Using address data to compare accuracy of accessibility to health services

Amy Burden[*1], Richard Fry[1], Sarah Rodgers[1], Daniel Grinnell[2]

[1]Farr Institute @CIPHER, College of Medicine, Swansea University, Swansea, SA2 8PP, UK
tel. +44 (0)1792 606279
email. Amy@chi.swan.ac.uk
[2]Universities' Police Science Institute
Cardiff University School of Social Sciences, 1-3 Museum Place, Cardiff, CF10 3BD
UK

KEYWORDS: Accessibility, Aggregation Error, Address Data, Secure Anonymised Information Linkage

1.Introduction

It is an international priority that healthcare be equally accessible to all [1–7]. This is because equitable access to healthcare is strongly linked with reducing ill health and suffering [11]. In order to provide the population with the best service, it is important for researchers, policy makers and planners to be aware of current healthcare accessibility needs. Modelling geographical accessibility in a GIS is one way to assess accessibility needs but there are data and methodological limitations. At present, spatially aggregated data sets are commonly used. Datasets are often only available as aggregate data but it is also used as a way of protecting the privacy of individuals by collating them into a non-identifiable spatial units and reducing computational and storage requirements [8]. However, spatial aggregation introduces ecological fallacy by smoothing local variation leading to erroneous results [9]. Anonymised data linkage systems make it possible to analyse geographical and health data based on individual residences [10]. Anonymised databanks are, however, relatively new and not a widely available alternative. It is therefore important to be aware of the error associated with spatially aggregate data. The method and results presented here contribute towards quantifying the error associated with aggregate data.

2. Background

Geographical accessibility describes how easily the population can travel to health services and it is an important and relevant issue in public health [13, 16-24]. In practice, equal geographical access to healthcare facilities is unrealistic to attain [19]. Health services are more concentrated in areas that are more densely populated so to serve an optimum catchment of the population. Usually urban populations have shorter distances to travel to health services compared to rural populations.

For successful policy and infrastructure planning it is important for researchers to be aware of current accessibility needs and communicate this effectively to planners. GIS software can be used to generate meaningful and accurate representations of geographical accessibility to health services [12–17]. Representations of modelled road networks can be used to calculate network distances between points; including turns, road and speed limits.

Anonymised data linkage systems make it possible to analyse geographical and health data based on individual residences [10]. This data type produces realistic representations that are less susceptible to bias. A successful example is the Secure Anonymised Information Linkage databank (SAIL), which uses a split file method to anonymously link environment and health data [10]. The SAIL databank contains health data for over 3 million Welsh residents and has been recognised as a powerful tool used in public health research [8, 18]. Spatially referenced environmental data are anonymised by a trusted third party before linking to individual level health data using Residential Anonymous Linking Fields [10, 19–21]. This maintains geoprivacy while allowing trusted researchers to analyse data at individual level within a secure environment. Perhaps the biggest challenge in encouraging individual level analysis is the availability of data. Although databanks such as SAIL adhere to data protection legislation and confidentiality guidelines, many data sets are not widely available at individual level due to data protection constraints, the lack of available data at this resolution, and computational problems associated with large datasets [22–24].

Most accessibility studies use data that has been aggregated for small areas. Spatially aggregated data sets are a popular choice because 1) aggregate data has been used as a way of protecting the privacy of individuals by collating them into a non-identifiable spatial units; 2) often the datasets in use may only be available at lower spatial resolutions; 3) aggregation reduces computational and storage requirements [8]. However, spatial aggregation introduces ecological fallacy, smoothing local variation, potentially leading to erroneous results [9].

This report takes a novel perspective by calculating network distances from address level data to nearest GP surgery. These distances are used to represent “real life” distances travelled by the population. These “gold standard” distances are compared with distances calculated from centroids of aggregated spatial units. Four different size spatial units are used in distance calculations. Geometric and population weighted centroid types are used in the analysis. The difference between urban and rural spatial units are also compared.

3. Data & Methods

The spatial units were obtained from the 2011 UK Census of Population, Office for National Statistics (ONS) [25]. These were Postcode, Output Area (OA), Lower Super Output Area (LSOA), and Middle Super Output Area (**Table 1.**). Postcode data from Postcode Address File, supplied by Royal Mail, provided boundary polygons for each postcode. The road network for Swansea was provided by the OS MasterMap Integrated Transport Network (ITN) Layer [26].

Address points were linked to their containing aggregation units (**Figure 1.**), retaining the address point to area level relationships. Geographically and population weighted centroids were calculated for each boundary dataset. From each centroid, the distance to the nearest GP surgery was calculated for Euclidean and Network distances. Each address was classified as Rural or Urban, according to the ONS rurality index.

Table 1. Sample of comparable international spatial unit

Spatial Unit	Average Population	Comparable International Units
Postcode	50	Japan: *Prefecture*
OA	100	Australia: *Meshblock*
LSOA	1500	Japan: *Municipality*
MSOA	7500	USA: *ZIP Codes*; Australia: *SA2s*

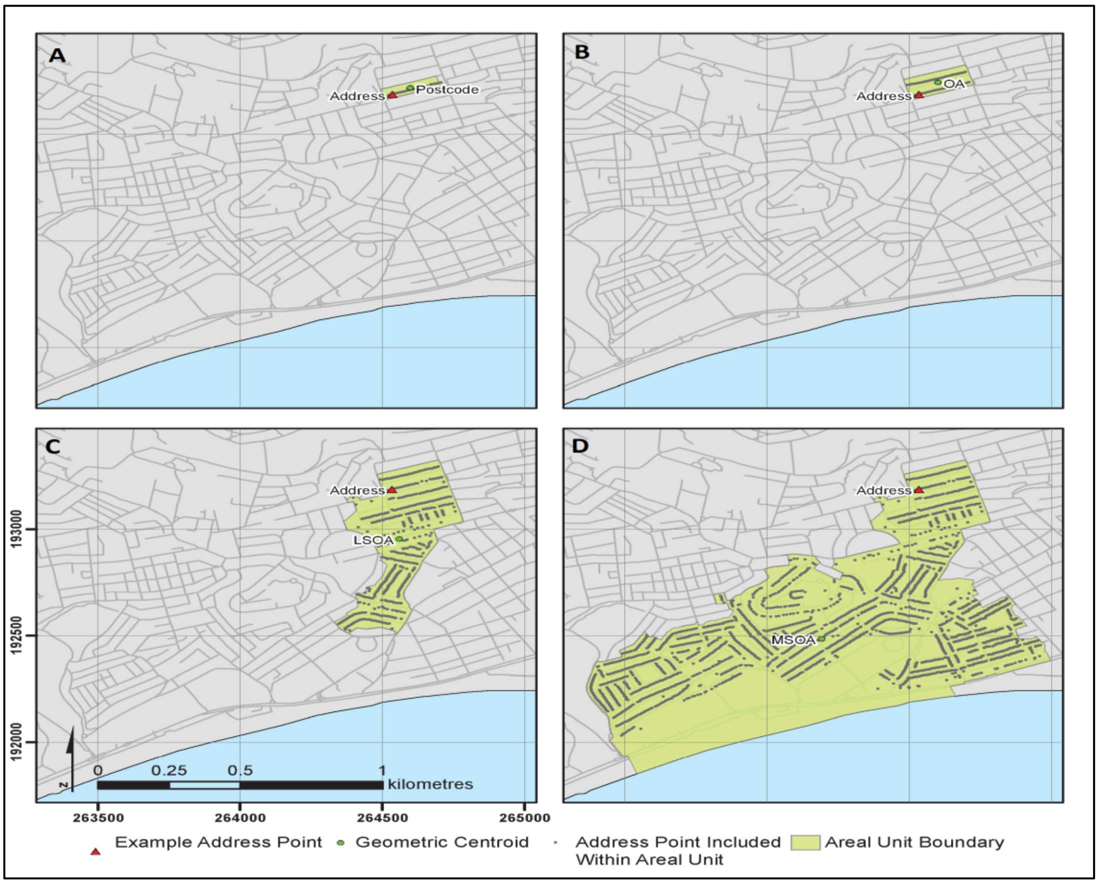


Figure 1. Aggregation unit polygons and the geometric centroids associated with an example address point. Level of generalisation: A= postcode, B= OA, C= LSOA, D= MSOA. The total number address points also encompassed by the each polygon are postcode= 64, OA= 107, LSOA= 657, MSOA= 3422.

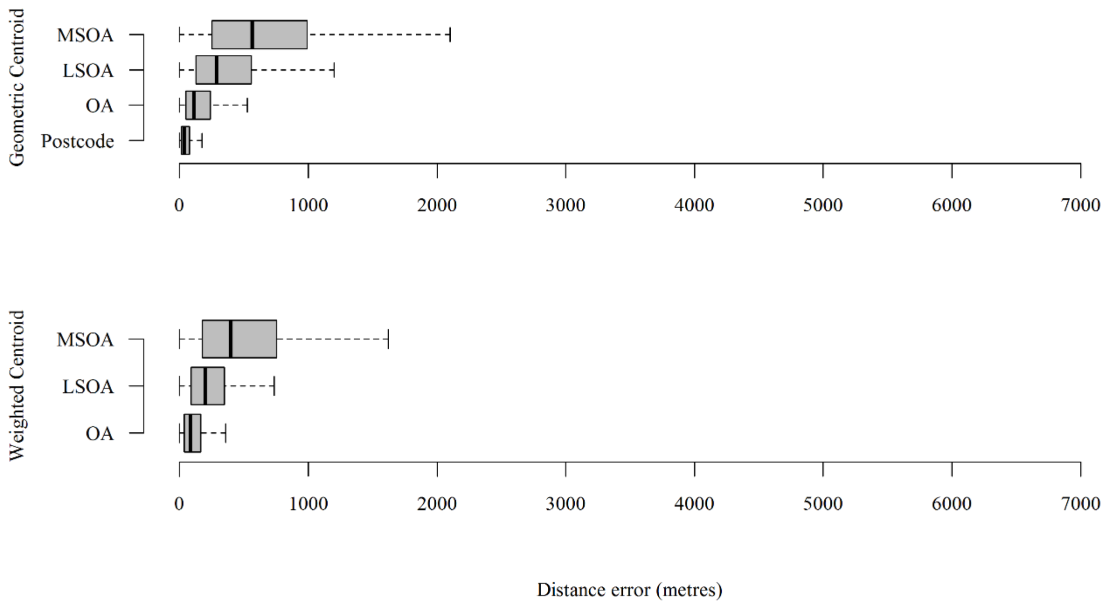
4. Results

As the spatial unit size increases, there is a greater distance difference from 'real life' distances. Weighted centroids perform better than geometric centroids. Distance calculations using population weighted centroids have lower range and median values than the distances calculated using geometric centroids. The lowest errors (RMSE) in the largest spatial units are calculated using weighted centroids. For every spatial unit size in urban populations, population weighted centroids perform the best with 4-54% smaller errors than corresponding geometric centroid distances. Population weighted centroids also have smaller range and median error values than the distances calculated using a geometric centroids. For LSOAs and MSOAs, errors are greater by 30.8% and 32.3%, respectively, for geometric centroids and network distances, compared to geometric centroids and Euclidean distances.

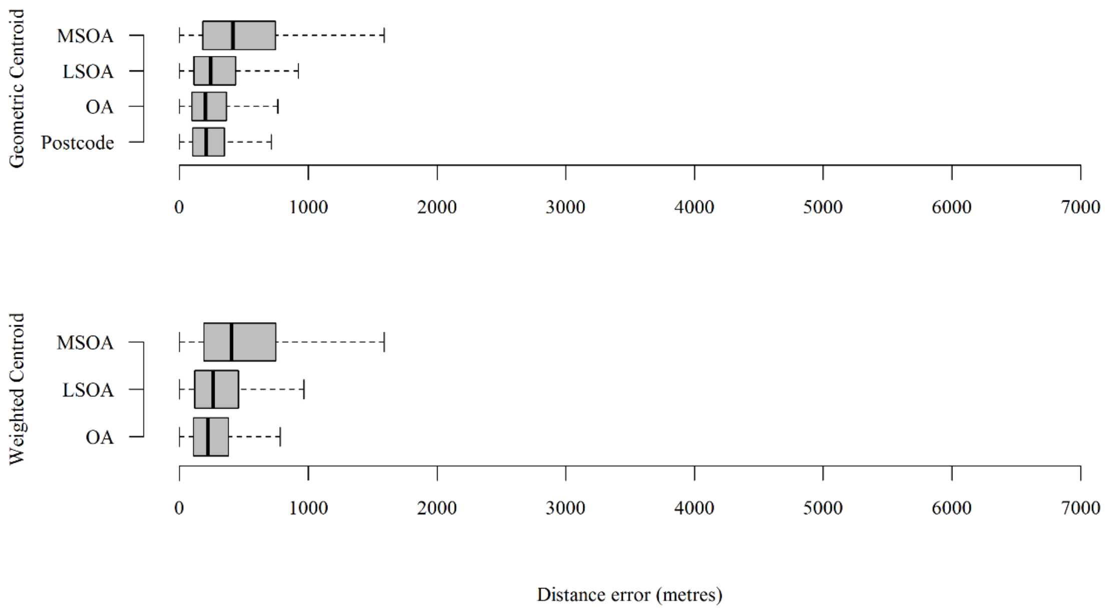
The smallest spatial units (Postcode and OA) have the most accurate representations from network distances calculations. Euclidean distances perform best at the largest spatial unit scale.

Urban areas produce less varied distance measures and smaller positional errors than rural areas. Rural regions see a larger level of positional error in every measure compared to urban regions.

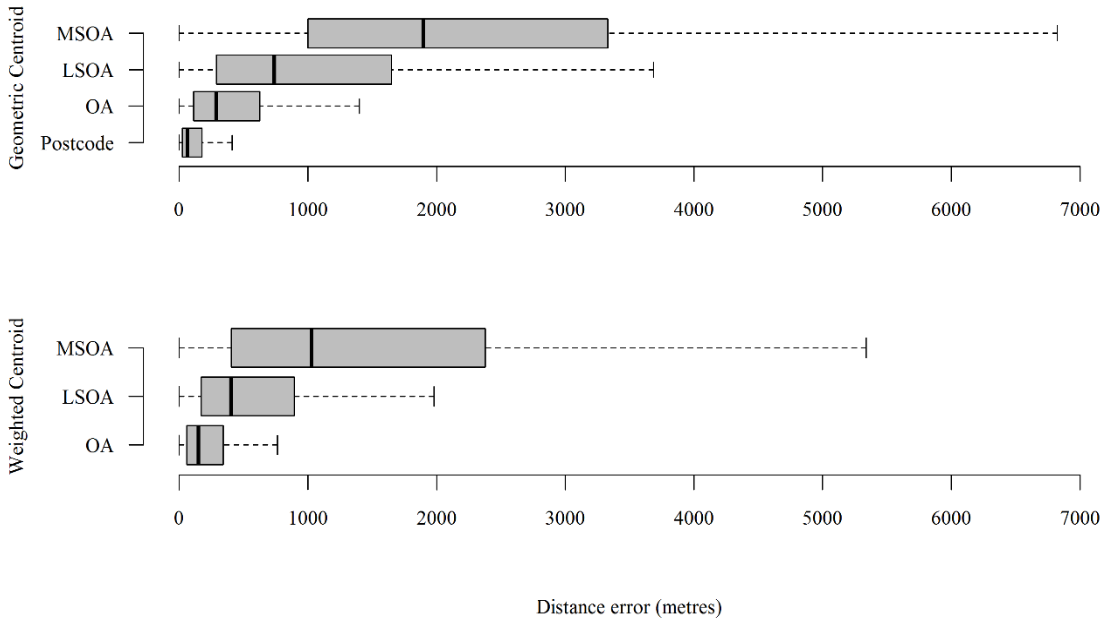
a. Urban areas using network distance calculations and two centroid types.



b. Urban areas using Euclidean distance calculations and two centroid types.



c. Rural areas using network distance calculations and two centroid types.



d. Rural areas using Euclidean distance calculations and two centroid types.

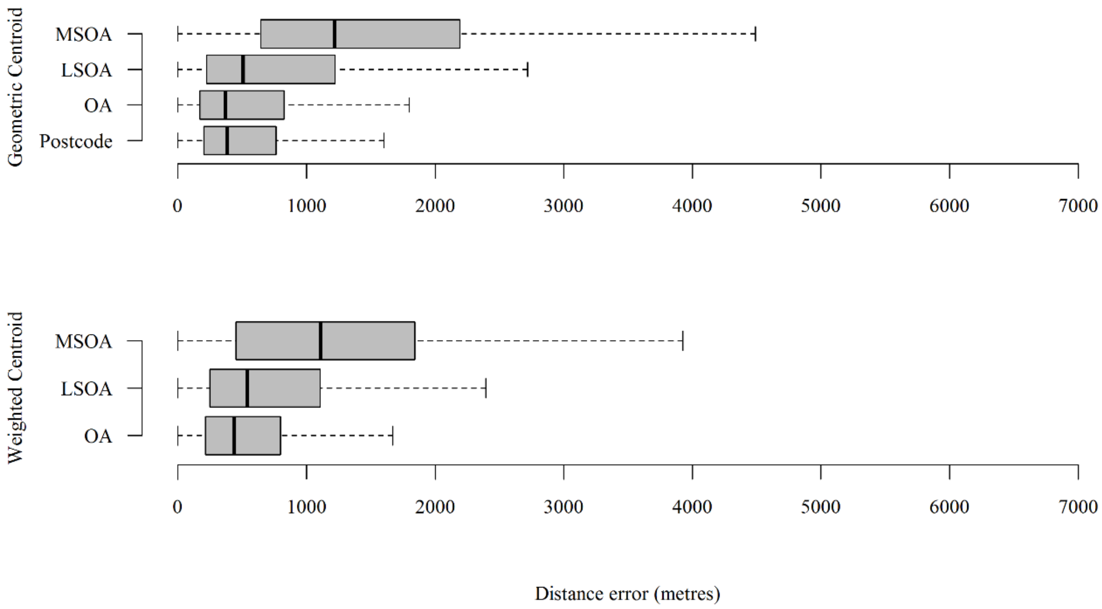


Figure 2. Distance error represents the difference, in metres, between 'real life distance travelled' and distances calculated for each spatial unit, in rural and urban areas, using Euclidean and network distance calculations and two centroid types.

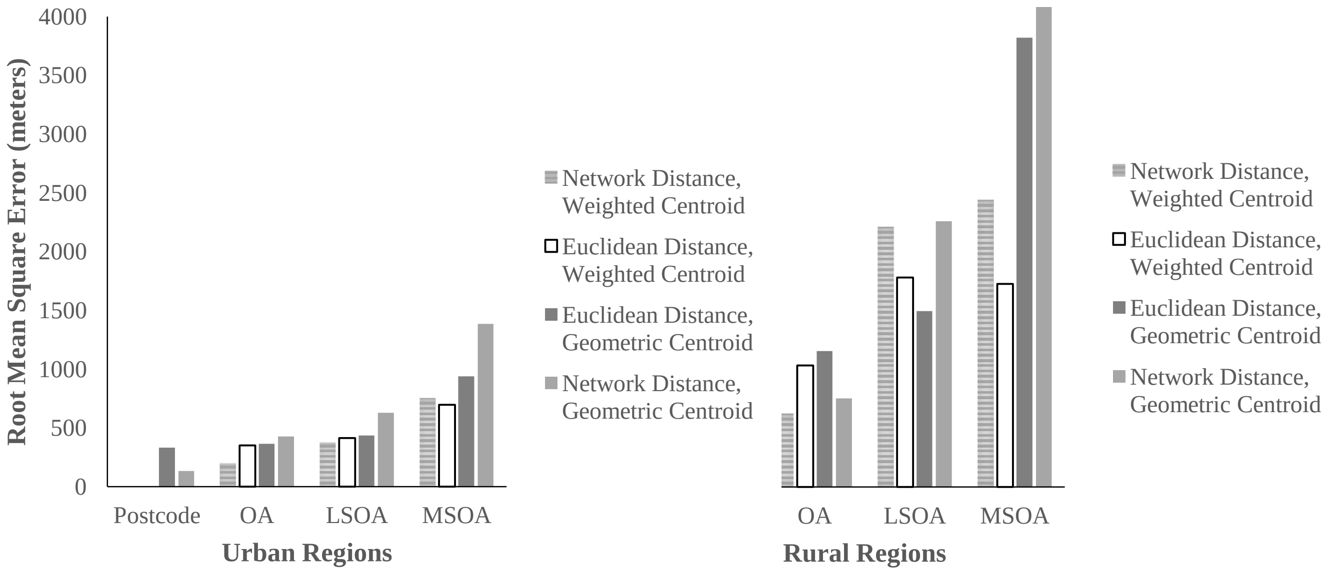


Figure 3. Root mean square error between 'real life distance travelled' and the corresponding centroid location of spatial unit, in meters.

Urban position error (m)

	Postcode	OA Geom.	LSOA Geom.	MSOA Geom.	OA Weighted	LSOA Weighted	MSOA Weighted
Network	35	110	290	565	85	195	400
Euclidean	NA	210	205	240	415	220	260

Rural position error (m)

	Postcode	OA Geom.	LSOA Geom.	MSOA Geom.	OA Weighted	LSOA Weighted	MSOA Weighted
Network	65	286	740	1895	150	400	1030
Euclidean	NA	380	370	510	1220	435	540

Table 1. Positional errors relative to address associated with each aggregate unit type to nearest 5m. Calculated from median distances. Note that no postcode weighted centroid was calculated because it is not a widely used product.

5. Discussion and Conclusions

The increased methodological complexity of calculating network distances based on small spatial units with weighted centroids is justified because of the small associated error for urban areas (35m, Urban Postcode/ 65m, Rural Postcode). However, distances calculated using geometric centroids in urban areas do not benefit from more sophisticated network distances (565m, MSOA Geom Network/240m, MSOA Geom. Euclid). Euclidean distances perform best at the largest spatial unit scale; providing adequate representations for larger spatial units (540m, MSOA Rural Euclidean/ 1030m MSOA Rural Network).

Rural areas are likely to introduce positional errors more than double the equivalent method in urban areas (290m, Urban Network Geom. LSOA/ 740m, Rural Network Geom. LSOA).

Where possible the lowest aggregation unit size should be used for accessibility calculations. These values can then be summed into larger aggregation units if necessary for representation at different geographic scales or for linking to lower resolution datasets. It is more desirable to use individual data than a single point to represent an average value for an area. The former method will highlight variation within the spatial unit. Calculating accessibility at the household level allows researchers to aggregate data to any spatial unit required. Ecological fallacy will be reduced because data may be aggregated centred on each individual/household.

If available, the use of individual level data should be encouraged. Health databanks now have the facility to link exposures to each residence and then to individual level health data [27]. If aggregate data are to be used, the unit size being analysed should be recorded and the associated flaws discussed and justified.

References

1. Meyer SB, Luong TCN, Mamerow L, Ward PR: **Inequities in access to healthcare: analysis of national survey data across six Asia-Pacific countries.** *BMC Health Serv Res* 2013, **13**:238.

2. Ruiz-Casares M, Rousseau C, Derluyn I, Watters C, Crépeau F: **Right and access to healthcare for undocumented children: addressing the gap between international conventions and disparate implementations in North America and Europe.** *Soc Sci Med* 2010, **70**:329–36.

3. Stronks K, Ravelli a C, Reijneveld S a: **Immigrants in the Netherlands: equal access for equal needs?** *J Epidemiol Community Health* 2001, **55**:701–7.

4. Norredam M, Nielsen SS, Krasnik A: **Migrants' utilization of somatic healthcare services in Europe--a systematic review.** *Eur J Public Health* 2010, **20**:555–63.

5. Meng Q, Xu L, Zhang Y, Qian J, Cai M, Xin Y, Gao J, Xu K, Boerma JT, Barber SL: **Trends in access to health services and financial protection in China between 2003 and 2011: a cross-sectional study**. *Lancet* 2012, **379**:805–814.

6. Gulliford M, Figueroa-Munoz J, Morgan M, Hughes D, Gibson B, Beech R, Hudson M: **What does "access to health care" mean?** *J Health Serv Res Policy* 2002, **7**:186–188.

7. Hu R, Dong S, Zhao Y, Hu H, Li Z: **Assessing potential spatial accessibility of health services in rural China: a case study of Donghai County.** *Int J Equity Health* 2013, **12**:35.

8. Lyons R a, Jones KH, John G, Brooks CJ, Verplancke J-P, Ford D V, Brown G, Leake K: **The SAIL databank: linking multiple health and social care datasets.** *BMC Med Inform Decis Mak* 2009, **9**:3.

9. Luo L, McLafferty S, Wang F: **Analyzing spatial aggregation error in statistical models of late-stage cancer risk: a Monte Carlo simulation approach.** *Int J Health Geogr* 2010, **9**:51.

10. Rodgers SE, Demmler JC, Dsilva R, Lyons RA: **Protecting health data privacy while using residence-based environment and demographic data.** *Heal place* 2012, **18**:209–17.

11. Prioska Ostlin, Ted Schreckerm Ritu Sadana, Josiane Bonnefoy, Lucy Gilson, Clyde Hertzman, Michael Kelly, Tord Kjellstrom, Ronald Labonte, Olle Lundberg, Carles Muntaner, Jennie Popay, Gita Sen ZV: *Priorities for Research on Equity and Health : Implications for Global and National Priority Setting and the Role of WHO to Take the Health Equity Research Agenda Forward*. Geneva; 2010(November).

12. Gatrell AC, Wood DJ: **Variation in geographic access to specialist inpatient hospices in England and Wales.** *Health Place* 2012, **18**:832–40.

13. Lovett A, Haynes R, Sünnenberg G, Gale S: **Car travel time and accessibility by bus to general practitioner services: a study using patient registers and GIS.** *Soc Sci Med* 2002, **55**:97–111.

14. Pearce J, Witten K, Bartie P: **Neighbourhoods and health: a GIS approach to measuring community resource accessibility.** *J Epidemiol Community Health* 2006, **60**:389–95.

15. Luo W: **Using a GIS-based floating catchment method to assess areas with shortage of physicians**. *Health Place* 2004, **10**:1–11.

16. Gu W, Wang X, McGregor SE: **Optimization of preventive health care facility locations.** *Int J Health Geogr* 2010, **9**:17.

17. Wang L, Roisman D: **Modeling Spatial Accessibility of Immigrants to Culturally Diverse Family Physicians**. *Prof Geogr* 2011, **63**:73–91.

18. Ford D V, Jones KH, Verplancke J-P, Lyons R a, John G, Brown G, Brooks CJ, Thompson S, Bodger O, Couch T, Leake K: **The SAIL Databank: building a national architecture for e-health research and evaluation.** *BMC Health Serv Res* 2009, **9**:157.

19. Jones KH, Ford D V, Jones C, Dsilva R, Thompson S, Brooks CJ, Heaven ML, Thayer DS, McNerney CL, Lyons R a: **A case study of the Secure Anonymous Information Linkage (SAIL) Gateway: a privacy-protecting remote access system for health-related research and evaluation.** *J Biomed Inform* 2014(January):1–9.

20. Hutchings H a, Evans A, Barnes P, Demmler J, Heaven M, Hyatt M a, James-Ellison M, Lyons R a, Maddocks A, Paranjothy S, Rodgers SE, Dunstan F: **Do children who move home and school frequently have poorer educational outcomes in their early years at school? An anonymised cohort study.** *PLoS One* 2013, **8**:e70601.

21. Exeter DJ, Rodgers S, Sabel CE: **“Whose data is it anyway?” The implications of putting small area-level health and social data online.** *Health Policy* 2014, **114**:88–96.

22. Saporito S, Chavers JM, Nixon LC, McQuiddy MR: **From here to there: Methods of allocating data between census geography and socially meaningful areas**. *Soc Sci Res* 2007, **36**:897–920.

23. Gotway CA, Young LJ: **Combining Incompatible Spatial Data**. *J Am Stat Assoc* 2002, **97**:632–648.

24. Jones SG, Ashby AJ, Momin SR, Naidoo A: **Spatial implications associated with using Euclidean distance measurements and geographic centroid imputation in health care research.** *Health Serv Res* 2010, **45**:316–27.

25. **Office for National Statistics (ONS)** [http://www.ons.gov.uk/ons/index.html]

26. **AddressBase Premium | Business and government | Ordnance Survey** [http://www.ordnancesurvey.co.uk/business-and-government/products/addressbase-premium.html]

27. Lyons R a, Ford D V, Moore L, Rodgers SE: **Use of data linkage to measure the population health effect of non-health-care interventions.** *Lancet* 2013, **6736**:10–12.

Biographies

Amy Burden is a Research Assistant in GIS at the Farr Institute @ CIPHER, Swansea University. A recent graduate from the University of Leicester, her research interests include, accessibility modelling and health geography.

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.

Daniel Grinnell is a Research Officer at the Universities Police Science Institute (UPSI), Cardiff. He previously worked as part of the team at the Farr Institute @ CIPHER, Swansea University.