reality. It is useful, particularly in the commercial search directories to see whether certain outlet types are more likely to advertise their company. Restaurant and pub/bar were the most frequently observed outlets in the

area studied (112 and 136 outlets respectively) and also were found to have a moderate to good percentage sensitivity range across the data sources.

Strengths and limitations

Two small area geographies (LSOAs) were selected and combined to create the 'study area'. As far as possible, these areas were selected based on their homogeneity - they were similar in physical size, they were uniformly urban, and they both contained high numbers of food outlets per square kilometre. Further reasons for their inclusion included their proximity to the research group and to each other, and their compact dimensions. While it may be preferable for further study to focus on areas that are homogeneous in terms of their socio-economic standing, these results are not compromised as a result of the variance shown here in this regard. The sample numbers would not have been substantial enough to support a stratification of the analysis by high/low socio-economic status. The ability to generalise these findings should extend to UK urban areas, but future research should focus on the rural food environment and the ease/difficulty of collecting food environment data therein, which has been relatively unexplored to date.

The advantages of using the Yellow Pages and Yell.com as commercial data sources are that, food outlets can be listed for free (alphabetically) or they can pay to have a larger advertisement in the directories and they are updated annually. This results in a large variety of outlets being incorporated including small outlets with low revenue. In the US, Wang et al (2006) used commercial databases which generally did not include retail food stores with an annual sales volume of under \$500,000. This would not provide a true representation of the stores in a particular area as small retail food stores may not be included.

In this analysis the same classification terms were used for both commercial data sources, which improved reliability when searching outlets. Companies can choose the classification term under which they are listed instead of being grouped under large categories used by other data sources. This gives a better and less superficial representation of the type of outlet present and aids with researcher classification. The Newcastle City Council classifications were broad and without visual validation it would be difficult to determine type of outlet from within the council category.

The Yellow Pages are published annually and therefore any outlet changes during this period are not reflected in this paper volume. Therein the main limitation of printed search directories is that they cannot be automatically changed to account for closures, changes of name and premises, and can only be updated when a new version is released. The annual publication date is area specific (Yell.com Customer Services, 2009).

Existing customers of the Yellow Pages and Yell.com (paid or free-listed) are contacted by a sales assistant annually (before publication date for the Yellow Pages or a year after registration for Yell.com customers) and are asked if they would like to extend their advertisement, change their contact details or be removed from the

listings (Yell.com Customer Services, 2009). This ensures a reasonable amount of confidence regarding the outlets that are listed close to the publication date; however the extent to which this is useful remains in doubt. A limitation of using commercial search directories (as was also found by Wang et al (2006)) is that store addresses can be that of the company Head Quarters and not the actual store address. This results in the data source count being higher than the actual field validation count.

Yell.com can be more easily updated as it is an online service. However, the onus lies with the advertised company to notify Yell.com of any changes within the 12 month contract period for updates to occur otherwise Yell.com only updates the details of each outlet with the renewal or cancellation of advertisements. A limited number of randomly generated results (100) are shown when a term or outlet name is searched online which may lead to problems when searching an area with a large number of similar types of outlets (Yell.com Customer Services, 2009). Paid adverts also have precedence over free listings leading to bias as paid adverts will appear first in outlet searches (Yell.com Customer Services, 2009).

The data obtained from Newcastle City Council is a reliable source to use as food outlets are required to register by law, report changes and to unregister, where necessary, with the Environmental Protection Officer within the local authority (Food Standards Agency, 2008a). This is not an optional listing unlike the commercial search directories.

The main limitation of the commercial data sources used is the frequency with which they are updated. This can account for a high proportion of outlets being listed in addition to those actually present in the fieldwork (Yellow Pages n=69, Yell.com n=83). Both commercial data sources are updated annually, however Yell.com can be updated immediately to amend or remove an existing listing or to add a new listing if informed by the customer. Customer services at Yell.com advised that more people are choosing to advertise using the online service in preference to the Yellow Pages. The reasons suggested for this shift were that customers perceive the Internet to be an easily accessible source for their client base as more people have access to the Internet than in previous years and customers are able to advertise at any time of the year with Yell.com as registration is instant (registration with the Yellow Pages has to be at specific time periods before publication) (Yell.com Customer Services, 2009). These higher rates of advertising may be an explanation for Yell.com having the highest frequency of outlets not present in the fieldwork in comparison to the other data sources (n=83).

There were 48 additional outlets listed in the council data that were not found in the fieldwork. Food businesses should notify the local authority of changes (The Chartered Institute of Environmental Health Policy Officer, 2009). Notifying the local authority is not often a priority for closing businesses however it is a legal requirement to register and notify change (Article 6 Regulation 852/2004) and it is therefore an offence not to conform to regulations (The Chartered Institute of Environmental Health Policy Officer, 2009). Registration is required 28 days before

opening an outlet selling food and/or food products (Food Standards Agency, 2008a). The time lag between an outlet registering and opening may be a reason for an increased number of outlets being present in the data source. Failure to register a business could be as a result of lack of awareness of the need to register or deliberate avoidance, perhaps through a disinclination for an inspection visit (The Chartered Institute of Environmental Health Policy Officer, 2009).

The effects of infrequent directory updates are intrinsic to the fact that a number of outlets (n=141) were not present in the fieldwork. This indicates that the three additional data sources used are not as up-to-date as the field validation method resulting in an exaggerated number of outlets present. An increased number of outlets may have been observed if the data sources used were from vastly different time points and this would further limit the results obtained. For validity of results the time lag must be as short as possible between obtaining data and obtaining results.

The development of the food outlet classification tool and exploring the validity and acceptability of secondary data sources have been the first steps within a larger body of work around the food environment, food choices and adiposity. Further work is currently underway to test the classification tool in urban and rural settings as well as areas of high and low deprivation to ensure the entire foodscape is represented in secondary data and can be classified.

Conclusion

Tackling the obesogenic environment and specifically the food environment requires a multi-disciplinary approach: this research drew on expertise from nutrition, policy and geography. Food environments are believed to play a significant role in the obesity epidemic and robust research methods are required to establish which factors or aspects of the food environment are relevant to food choice and to adiposity. Collecting and classifying data on the food environment relies heavily on the quality of the secondary data and researchers should, where possible, conduct some field validation in order to validate their data. These results have shown that multiple data sources must be used in order to achieve an acceptable and robust alternative to field validation. The data used must be from a range of sources in order to incorporate all food outlets, independent of revenue.

References

- Booth, K. M., Pinkston, M. M. and Poston, W. S. C. (2005) 'Obesity and the Built Environment', *Journal of the American Dietetic Association*, 105, (5, Supplement 1), pp. 110-117.
- Brownson, R. C., Hoehner, C. M., Day, K., Forsyth, A. and Sallis, J. F. (2009) 'Measuring the built environment for physical activity: state of the science', *American Journal of Preventative Medicine*, 36, (4), pp. S99-S123.
- Burgoine, T., Lake, A. A., Stamp, E., Alvanides, S., Mathers, J. C. and Adamson, A. J. (2009) 'Changing foodscapes 1980-2000, using the ASH30 Study', *Appetite*, 53, (2), pp. 157-165.
- Commission for Rural Communities. (2004) *What is Rural?* London: Countryside Agency
- Contento, I. R., Williams, S. S., Michela, J. L. and Franklin, A. B. (2006) 'Understanding the food choice process of adolescents in the context of family and friends', *Journal of Adolescent Health*, 38, (5), pp. 575-582.
- Cross-Government Obesity Unit, D. o. H. a. D. o. C., Schools and Families, . (2008) *Healthy Weight, Healthy Lives: a Cross-Government Strategy for England*. London:
- Donkin, A. J. M., Dowler, E. A., Stevenson, S. J. and Turner, S. A. (2000) 'Mapping access to food in a deprived area: the development of price and availability indices', *Public Health Nutrition*, 3, (01), pp. 31-38.
- EDINA (2009) *Digimap Collections*. Available at: <http://edina.ac.uk/digimap> (Accessed: 16/03/09).
- Evenson, K. R., Sotres-Alvarez, D., Herring, A. H., Messer, L., Laraia, B. A. and Rodriguez, D. A. (2009) 'Assessing urban and rural neighbourhood characteristics using audit and GIS data: derivation and reliability constructs', *International Journal of Behavioural Nutrition and Physical Activity*, 6, (44), pp. 1-33.
- Food Standards Agency. (2008a) *Food Law: Code of Practice (England)*. FSA
- Food Standards Agency. (2008b) *Starting up –Your first steps to running a catering business*. London: Food Standards Agency
- Foresight. (2007) *Tackling Obesities: Future Choices – Project report*. London: Government Office for Science
- Glanz, K., Sallis, J., Saelens, B. and Frank, L. (2005) 'Healthy Nutrition Environments: Concepts and Measures.' *American Journal of Health Promotion*, 19, (5), pp. 330-333.
- Google (2008) *Google Maps UK*. Available at: <http://maps.google.co.uk/maps?hl=en&tab=w1> (Accessed: 05/08/08).

- Holsten, J. E. (2009) 'Obesity and the community food environment: a systematic review', *Public Health Nutrition*, 12, (03), pp. 397-405.
- Lake, A. and Townshend, T. (2006) 'Obesogenic environments: exploring the built and food environments', *The Journal of the Royal Society for the Promotion of Health*, 126, (6), pp. 262-267.
- Lytle, L. A. (2009) 'Measuring the Food Environment: State of the Science', *American Journal of Preventive Medicine*, 36, (4, Supplement 1), pp. S134-S144.
- Macdonald, L., Ellaway, A. and Macintyre, S. (2009) 'The food retail environment and area deprivation in Glasgow City, UK', *International Journal of Behavioral Nutrition and Physical Activity*, 6, (1), pp. 52.
- Maddock, J. (2004) 'The relationship between obesity and the prevalence of fast food restaurants: state-level analysis', *American Journal of Health Promotion*, 19, (2), pp. 137-43.
- McKinnon, R. A., Reedy, J., Morrisette, M. A., Lytle, L. A. and Yaroch, A. L. (2009) 'Measures of the Food Environment: A Compilation of the Literature, 1990-2007', *American Journal of Preventive Medicine*, 36, (4, Supplement 1), pp. S124-S133.
- Morland, K. B. and Evenson, K. R. (2009) 'Obesity prevalence and the local food environment', *Health & Place*, 15, (2), pp. 491-495.
- Office for National Statistics (2005) *Land Use Statistics (Generalised Land Use Database)*, [Online]. Available at: <http://www.neighbourhood.statistics.gov.uk> (Accessed: 01/08/2008).
- Office of Public Sector Information. (2000) 'Freedom of Information Act Chapter 36 '.
- Paquet, C., Daniel, M., Kestens, Y., Leger, K. and Gauvin, L. (2008) 'Field validation of listings of food stores and commercial physical activity establishments from secondary data', *International Journal of Behavioral Nutrition and Physical Activity*, 5, (1), pp. 58.
- Public Health & Environmental Protection Regeneration Directorate at Newcastle City Council. (2008) Outlet listings for Newcastle, 08/2008.
- Sharkey, J. R. (2009) 'Measuring Potential Access to Food Stores and Food-Service Places in Rural Areas in the U.S', *American Journal of Preventive Medicine*, 36, (4, Supplement 1), pp. S151-S155.
- StataCorp. (2008) 'Stata'.

- Story, M., Kaphingst, K. M., Robinson-O'Brien, R. and Glanz, K. (2008) 'Creating healthy food and eating environments: policy and environmental approaches', *Annual Review of Public Health*, 29, pp. 253-72.
- The Chartered Institute of Environmental Health Policy Officer. (2009) Public Health and Environmental Protection Policies,
- The European Parliament and of The Council of the European Union. (2004) *Regulation (EC) No 852/2004 of The European Parliament and of The Council of 29 April 2004 on the hygiene of foodstuffs.*
- Townshend, T. and Lake, A. A. (2009) 'Obesogenic urban form: Theory, policy and practice', *Health & Place*, 15, (4), pp. 909-916.
- Wang, C., Cubbin, C., Ahn, D. and Winkleby, M. A. (2008) 'Changes in neighbourhood food store environment food behaviour and body mass index 1981–1990', *Public Health Nutrition*, 11, (9), pp. 963-970.
- Wang, M., Gonzales, A., Ritchie, L. and Winkleby, M. (2006) 'The neighborhood food environment: Sources of historical data on retail food stores', *The International Journal Of Behavioral Nutrition and Physical Activity*, 3, pp. 15 doi:10.1186/1479-5868-3-15.
- White, M. (2007) 'Food access and obesity', *Obesity Reviews*, 8, (s1), pp. 99-107.
- White, M., Bunting, J., Raybould, S., Adamson, A. J., Williams, L. and Mathers, J. C. (2004) (2004) 'N09010: Do 'food deserts' exist? A multi-level, geographical analysis of the relationship between retail food access, socio-economic position and dietary intake. Newcastle: Newcastle University
- Yell.com Customer Services. (2009) Telephone conversation with Yell.com Customer Services,
- Yellow Pages Online (2008) *Yell.com*. Available at: <http://www.yell.com/> (Accessed: 20/08/08).

Table 1 Classification terms for food outlets used in the investigation

Classification
1 – Restaurant
2 – Pub/Bar
3 – Convenience
4 – Supermarket
5 – Takeaway Food
6 - Work Place/Education
7 - Hotels/Function Rooms/Associations
8 – Medical
9 – Entertainment
10 - Department Stores
11 – Discount Stores
12 - Fast Food
13 – Pizzeria
14 - Non-Food Stores/Novelty Items
15 - Food Production Services
16 - Sandwich Shop
17 - Café/Coffee Shop
18 – Specialist
19 - Specialist Traditional
20 - Baker-Retail
21 - Health and Leisure

Table 2 Calculation of positive predictive values (ppv) and sensitivity

		Fieldwork	
		<i>outlet present</i>	<i>outlet absent</i>
Data Source	<i>outlet present</i>	true positive (TP)	false positive (FP)
	<i>outlet absent</i>	false negative (FN)	true negative (TN)

$$PPV = TP / (TP + FP)$$

$$Sensitivity = TP / (TP + FN)$$

Table 3 Sensitivity categories used in the analysis of results, adapted from Paquet et al (2008) and number of classification groups within these sensitivity categories according to data source fitting within the sensitivity range

Sensitivity categories	Sensitivity range (%)	Council data	Yellow Pages	Yell.com
Very poor	<20	0	2	3
Poor	21-30	1	3	1
Fair	31-50	1	5	5
Moderate	51-70	0	3	6
Good	71-90	12	3	1
Excellent	>90	1	0	0

Table 4 Frequency and percentage results of outlet classifications present in the fieldwork (n=617)

Classification	Frequency	Percentage (%)
1 Restaurant	112	18.2
2 Pub/Bar	136	22.0
3 Convenience	35	5.7
4 Supermarket	4	0.6
5 Takeaway Food	37	6.0
6 Work Place/Education	29	4.7
7 Hotels/Function Rooms/ Associations	31	5.0
8 Medical	7	1.1
9 Entertainment	12	1.9
10 Department Stores	15	2.4
11 Discount Stores	2	0.3
12 Fast Food	7	1.1
13 Pizzeria	1	0.2
14 Non-Food Stores/Novelty Items	21	3.4
15 Food Production Services	0	0.0
16 Sandwich Shop	35	5.7
17 Café/Coffee Shop	69	11.2
18 Specialist	11	1.9
19 Specialist Traditional	31	5.0
20 Baker-Retail	20	3.2
21 Health and Leisure	1	0.2
Not classified*	1	0.2
Total	617	100.0

*Not classified - insufficient information known about the outlet in order to classify

Table 5 Frequency and percentage of outlets not present in the fieldwork by data source (n=141)

Data Source	Frequency	Percentage (%)
Council Data	48	34.0
Yell.com	83	58.9
Yellow Pages	69	48.9

Table 6 Frequency and percentage of classifications when the outlets were not present in the fieldwork (n=141)

Classification	Frequency	Percentage (%)
Restaurant	27	19.1
Pub/Bar	29	20.6
Convenience	9	6.5
Supermarket	1	0.7
Takeaway Food	5	3.5
Work Place/Education	4	2.8
Hotels/Function Rooms/ Associations	6	4.3
Medical	1	0.7
Entertainment	1	0.7
Department Stores	0	0
Discount Stores	0	0
Fast Food	1	0.7
Pizzeria	0	0
Non-Food Stores/Novelty Items	4	2.8
Food Production Services	3	2.1
Sandwich Shop	3	2.1
Café/Coffee Shop	13	9.2
Specialist	3	2.1
Specialist Traditional	5	3.5
Baker-Retail	9	6.5
Health and Leisure	0	0
Not classified*	17	12.1
Total	141	100

*Not classified - insufficient information known about the outlet in order to classify

Table 7 Positive predictive values (PPV) for the data set (n = 617)

Data Sources	PPV (%)
Fieldwork v Council data	91.5
Fieldwork v Yell.com	79.1
Fieldwork v Yellow Pages	82.4

Table 8 Sensitivity and significance values using Council Data as a 'gold standard'

Data Sources	Sensitivity (%)	Sensitivity Category	Difference in sensitivity (%)	Significance (P-Value)
Yell.com v Council Data	51.2	Moderate	32.74	<0.001
Yellow Pages v Council Data	50.9	Fair	31.12	<0.001
Yell.com and Yellow Pages v Council data	56.0	Moderate	0	0.569