

Geocoding routinely collected administrative data to measure access to alcohol outlets in Wales

Jennifer Morgan¹, Richard Fry², Scott Orford³, Sarah Rodgers² and David Fone¹

¹ Institute of Primary Care & Public Health, School of Medicine, Cardiff University, Heath Park, Cardiff, CF14 4YS, UK, ² Health Information Research Unit, School of Medicine, Swansea University, Swansea, SA2 8PP, UK, ³ Wales Institute of Social and Economic Research, Data and Methods (WISERD), Cardiff University, 46 Park Place, Cardiff, CF10 3BB, UK

KEYWORDS: Accessibility, Alcohol, Health, Secure Anonymised Information Linkage, Administrative Data

1. Introduction

A substantial level of excess alcohol consumption results in a wide range of harm and the potential impact on health at the population level of a reduction in consumption is considerable. A proposed policy for reducing alcohol consumption is reducing the availability of alcohol and the density of alcohol outlets. This paper discusses work undertaken to collate, geocode and analyse changes in the number of alcohol outlets across Wales for use in a longitudinal analysis of outlet density.

2. Background

Excess alcohol consumption has many adverse effects on health including liver cirrhosis (Leon & McCambridge 2006), cancer of the oral cavity, pharynx, larynx, oesophagus, liver (IARC 1998) and breast (Narod 2011), high blood pressure (Klatsky & Gunderson 2008), and stroke (Reynolds et al. 2003). There is also an increased risk of harm resulting from violence including homicide (Parker 1998), suicide (Ramstedt 2001), road traffic accidents (del Rio et al. 2002), domestic violence (Abramsky 2011), and attendance at Emergency Departments and Minor Injuries Units for treatment of violence-related injuries (Sivarajasingam et al. 2009).

Binge drinking is typically defined as consuming double the guideline limits in a single day during the previous week (McAlaney & McMahon 2006). It is an increasing problem, particularly rising in young women (Smith & Foxcroft 2009), and is associated with anti-social behaviour (Hayward & Sharp 2005) and around one-half of all violent crimes in the UK (Richardson & Budd 2003; Flatley et al. 2010). Binge drinking places an acute burden on the NHS as around 20–40% of all people presenting to accident and emergency departments are intoxicated, increasing to 80% after 12am (Hayward & Sharp 2005). Epidemiological evidence also suggests that binge drinking may reduce the benefits of moderate alcohol consumption on coronary heart disease risk (Pletcher et al. 2005; Bagnardi et al. 2008).

Recent data show that around 37% of men and 29% of women exceed current UK guidelines for safe levels of alcohol consumption (defined in women as > 3 units per day; men > 4 units

per day) and 20% of men and 13% of women binge drink (defined in women as >6 units per day; men >8 units per day) (Robinson & Harris 2011). Given the wide range of harm resulting from this substantial level of excess consumption, the potential impact on health at the population level from a reduction in consumption is considerable. One of the principal evidence-based policies recently recommended by the British Medical Association to reduce alcohol consumption is to reduce easy access to alcohol through controls on hours of sale and outlet density (British Medical Association 2008; Babor 2008).

Overall, there is cross-sectional evidence to suggest that higher outlet density is associated with alcohol-related harm, but evidence from longitudinal studies for associations between harm and a change in outlet density is less well established (Fone et al. 2012). This research forms part of a wider study set in Wales investigating associations between change in alcohol outlet density and alcohol-related harm to population health, in response to the National Institute for Health Research (NIHR) commissioning brief PHR no. 09/3007.

3. Alcohol outlet data collection

Alcohol outlet refers to any premise or club licensed by a Local Authority (LA) to sell alcohol. The 22 Welsh LAs are responsible under The Licensing Act 2003 for maintaining public registers of all licensed alcohol outlets. These licenses may be in the form of a premise licence, a club licence or a temporary event notice. Under The Act LAs are required to record the name and address of the outlet (Home Office 2012). These records are available for public inspection in council offices following the act coming into force in November 2005.

Requests were made for the names and location of all licensed alcohol outlets within each LA between Nov 2005 and Dec 2011. Data requested for each outlet consisted of: the date permission was granted or the licence became active (referred to as *Start Date*), the licence expiry date or an indicated date of outlet closure (*End Date*), whether this premise is licensed for ON and/or OFF premise sales (*On/Off Status*), the hours permissible to sell alcohol or general opening hours of the outlet (*Opening Hours*) and the type of premise as assigned by the LA (*Premise Type*) if available. All LAs were able to provide an actual or approximate license issue date allowing us to summarise the number of outlets annually. Several LAs were unable to provide precise outlet closure dates, so the date of the last interaction with the outlet was used to generate an approximate end date. Half of the LAs were able to provide the On/Off status of outlets, and 9 were able to provide opening hours. Each LA presented its own unique challenges for data collation, such as the need to collect records in person, and collation from multiple PDFs and text files with inconsistent formatting, or print outs into a useable format. As a result extensive processing was required to gather all files, transpose the information into useable outlet records, and collate into a single database of outlets. Tools and techniques required included web scraping, bespoke software for PDF manipulation, string manipulation and optical character recognition.

4. Geocoding outlet licence data

The main purpose of collecting the alcohol outlet data was to develop an outlet density measure over time for the population of Wales. This density measure can then be combined with longitudinal health records held within the Secure Anonymised Information Linkage (SAIL) at Swansea University to assess the impact of change in alcohol outlet density. To realise an alcohol outlet density for the population of Wales it was necessary to geocode the

outlet data so that it could be used in a GIS. Geocoding software was developed which used key address identifiers such as premise name, premise number, street name, town and postcode to match premise data to Ordnance Survey AddressBase Premium (ABP) (Ordnance Survey, 2012). Each record was assigned a unique identifier from the ABP Unique Property Reference Number so that the address details could be linked. It was assumed, perhaps naively, that this would be a relatively straightforward process given the legal status of the outlet licencing documents. Of the 19,700 individual licence records only 52% could be automatically matched using the geocoding software, with the lowest match rate 35% and the highest 86%. The main reasons for a failure to match an address record were an absent or inaccurate premise name or street number and an absent or inaccurate postcode. The lack of these details made it impossible to disambiguate similar address records automatically and therefore required a manual matching process to take place for the unidentifiable records. However, there was a marked difference between the licence records obtained from more urbanised LA's and those with a more rural setting. The average match rate for the more urbanised LA's (Cardiff, Swansea & Newport) in Wales was 72%. We cannot definitively say why there is this marked difference, however one could suppose that the night time economies in these areas are larger and as such a LA need to accurately know where a premise exists from a planning and policing perspective is probably greater than the more rural areas. This need may lead the manual adjustment of addresses by the LA and/or the use of addressing software to match records when the licence data is captured by the LA.

5. Change in alcohol outlet density

Measuring change in alcohol outlet density can be approached using a number of different methodologies ranging from a simplistic number of outlets per unit of population approach to more sophisticated distance weighted gravity models. In this study we have decided to look at a range of these models to assess the impact of change in alcohol outlet density on population health. These models are a density measure calculated by summing the number of outlets per LSOA divided by population, a kernel density measure of outlets by LSOA and two cumulative opportunity gravity models. The last two models (models 3 & 4) are in essence variations on the enhanced two-step floating catchment analysis (E2SFCA) methodology which has been used to measure accessibility by a number of authors in a variety of scenarios (Luo et al, 2009; Higgs et al. 2012; Dai and Wang 2011). Model 3 uses a simple average distance from all residences within a catchment area, in this case a 10 minute walking time, to an outlet. This figure acts as the service provision ratio (outlet: average distance from population centroid(residence)) score defined in step 1 of the E2SFCA under the assumption that distance is the biggest determinant in accessibility when measuring access to alcohol outlets. Other factors such as outlet capacity, opening hours, type of outlet are less tangible and therefore harder to quantify. An attempt to quantify these less tangible aspects of alcohol outlets is made in model 4 by weighting the outlet by type and opening hours as part of the calculation (weighted outlet: average distance from population centroid (residence)). As this study requires the change in alcohol outlet density to be calculated a consistent window needed to be selected to represent each year and allow change in availability to be evaluated. However, a range of start and end dates were provided by LAs which can influence this calculation. There are several options for calculating change: using the number of outlets operating at a fixed point in time (such as 1st January), a count of the number of outlets open throughout the year, or outlets weighted to reflect the proportion of

the year they are open. Quarterly summaries offer insight into the seasonality within the data as well as annual change and therefore was chosen as the time period for analysis with accessibility scores calculated for each quarter. These scores are then calculated and encrypted and deposited in SAIL to be linked to health records. Each individual (linked by residence) will have four accessibility scores for each quarter between 2005 and 2011, enabling us to analyse the impact in the change of alcohol outlet density over time through multilevel modelling.

6. Initial Results

Due to the nature of the data collation problems this research has only recently been able to analyse the data and outputs from the GIS models. Figure 1 illustrates a basic density measure, average distance density measure and an E2SFCA (aggregated to output area level for mapping purposes) density measure for two quarters in 2008 in Cardiff. A small increase in the number of licenced premises (20) in June of 2008 has resulted in different accessibility scores for some areas, particularly to the north-western areas of Cardiff. There are two main blocks of high accessibility (the city centre and the southerly edge around Cardiff Bay) where outlet numbers are high but the resident population is low due to commercial and industrial interests.

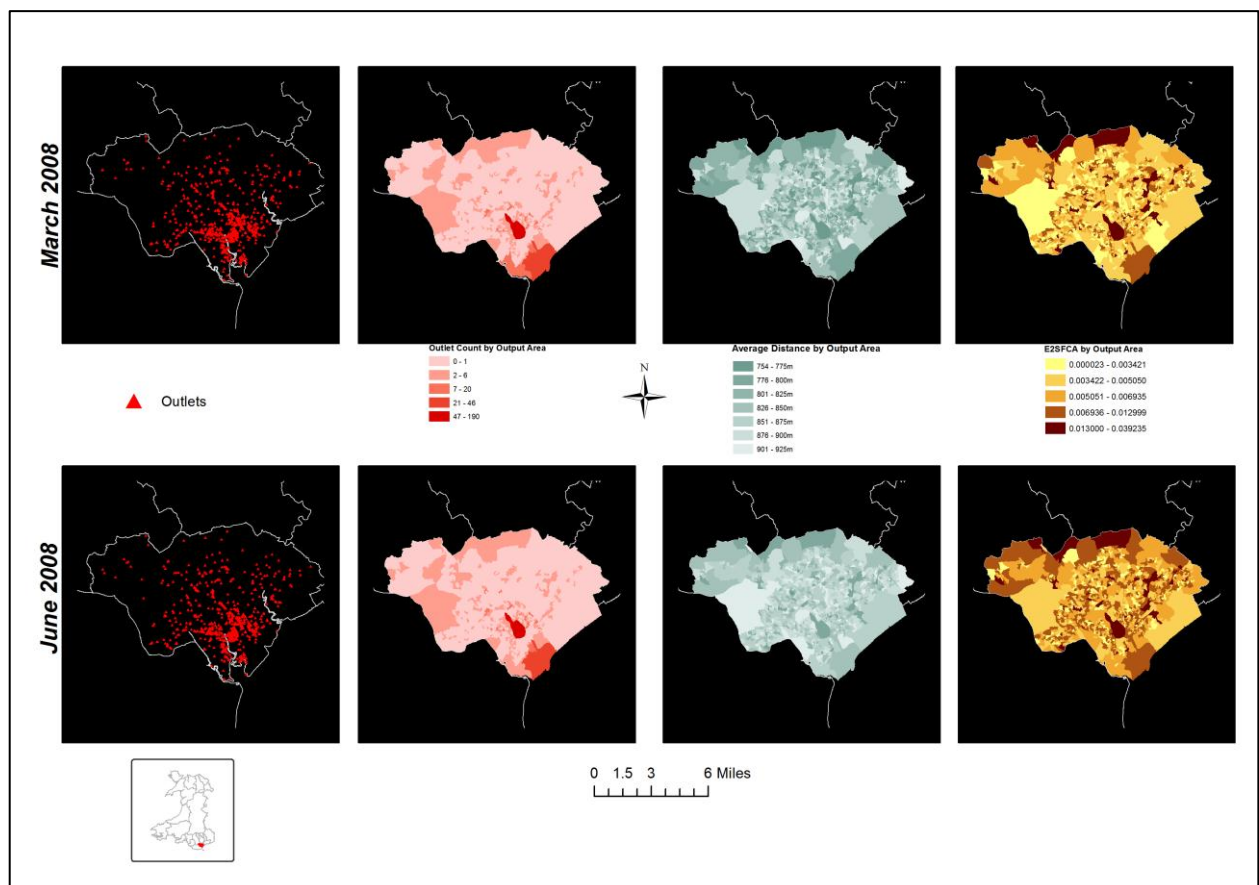


Figure 1: Initial outputs from GIS models for two different quarters in 2008

Once the analysis is complete at an all-Wales level for each quarter, the next step is to link

the accessibility scores to longitudinal health data held within the SAIL databank. This is due to be completed early in the New Year.

7. Conclusions

The research relies upon license premises data which are routinely collected as a statutory function of Local Authorities. It therefore relates to the types of administrative data that are increasingly being promoted as important sources of socio-economic data (e.g. ESRC National Data Strategy 2009-12), the 'Beyond 2011' census programme and part of the Government's 'Big Data' opportunity in the public and policy sectors. However, similar to much of the administrative data published as part of the Government's open data initiative (see BBC, 2012), the license data as currently published are largely unusable without significant cleansing, re-formatting and manual correction. This is compounded by the unsystematic nature of its collation resulting in twenty-two different data assemblage methods, one for each Local Authority in Wales. This is despite the statutory nature of the information laid down by the Home Office. However, it has been shown in our initial results that the data held by local authorities has the potential to provide a rich source of geographic information which can help researchers and policy makers tackle a wide range of issues within society.

8. Acknowledgements

This project is funded by the National Institute for Health Research Public Health Research (NIHR PHR) programme (project number 09/3007/02). Please visit the PHR programme website for more information.

The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the PHR programme, NIHR, NHS or the Department of Health.

We would like to acknowledge the local authorities for providing the data making this research possible.

9. References

- Abramsky, T., Watts, C. H., Garcia-Moreno, C., Devries, K., Kiss, L., Ellsberg, M., Jansen, H. & Heise, L. 2011. What factors are associated with recent intimate partner violence? Findings from the WHO multi-country study on women's health and domestic violence. *BMC Publ Health*, 11, 109.
- Babor, T. F. 2008. Tackling alcohol misuse in the UK (editorial). *BMJ*, 336, 455.
- Bagnardi, V., Zatonski, W., Scotti, L., La Vecchia, C. & Corrao, G. 2008. Does drinking pattern modify the effect of alcohol on the risk of coronary heart disease? Evidence from a meta-analysis. *J Epidemiol Community Health*, 62, 615 - 619.
- BBC. 2012. Government on-line data ignored by 'armchair auditors'
<http://www.bbc.co.uk/news/uk-politics-20221398>, 09-11-12
- British Medical Association 2008. *Alcohol misuse: tackling the UK epidemic*, London, BMA.
- Dai D, Wang F, 2011, "Geographic disparities in accessibility to food stores in southwest Mississippi" *Environment and Planning B: Planning and Design* 38(4) 659 – 677
- Del Rio, M. C., Gomez, J., Sancho, M. & Alvarez, F. 2002. Alcohol, illicit drugs and medicinal drugs in fatally injured drivers in Spain between 1991 and 2000. *Forensic Sci Int*, 127, 63 - 70.
- ESRC National Data Strategy: UK Strategy for Data Resources for Social and Economic Research, 2009-12
- Flatley, J., Kershaw, C., Smith, K., Chaplin, R. & Moon, D. 2010. *Crime in England and Wales 2009/10*, London, Home Office.
- Fone DL, Dunstan FD, Webster C, Lyons RA, Rodgers S, Lee S, Shiode N, Orford S, Sivarajasingam V, Weightman A, Morgan J, Fry R. 2012 Change in alcohol outlet density and alcohol-related harm to population health. *BMC Public Health*, 12:428.
- Hayward, R. & Sharp, C. 2005. Young people, crime and anti-social behaviour: findings from the 2003 Crime and Justice Survey. *Home Office Research Findings*, 245, 1473 - 8406.
- Home Office. 2012. Premises Licence Application
<http://www.homeoffice.gov.uk/publications/alcohol-drugs/alcohol/alcohol-licensing>, 14-11-12
- International Agency for Research on Cancer (Iarc) 1998. *Alcohol Drinking*, Lyon, IARC.
- Klatsky, A. L. & Gunderson, E. 2008. Alcohol and hypertension: a review. *J Am Soc Hypertens*, 2, 307 - 317.
- Langford, M., Fry, R. & Higgs, G., 2012. Measuring transit system accessibility using a modified two-step floating catchment technique. *International Journal of Geographical Information Science*, (January 2012), pp.1–22.
- Leon, D. A. & Mccambridge, J. 2006. Liver cirrhosis mortality rates in Britain from 1950 to

2002: an analysis of routine data. *Lancet*, 367, 52-56.

Luo, W. and Qi, Y., 2009. An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. *Health and Place*.15, pp.1100-1107

Mcalaney, J. & McMahon, J. 2006. Establishing rates of binge drinking in the UK : Anomalies in the data. *Alcohol Alcohol*, 41, 355 - 357.

Narod, S. A. 2011. Alcohol and risk of breast cancer (editorial). *JAMA*, 306, 1920 - 1921.

Ordnance Survey, 2012: AddressBase™ Premium - CSV. Technical specification. [<http://www.ordnancesurvey.co.uk/oswebsite/docs/technical-specifications/addressbase-premium-technical-specification-csv.pdf>]

Parker, R. N. 1998. Alcohol, homicide, and cultural context: a cross-national analysis of gender-specific homicide victimization. *Homicide Studies*, 2, 6.

Pletcher, M. J., Varosy, P., Kiefe, C. I., Lewis, C. E., Sidney, S. & Hulley, S. B. 2005. Alcohol Consumption, Binge Drinking, and Early Coronary Calcification: Findings from the Coronary Artery Risk Development in Young Adults (CARDIA) Study. *Am J Epidemiol*, 161, 423 - 433.

Ramstedt, M. 2001. Alcohol and suicide in 14 European countries. *Addiction*, 96, 59 - 75.

Reynolds, K., Lewis, L. B., Nolen, J. D. L., Kinney, G. L., Sathya, B. & He, J. 2003. Alcohol Consumption and Risk of Stroke: A Meta-analysis. *JAMA*, 289, 579 - 588.

Richardson, A. & Budd, T. 2003. *Alcohol, Crime and Disorder: A Study of Young Adults*, London, Home Office.

Robinson, S. & Harris, H. 2011. Smoking and drinking among adults, 2009. A report on the 2009 General Lifestyle Survey [Internet].

Sivarajasingam, V., Morgan, P., Matthews, K., Shepherd, J. P. & Walker, R. 2009. Trends in violence in England and Wales 2000-2004: An Accident and Emergency Perspective. *Injury*, 40, 820 - 825.

Smith, L. & Foxcroft, D. 2009. *Drinking in the UK. An exploration of trends*, York, Joseph Rowntree Foundation.

8. Biographies

Jennifer Morgan is a research associate at the Institute of Primary Care and Public Health, School of Medicine, Cardiff University. Her research interests include analysis of health data, healthcare systems modelling and simulation.

Richard Fry is a senior research fellow in GIS at the Health Information Research Unit, Swansea University. His research interests include accessibility modelling, health geographies, data integration and linkage, OpenSource and Web GIS.

Scott Orford is a senior lecturer in Spatial Analysis and GIS in the School of City and Regional Planning, Cardiff University. His research interests include the analysis and modelling of socio-economic process, innovation in data integration and data construction and the mapping and spatial analysis of qualitative data.

Sarah Rodgers is a lecturer 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.

David Fone is professor of health sciences research at the School of Medicine, Cardiff University. His research interests are primarily in the analysis of record-linked electronic cohort studies and multilevel analyses of place and health outcomes.