# Improving the accuracy of GIS generated environmental exposures for children's routes to school

A Mizen*[1], S.E.Rodgers [1] and R.J.Fry [1]

[1] Swansea University Medical School

January 12, 2018

**Summary**

Modelling the daily exposure environment assists epidemiologists and public health researchers in providing evidence for policy on a range of public health issues. The dose-response relationship between exposure to *food and dietary intake*, however, has not been widely investigated. Previous research found that GIS generated routes are a good proxy for the distance that a child travels but not for the environment they are exposed to. This investigation evaluated whether a weighted network analysis improved the accuracy of GIS modelled school walking routes and associated exposures.

**KEYWORDS:** GIS, GPS, Retail Food Environment, Weighted Network, Obesity

## 1. Introduction

This study has developed a methodology to generate population level exposure to the 'retail food environment' (RFE) along children's walking commute to school and home.
This study:

- Investigated which environmental characteristics differ along GPS routes to and from school, and shortest network routes (SNR) to and from school;
- Incorporated the environmental characteristics that differ to GIS methodology by using a weighted network to emulate route choice more realistically;
- Used the weighted network to model routes to and from school in a GIS for individuals that provided GPS data;
- Evaluated the accuracy of the routes to and from school modelled using the weighted network.

The RFE is receiving greater attention from researchers and policy makers as a way to explore obesity-related behaviours. Many studies that have investigated the link between the RFE and obesogenic behaviours have tended to focus on a single activity space that individuals will interact with in their day to day lives. For example, studies have focussed on the RFE either around the home or school (Davis & Carpenter 2009; Fraser & Edwards 2010; Smith et al. 2013; An & Sturm 2012; Buck et al. 2013). The majority of studies that have focussed on the link between childhood obesity and the RFE have focussed on the RFE surrounding schools, and policies have been adopted by Local Authorities (LAs) in the UK limiting fast food outlets within 400m of school premises (Dr Foster Intelligence and Land Use Consultants 2011). However, a systematic review published in 2014 (Williams et al. 2014) found little evidence to suggest that the RFE surrounding schools influences food purchases and consumption. The review did find some evidence that the RFE may have an effect on body weight but the authors suggested that this may be a result of residual confounding. There is a conflicting evidence base amongst RFE studies and this has been attributed to the methodological inconsistencies between studies. Studies have used different data sources, defined the RFE in different ways, and used different GIS methods to define the RFE (Wilkins et al. 2017).

Children's journeys through the RFE on their way to and from school have been investigated by two studies in the UK (Harrison et al. 2014; Griffiths et al. 2014). These studies have used GIS to predict

* amy@chi.swan.ac.uk

exposure to the RFE along a child's commute to and from school. Harrison et al. (Harrison et al. 2014) and Griffiths et al. (Griffiths et al. 2014) modelled children's commutes to and from school using shortest network analysis in a GIS. Griffiths et al. concluded that there "is little support for the notion that exposure to food outlets in the home, school and commuting neighbourhoods increase the risk of obesity in children". However, GIS-generated commutes have been found to be a suitable proxy for the distance a child travels but not for calculating the RFE they are exposed to (Harrison et al. 2014).

The Foresight report (Butland et al. 2007) highlighted the complexity of the obesity epidemic by stating that change is required at individual and population levels, across multiple sectors. Therefore, studies investigating obesity and obesogenic behaviours should acknowledge the complexity of the issue and the study design should reflect this. This study investigated whether developing a more sophisticated GIS method could be used to improve the accuracy of GIS modelled RFE exposures along children's routes to and from school. This could allow population modelling of individual level environment and health data.

## 2. Methodology

All GPS processing and GIS analysis was undertaken in PostGIS (using pgadmin3 version 9.5). The methodology is summarised in Figure 1.

### 2.1 Data Sources

#### 2.1.1 GPS routes

I obtained a large sample of GPS data for 995 children aged 13-14 from researchers who had worked on a large-scale study called the PEAR study (Anon n.d.). The PEAR study was a cross-sectional study of 982 students from Bristol, South Gloucestershire, North Somerset and Bath and North East Somerset.

The PEAR data provided 949 walking routes to school and 976 walking routes home from school, for 884 individuals. The GPS data from the PEAR project were cleaned and prepared for analysis.

#### 2.1.2 Road network

The road network used to calculate the GIS modelled routes was generated from the OSM dataset. The road dataset was noded using pgRouting.

#### 2.1.3 GIS generated shortest network routes

Shortest Network Routes (SNRs) were calculated for the 884 PEAR participants that walked to school. The SNRs were calculated from home to school locations. The home point locations were provided by the PEAR research team. The name of the school attended by each participant was also provided by the PEAR research team. School point locations were downloaded from OS Mastermap (Ordnance Survey 2017b).

### 2.2 Environmental Characteristics of routes

Characteristics of the environment that have been documented in the literature as influencing route choice (Harrison et al. 2011; Dessing et al. 2016; Mölter & Lindley 2015) were calculated for both the GPS walking routes, and the corresponding GIS generated SNRs. Table 1 defines the environmental characteristics that were calculated.

### 2.3 Conditional Logistic Regression

A conditional logistic regression analysis was undertaken to investigate the environmental factors associated with route choice along child walking routes. A conditional logistic regression was undertaken to identify the most discriminatory environmental characteristics between GPS and SNR; that is, those characteristics that caused the greatest differences between the GPS routes and the modelled GIS SNRs.

### 2.4 Weighted Network

The results of the logistic regression models were used to assign costs (also known in the literature as impedances) ((Papinski & Scott 2011; Feng et al. 2010)) to the road network. Line geometries (vertices) that made up the road network were labelled based on the environmental characteristics they contained. The vertices were then allocated impedance values. The environmental characteristics that were statistically significantly greater on the GPS routes compared with SNR (e.g. traffic lights) resulted in a lower cost on the network. Environmental characteristics that were statistically significantly smaller along GPS routes compared to SNR were given a higher cost on the network (e.g. food outlets).

### 2.5 Modelling routes to and from school using a weighted network

Routes to school and routes home from school were generated in PostGIS using the Dijkstra algorithm (Bondy & Murty 1976) in pgRouting (Takubo et al. 2017; Feng et al. 2010) using the cost weighted network. The weighted network routes (WNR) to and from school were then used to calculate the associated exposures to the RFE along the commute to and from school. The exposure to the RFE was calculated in PostGIS and defined as the number of outlets along the route within 100m of the route (Panter et al. 2010; Harrison et al. 2014; Harrison et al. 2011; Burgoine et al. 2015).

### 2.6 Multilevel Regression Model

A multilevel regression model was fitted to assess the association between the GPS exposures and weighted network exposures (WNE). A multilevel model was used because the route data was hierarchical. Two random effects regression models were run. One regression model was fitted for the GPS exposures and WNE for routes to school. A second regression model was fitted for the GPS exposures and WNE for routes home from school. were calculated in R (version 3.3.3).

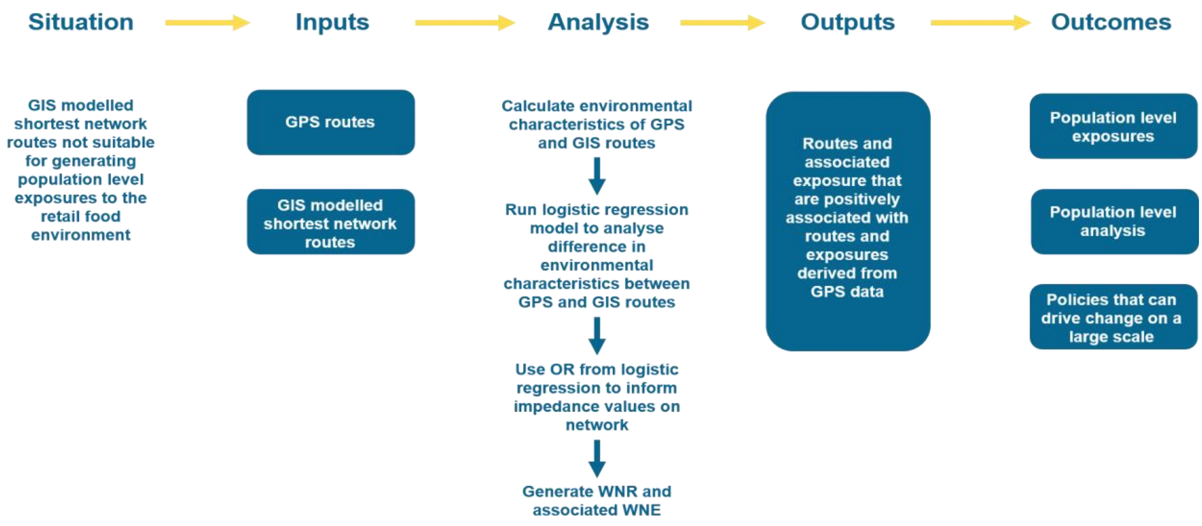


*Figure 1. Summary of methodology*

*Table 1. Environmental characteristics calculated along GPS and GIS modelled routes in PostGIS; method and data source*

| Environmental Characteristic | How to measure | Data Source |
|---|---|---|
| **Length of route (m)** | Calculate length of route | |
| **Green space along route (%)** | Percentage of route within 25m of a green space | Greenspace polygon data were downloaded from OpenStreetMap (OSM) (OpenStreetMap n.d.) |
| **Blue space along route (%)** | Percentage route within 25m of a blue space | Bluespace polygon data were downloaded from Ordnance Survey (OS) Open Rivers (Ordnance Survey 2017c), Meridian 2 (Ordnance Survey 2017a) and OSM (OpenStreetMap n.d.). The three datasets were merged to include areas where there were gaps in each dataset's coverage |
| **Traffic lights (n)** | Count of points along route | Traffic light point data were obtained from OSM (OpenStreetMap n.d.) |
| **Accidents (n)** | Count of points along route | Road traffic accident data was downloaded from Stats19 (Department of Transport 2016). Accidents that occurred between the school commuting hours (7:30-9:30 and 14:30-16:30) were extracted and represented as point data |
| **Type of street (%)**<br>**a. Main road**<br>**b. Residential**<br>**c. Minor Road**<br>**d. Footpath** | Percentage of route with this road type. Using OSM road classification | The OSM road types were aggregated into four road classifications that have been used in the literature (Dessing et al. 2016): main road, minor road, residential road and footpath |
| **Woodland** | Percentage of woodland along route (% within 25m) | Woodland polygon data were downloaded from OS Meridian 2 (Ordnance Survey 2017a) |
| **Exposure to RFE** | Count of unhealthy food outlets within 100m of route. Chapter 4 documents how the unhealthy food outlet data were collated and prepared | Postcode level food outlet point data were downloaded from the Food Standards Agency (Anon n.d.). A typology of opening times is documented in Appendix 9. This was so to account for outlets not being open all day. |

## 3. Results

Overall, 884 individuals provided GPS data on their walking route to school *or* walking routes home.

### 3.1 Routes to school

The results showed that for routes to school, there were significantly more blue spaces and traffic lights along the GPS routes. The Odds Ratio (OR) for main roads, residential, footpaths and minor roads showed that GPS routes contained a significantly smaller percentage of the route along these road types. The logistic regression results also showed that GPS routes had a significantly less exposure than the SNR for the walk **to school** (OR 0.918, 95% CI 0.834,0.976) and the walk home **from school** (OR 0.901, 95% CI 0.842,0.964).

### 3.2. Routes home from school

Similarly, there were significantly more blue spaces and traffic lights along the GPS routes than the SNR home from school. The percentage of the routes along main roads, residential, footpaths and minor roads was significantly less for GPS routes compared with the SNR. GPS walking routes home had significantly smaller exposures than the SNR (OR 0.918).

Impedance values were informed by the OR of the conditional logistic regression. The larger OR were assigned smaller impedance values. Each vertex was assigned a value based on the road type. Network vertices that had traffic lights or outlets along them were reassigned the impedance value set for traffic lights and outlets, as road type impedance values were assigned first.

A multilevel random effects model was fitted between exposure scores along the GPS route and exposure scores calculated along the weighted network. Routes were nested within individuals who were nested within schools. The WNE, both along routes to school and along routes home, were significantly associated with the exposures calculated from the GPS routes ($p < 0.001$). The regression coefficients and standard error (SE) are shown in Table 2.

*Table 2. Multilevel regression model*

| Fixed Effect | Value | SE | t-value | p-value |
|---|---|---|---|---|
| Route to school | | | | |
| Intercept | 0.81 | 1.17 | 0.48 | 0.63 |
| WNR Exposure | 1.42 | 0.07 | 19.43 | < 0.001 |
| Route home from school | | | | |
| Intercept | 0.66 | 0.34 | 1.95 | 0.05 |
| WNR Exposure | 1.15 | 0.04 | 25.65 | < 0.001 |

## 4. Conclusions

The most influential environmental factors associated with child walking routes to and from school were length, traffic light count, exposure and the proportion of the journey that was made up of a particular road type (main road, residential road, footpath and minor road). These characteristics have been used to inform impedances along a network in order to emulate route choice for children walking to and from school. This chapter has developed a methodology that can produce predictions of children’s exposure to the RFE along their walk to and from school with known accuracy. This is a novel methodology that provides great potential for developing the model to account for other modes of commuting to and from school or workplaces. This provides the potential to target public health interventions to the people who are most likely to achieve active travel to school and work (Audrey et al. 2015).

## 5. References


Akio Takubo, Anton Patrushev, Ashraf Hossain, Christian Gonzalez, Daniel Kastl, Dave Potts, David Techer, Ema Miyawaki, Florian Thurkow, Frederic Junod, Gerald Fenoy, Jay Mahadeokar, Jinfu Leng, Kai Behncke, Kishore Kumar, Ko Nagase, Manikanta Kondeti, Ma, V.V., 2017. pgRouting Project — Open Source Routing Library. Available at: http://pgrouting.org/ [Accessed September 11, 2017].

An, R. & Sturm, R., 2012. School and residential neighborhood food environment and diet among California youth. *American journal of preventive medicine*, 42(2), pp.129–35. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3298889&tool=pmcentrez&rendertype=abstract [Accessed January 20, 2014].

Anon, Food Standards Agency - Search for food hygiene ratings. Available at: http://ratings.food.gov.uk/open-data/en-GB [Accessed November 7, 2014a].

Anon, PEAR project. Available at: http://www.bristol.ac.uk/sps/research/researchprojectpages/pearproject/ [Accessed February 15, 2017b].

Audrey, S. et al., 2015. Study protocol: the effectiveness and cost effectiveness of an employer-led intervention to increase walking during the daily commute: the Travel to Work randomised controlled trial. *BMC Public Health*, 15(1), p.154. Available at: http://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-015-1464-4.

Bondy, J.A. & Murty, U.S.R., 1976. Graph theory with applications. *Operational Research Quarterly 19701977*, 290, p.270.

Buck, C. et al., 2013. Clustering of unhealthy food around German schools and its influence on dietary behavior in school children: a pilot study. *The international journal of behavioral nutrition and physical activity*, 10, p.65. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3686694&tool=pmcentrez&rendertype=abstract.

Burgoine, T. et al., 2015. Associations between BMI and home , school and route environmental exposures estimated using GPS and GIS : do we see evidence of selective daily mobility bias in children ? , 14, pp.1–12.

Butland, B. et al., 2007. Tackling Obesities : Future Choices – Project report. *Government Office for Science*, pp.1–161.

Davis, B. & Carpenter, C., 2009. Proximity of fast-food restaurants to schools and adolescent obesity. *American journal of public health*, 99(3), pp.505–10. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2661452&tool=pmcentrez&rendertype=abstract [Accessed January 27, 2014].

Department of Transport, 2016. Road Safety Data - Datasets. *Data.Gov.Uk*. Available at: https://data.gov.uk/dataset/road-accidents-safety-data [Accessed September 13, 2017].

Dessing, D. et al., 2016. Children's route choice during active transportation to school: difference between shortest and actual route. *International Journal of Behavioral Nutrition and Physical Activity*, 13(1), p.48. Available at: http://ijbnpa.biomedcentral.com/articles/10.1186/s12966-016-0373-y.

Dr Foster Intelligence and Land Use Consultants, 2011. *Tackling the Takeaways: A new policy to address fast-food outlets in Tower Hamlets*, London.

Feng, J. et al., 2010. The built environment and obesity: A systematic review of the epidemiologic evidence. *Health and Place*, 16(2), pp.175–190. Available at: http://dx.doi.org/10.1016/j.healthplace.2009.09.008.

Fraser, L.K. & Edwards, K.L., 2010. The association between the geography of fast food outlets and childhood obesity rates in Leeds, UK. *Health & place*, 16(6), pp.1124–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20691630 [Accessed August 18, 2014].

Griffiths, C. et al., 2014. A cross sectional study investigating the association between exposure to food outlets and childhood obesity in Leeds, UK. *The international journal of behavioral nutrition and physical activity*, 11, p.138. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4271469&tool=pmcentrez&rendertype=abstract.

Harrison, F. et al., 2011. Environmental correlates of adiposity in 9-10 year old children: Considering home and school neighbourhoods and routes to school. *Social Science and Medicine*, 72(9), pp.1411–1419.

Harrison, F. et al., 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. *International journal of health geographics*, 13(1), p.5. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3942764&tool=pmcentrez&rendertype=abstract [Accessed October 15, 2014].

Mölter, A. & Lindley, S., 2015. Influence of walking route choice on primary school children's exposure to air pollution - A proof of concept study using simulation. *Science of the Total Environment*, 530–531(2), pp.257–262. Available at: http://dx.doi.org/10.1016/j.scitotenv.2015.05.118.

OpenStreetMap, OSM Wales. Available at: http://download.geofabrik.de/europe/great-britain/wales.html [Accessed February 2, 2017].

Ordnance Survey, 2017a. meridian 2. Available at: http://digimap.edina.ac.uk/webhelp/os/data_information/os_products/meridian_2.htm [Accessed September 13, 2017].

Ordnance Survey, 2017b. OS MasterMap Topography Layer. Available at: https://www.ordnancesurvey.co.uk/business-and-government/products/topography-layer.html [Accessed November 15, 2017].

Ordnance Survey, 2017c. OS Open Rivers. Available at: https://www.ordnancesurvey.co.uk/business-and-government/products/os-open-rivers.html [Accessed September 12, 2017].

Panter, J.R. et al., 2010. Neighborhood, Route, and School Environments and Children's Active Commuting. *American Journal of Preventive Medicine*, 38(3), pp.268–278. Available at: http://dx.doi.org/10.1016/j.amepre.2009.10.040.

Papinski, D. & Scott, D.M., 2011. A GIS-based toolkit for route choice analysis. *Journal of Transport Geography*, 19(3), pp.434–442. Available at: http://dx.doi.org/10.1016/j.jtrangeo.2010.09.009.

Smith, D. et al., 2013. Does the local food environment around schools affect diet? Longitudinal associations in adolescents attending secondary schools in East London. *BMC public health*, 13(1), p.70. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23347757.

Wilkins, E.L. et al., 2017. Using Geographic Information Systems to measure retail food environments: Discussion of methodological considerations and a proposed reporting checklist (Geo-FERN). *Health and Place*, 44(October 2016), pp.110–117.

Williams, J. et al., 2014. A systematic review of the influence of the retail food environment around schools on obesity-related outcomes. *Obesity reviews : an official journal of the International Association for the Study of Obesity*, 15(5), pp.1–16. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24417984 [Accessed January 21, 2014].

## 6. Acknowledgements

The work was undertaken with the support of The Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement (DECIPHer), a UKCRC Public Health Research Centre of Excellence. Joint funding (MR/KO232331/1) from the British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, the Welsh Government and the Wellcome Trust, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged.

Thanks to the researchers at the MRC funder project, PEAR, for providing the GPS data that is used in this project.

## 7. Biography

Amy Mizen is a Research officer at the Farr Institute @ CIPHER, Swansea University. She passed her viva in January 2018. Her PhD project was 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 National Centre for Population Health and Wellbeing Research, Swansea University. His research interests include accessibility modelling, health geographies, data integration and linkage, OpenSource and WebGIS.

Sarah Rodgers is a Professor in Spatial Epidemiology and an investigator at theMRC 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.