The Quality of Local Authority Spatial Data

Amy Mizen*[1], Sarah Rodgers[1] and Richard Fry[1]

[1]Farr Institute, College of Medicine, Swansea University

November 7, 2014

1. Introduction

The obesity epidemic is understood to be caused and influenced by multiple factors (Jones, 2014; Finegood, 2010). Amongst obesity research, the food environment is increasingly acknowledged as having a significant influence on obesity levels (Papas, 2007). Poor quality food environments, describe areas where there is an abundance of unhealthy food sources and poor availability of fresh and healthy foods. The availability of unhealthy food is believed to be promoting an unhealthy lifestyle and an over consumption of energy dense foods. This is hypothesised to contribute to rising rates of obesity.

This investigation will evaluate the quality of data, including addresses, provided by local authorities (LA). We will endeavour to collect a time series of data to be used in a natural experiment. We will model the food environment of child home-to-school routes at household level. This environmental data will then be linked to routine health data, which will allow the analysis of the relationships between child health and environmental exposures.

2. Background

The global obesity epidemic is a public health priority in countries around the world (Lobstein, 2004; Han, 2010). There is a serious concern for the issue of *childhood* obesity because of the health impacts later on in life. Many studies support the idea that obese children are at much higher risk of becoming obese adults than children and adolescents who have a healthy weight (Biro, 2010; Serdula, 1993). Of increased interest is the impact of the food environment on obesity. Poor food environments describe areas where there is poor quality of food available to residents (Dean, 2011). Poor food environments are usually densely populated with fast food outlets and lack in supermarkets or shops selling fresh fruits and vegetables. Such areas have also been referred to as obesogenic environments which do not just describe areas lacking healthy food outlets but also deficient in access to green spaces and walking opportunities (Egger, 1997; Williams, 2014).

The prevalence of obesogenic environments surrounding schools is of concern because of the threat to children's health (Austin, 2005). An increased density of fast food outlets near to schools encourages the increased consumption of fast food and also contributes to lifelong unhealthy attitudes towards food (Egger, 1997). Many studies that have previously investigated the impact of food environment and exposure have used "fixed" spatial units. However, to better understand causes of childhood obesity, there is a need to investigate the environment that children experience throughout their day, rather than simply where they live (Kestens,2010). Harrison et al. (2014) highlighted the importance of further investigating the exposure environment that children experience on their journey to school.

In order to accurately model a spatial environment however, it is imperative that accurate data are imported into a GIS. Equally important is an awareness of the data limitations and implications this may have on proceeding analyses (Maynooth University, 2014; Devillers, 2007; Devillers, 2010). Initially, a data quality assessment will be carried out on data provided by a typical local council and

* Amy@chi.swansea.ac.uk

the Food Standards Agency. These datasets will be cleaned and then compared in order to see whether local council datasets will be able to add valuable information to our model. If the local council data are found to be useful, this will add to a sophisticated model of the exposure environment that children experience on their way to school. There are currently no longitudinal investigations of the food exposure environment within the UK (CEDAR UK, 2014).

Spatial data are becoming increasingly important in policy research (Pirog, 2014). However, obesogenic environments are little researched and so local councils in the UK currently do not have the evidence they need to prevent additional fast food outlets from opening. In order to help planners and councillors, it is important to create high resolution spatial patterns of the exposure environment that children are subjected to on their way to school. Currently, council planners believe there is harm caused from an overabundance of food outlets in an area and deny planning applications. However, appeals against blocked planning applications are often successful due to a lack of evidence of the harmful effects.

We anticipate that this study will lead to an increase in data standards within local councils and standardised national datasets; through working with the Welsh Collaboration of Health and Environment, a group comprised of environmental health protectorate directors, local council health improvement officers, academics, and public health practitioners. To our knowledge this is the first study combining council-sourced GIS data and routinely collected health data to investigate obesogenic environments. To date, comparable studies have had to recruit participants and collect data prospectively.

3. Data & Methods

The main purpose of collecting the food outlet data is to develop a density model of food outlets for home-to-school routes. The densities will then be combined with health data held within the Secure Anonymised Information Linkage (SAIL) at Swansea University to evaluate the impact of the food environment on children's Body Mass Index (BMI).

Datasets of food outlets were obtained from Swansea County Council and Monmouthshire County Council. Food outlet data from the Food Standards Agency (FSA) were also obtained (FSA, 2014). The FSA dataset was taken to be the reference data, or "gold standard," as it is a national, routinely collected dataset. The council datasets were compared with the FSA data to see whether they are suitable datasets to use in building the density model.

The outlet data from Swansea and Monmouthshire Councils were received in different formats. In order to optimise comparability, the outlet data were formatted to csv files and where possible the same column names were allocated. Character columns were transformed in to lowercase, to minimise differences between datasets. The council data sets were matched with FSA data by the only common field which was outlet name.

Unmatched records were identified and sought to solve why they did not match any FSA outlets.

4. Results

Unitary Authority	**Auto Matched (%)**	**Manual Matched (%)**	**Rural Matched (%)**	**Urban Matched (%)**
Swansea	96	100	99	97
Monmouthshire	4	89	90	88

Table 1. Summary of matching progress

The rate of match is shown in Table 1. Auto matched describes the percentage of outlets that were the same as outlets contained in the FSA data. Manual matched refers to the percentage of outlets that match the FSA data after manual corrections have been made to the unmatched records. Swansea Council had a much greater rate of matching the FSA data compared to Monmouthshire. Of the 1361 food outlets in Swansea, 1307 were correctly automatically matched. Of the 47 unmatched outlets from Swansea, 42 records were found have spelling errors. The remaining five outlets had been closed down and the council dataset had not been updated. Monmouthshire had a far lower auto match rate. The manual match rate saw a significant increase in matching.

After cleaning the data, rural areas were more likely to be matched than urban areas.

For the wider investigation, food outlet data for all 22 Welsh LAs has been requested. Ten LAs are currently collating the datasets. Permission from the remaining ten LAs is still being sought. At the conference we aim to present the match rates of the data obtained to date and present a preliminary density model to give an idea of the potential of this study.

5. Discussion and Conclusions

The LA data was susceptible to errors. Human error was the biggest contributor to error in the LA datasets i.e. spelling mistakes. Furthermore, the data providers from the LAs had to bring together datasets from various departments in order to collate these datasets. There is no uniform method of data collection between LAs, or even between departments within an LA. This complicates the comparison of data between LAs as it means that auto matching cannot be depended on to produce comparable results.

In the FSA data, the spatial locators are postcode centroids. Post code centroids have a spatial accuracy of approximately 100m. However, UPRNs which are being provided with the LA data gives the address level coordinates which will allow for higher resolution analyses when constructing the density model. UPRNs allow for the unique identification of a property which removes the possibility of error. As seen in our results, there can be many ways to format an address, but only one UPRN.

An advantage of using LA data is that this promotes engagement between LAs and academia; which not only promotes the importance of data that LAs record but also encourages better data collection and storage. Investigations such as this one encourages LAs to communicate with one another and share knowledge and data collection methodologies. Furthermore, LA data has proven to provide better typologies of food outlets stores compared to the FSA data. Along with opening and closing times of the outlets. The LA data was longitudinal whereas FSA is not which is very valuable for planning policy and public health research.

From the analysis of two local authority food outlet datasets, it can be concluded that despite potential flaws, LA data can provide researchers with valuable information that may not be captured in national datasets. Care should be taken when using such data sets but used correctly, LA data can add value to larger, uniformly collected national datasets. The value of local council data should be promoted so to encourage more stringent data capture and recording methods.

6. Acknowledgements

Many thanks to Swansea County Council and Monmouthshire County Council for sharing their data.

7. References

Biro FM, Wien M (2010) **Childhood obesity and adult morbidities** 91:1499–1505.

Austin SB, Melly SJ, Sanchez BN, Patel A, Buka S, Gortmaker SL: **Clustering of fast-food restaurants around schools: a novel application of spatial statistics to the study of food environments.***Am J Public Health* 2005, 95:1575–81.

Cetateanu A, Jones A (2014) **Understanding The Relationship Between Food Environments, Deprivation And Childhood Overweight And Obesity: Evidence From A Cross Sectional England-Wide Study**. *Health & Place* 27 : 67-76

Dean WR, Sharkey JR (2011) **Rural and urban differences in the associations between characteristics of the community food environment and fruit and vegetable intake.***J Nutr Educ Behav* 43:426–33.

Devillers R, Stein A, Bédard Y, Chrisman N, Fisher P, Shi W (2010) **Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities**. *Trans GIS.* 14:387–400.

Devillers R, Bédard Y, Jeansoulin R, Moulin B (2007)**Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data**. *Int J Geogr Inf Sci* 21:261–282.

Egger G, Swinburn B (1997) **An "ecological" approach to the obesity pandemic.***BMJ*. 315:477–80.

Finegood DT, Merth TDN, Rutter H (2010) **Implications of the foresight obesity system map for solutions to childhood obesity.***Obesity (Silver Spring)* 18 **(1)** :13–6.

Food Standards Agency (2014) **Search for food hygiene ratings** [http://ratings.food.gov.uk/open-data/en-GB] [Accessed October 12th 2014]

Han JC, Lawlor D a, Kimm SYS (2010) **Childhood obesity.***Lancet.* 375:1737–48.

Harrison F, Burgoine T, Corder K, van Sluijs EMF, Jones A (2014) **How well do modelled routes to school record the environments children are exposed to?: A cross-sectional comparison of GIS-modelled and GPS-measured routes to school.***Int J Health Geogr* 13:5.

Kestens Y, Lebel A, Daniel M, Thériault M, Pampalon R (2010) **Using experienced activity spaces to measure foodscape exposure.***Health Place* 16:1094–103.

Lobstein T, Baur L, Uauy R, Obesity I (2004) **Obesity in children and young people : a crisis in**. 5:4–85.

Maynooth University (2014) Data Quality in GIS Available from http://www.nuim.ie/staff/dpringle/gis/gis11.pdf [Accessed October 22nd, 2014]

Papas M a, Alberg AJ, Ewing R, Helzlsouer KJ, Gary TL, Klassen AC (2007) **The built environment and obesity.***Epidemiol Rev* 29:129–43.

Pirog MA (2014) **Data will derive innovation in public policy and management research in the next decade**. *J Policy Anal Manag* 33:537–543.

Serdula M, Ivery D, Coates R, Freedman D, Williamson D, Byers T (1993) **Do obese children become obese adults**. *Preventive* 22:167–177.

Williams J, Scarborough P, Matthews a, Cowburn G, Foster C, Roberts N, Rayner M (2014) **A systematic review of the influence of the retail food environment around schools on obesity-related outcomes.***Obes Rev* 15:359–74.

Biographies

Amy Mizen is a DECIPHER PhD student at the Farr Institute @ CIPHER, Swansea University. Beginning in October 2014, her PhD project is investigating the impact of modelled school travel routes on child health using GIS and routine linked data.

Richard Fry is a Senior Research fellow in GIS at the Farr Institute @ CIPHER, Swansea University. His research interests include accessibility modelling, health geographies, data integration and linkage, OpenSource and WebGIS.

Sarah Rodgers is an Associate Professor in Spatial Epidemiology and an investigator in the new MRC e-health centre of excellence, CIPHER, at Swansea University. Her research is aided by anonymised individually-linked health, and demographic data, and aims to influence policy to improve environments and positively impact physical and mental health.