

Spatial Characterization of Hypertension Clusters Using a Rural Australian Clinical Database

Rachel Whitsed¹, Ana Horta¹, Herbert F Jelinek^{2,3}, Faezeh Marzbanrad⁴

¹ Charles Sturt University, Institute of Land, Water and Society, Albury, Australia

² Charles Sturt University, School of Community Health, Albury, Australia

³ Macquarie University, Faculty of Medicine & Health Sciences, Clinical Medicine, Sydney, Australia

⁴ Dept. Electrical and Computer Systems Engineering, Monash University, Melbourne, Australia

Abstract

This study aimed to characterize the spatial distribution of hypertension (HT) clusters in a rural Australian city using self-reported HT data collected at a local health-screening clinic. HT status was recorded for 515 self-selected participants in a free health-screening program in Albury, New South Wales, Australia. We compared predictions of HT clusters computed using spatial scan statistic and Generalised Additive Model (GAM). We then implemented a new approach incorporating sensitivity analysis in GAM to combine cluster predictions at multiple span sizes. A statistically significant cluster for HT was identified in Albury centered to the north of the main urban center, with relative risk up to 2.29. The sensitivity analysis confirmed the cluster location and highlighted other potential HT clusters. Our approach allows detection of irregularly-shaped disease clusters and highlights potential clusters that may be overlooked using traditional methods. This is important in cases using local, small datasets where regularly-shaped or overly smoothed disease clusters may not provide enough detail to be suitable for targeting place-based interventions.

1. Introduction

Recent statistics for hypertension (HT) in Australia state that, in 2014-15, 23% of adults (4.1 million people) had measured high blood pressure (BP) [1], a slightly higher number than the 21.5% figure reported for 2011-12 [2]. Other reports describing HT in Australia mention that high BP increases with age [3]; half the people who reported having high BP also have heart or circulatory conditions [4]; and high BP is the most frequently managed problem in Australian general practice [5]. According to the HeartWatch Survey carried out in 2011, Australians living in rural areas (34.2%) are significantly more likely to have high BP than in metropolitan areas (31.3%) [6]. This rural/metropolitan dichotomy is increasingly concerning for health outcomes in Australia. Although national and state health surveys are useful to distinguish metropolitan

from rural areas in terms of incidence and prevalence of major health issues, the data provided is not enough to identify the problem at the local scale [7]. For HT (as for any other health issues), acting locally with the communities to improve health access can prevent hospitalizations, improve wellbeing and save money [8]. However, collating local health data results in a small dataset. Traditionally, epidemiological studies rely on large datasets representative of the national population. As such, small datasets are often disregarded due to lack of representativeness concerns. Acknowledging the importance of using local health datasets, we have developed an approach to assess whether the spatial distribution of self-selected participants at a local health-screening clinic is representative of the distribution of a rural city population [9].

The aim of this study is to use the same self-reported data to characterize the spatial distribution of HT in Albury. Previous studies suggest that in this rural Australian city, HT is substantially higher than the state and national averages [10, 11].

Most studies incorporating a spatial modelling approach apply regression algorithms with known risk factors to map HT [12, 13]. Others use the spatial scan model (SaTScan) [14] to identify HT clusters [15]. Here we compare SaTScan with the regression-based Generalised Additive Model (GAM) [16] to spatially characterize the occurrence of HT clusters. Our approach compares these two methods and integrates a sensitivity analysis, which reports clusters in a more reliable and robust manner than traditional GAM.

2. Methods

2.1. Study area and database

The study area is the urban area of Albury, a rural city in New South Wales, Australia with an urban population of 45,512 and an over-45 population of 18,350 ((2011 figures, [17]) (Fig. 1)). This study focuses on the over-45 population as a) this can be interpreted as the population at

risk and b) our database primarily consists of people aged over 45.

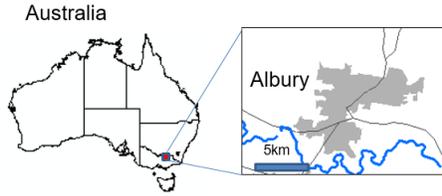


Figure 1. Location of Albury in Australia

Our dataset comprises 515 clinical data (2.8% of the over-45 population of Albury) collected from self-selected participants in a free health-screening program [10, 11], part of a research project studying chronic disease prevalence in rural areas. Amongst other data, presence of HT and age were recorded. Presence of HT and population over 45 were aggregated to meshblock level (the smallest geographical unit in the Australian census – approximately 30-40 households).

2.2. Spatial modelling algorithms

The models considered here are GAM [16] and SaTScan [14]. We also describe an implementation of GAM incorporating sensitivity analysis (GAM-SA). In all models we assume the presence of HT in each location is Poisson-distributed, and that under the null hypothesis, the expected number of cases with HT is proportional to the population over 45. All models calculate the relative risk (RR) of HT at all locations.

SaTScan uses circles or ellipses of different sizes and calculates a likelihood ratio statistic based on observed and expected cases [14]. Scan statistics are used to detect clusters by gradually scanning a circular or elliptical window across the region and evaluating observed and expected number of cases. The window with the maximum likelihood is the most likely cluster. For the Poisson model the likelihood function is [18]:

$$\left(\frac{c_{in}}{E(c_{in})} \right)^{c_{in}} \left(\frac{c_{out}}{E(c_{out})} \right)^{c_{out}} \quad (1)$$

where c_{in} is the observed number of HT cases inside the scan window, $E(c_{in})$ is the expected number of HT cases within the scan window, c_{out} is the observed number of HT cases outside the scan window and $E(c_{out})$ is the expected number of HT cases outside the scan window. Monte Carlo hypothesis testing is used to calculate the p -value for the windows of maximum likelihood, to test whether the null hypothesis can be rejected. We implemented spatial scanning using the SaTScan program [18].

GAM predicts a continuous RR surface of a disease, smoothed spatially [16], where statistically significant high RR locations can be interpreted as disease clusters. We

implemented GAM with a Poisson distribution and a log link, smoothed for location using a two-dimensional LOESS smooth [16]:

$$\log(E(c_i)) = S(x_i, y_i) + \log(n_i) \quad (2)$$

where $E(c_i)$ is the expected number of HT cases in meshblock i , S is a smooth function at the centroid of meshblock i with coordinates x_i and y_i , and n_i is the population aged over 45 in meshblock i . The optimal span size was determined by minimizing the Akaike's Information Criterion (AIC) – span is specified as proportion of data points used to define the analysis neighborhood, and the larger the span, the smoother the surface. Prediction significance was tested using 1000 permutations to estimate p -values. Clusters were defined as areas where the lower limit of the RR 95% confidence interval is greater than 1 [9, 19]. GAM was implemented using the MapGAM package [20] in R [21].

We applied sensitivity analysis to GAM (GAM-SA) by comparing cluster predictions at multiple span sizes, instead of just the optimal span size. We performed the algorithm for span sizes ranging from 0.05 to 0.95 at intervals of 0.05 and identified presence of clusters and RR within clusters at each span size. We then averaged the RR over multiple span sizes with the clusters to identify locations within the clusters that have consistently high RR at multiple span sizes.

3. Results

SaTScan returned an identical HT cluster for both circular and elliptical windows (Fig 2). Although both window shapes calculated the RR as 2.13 (Table 1), the calculated p -value differed. Smaller secondary clusters were also identified but they were not statistically significant.

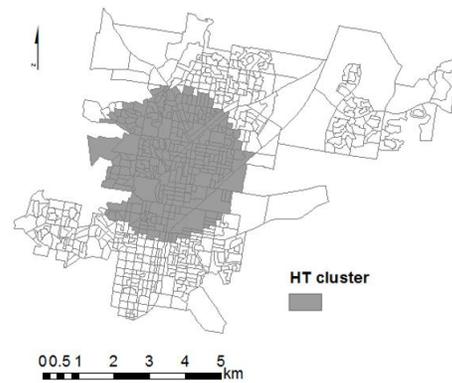


Figure 2. HT cluster identified using SaTScan (both circular and elliptical windows).

The optimal span size for GAM determined using AIC was 0.6. From this statistically significant clusters were identified as described above. A single cluster was identified (Fig. 3), with RR ranging from 1 to 2.07.

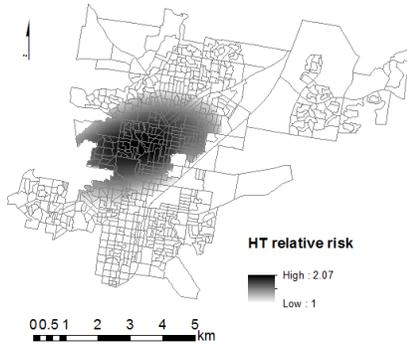


Figure 3. HT RR surface identified using GAM

Applying sensitivity analysis to GAM produced a mean RR surface across multiple span sizes (Fig. 4), ranging from 1 to 2.29. The areas outlined in black show potential cluster locations not calculated using GAM at optimal span size only. A cluster reliability surface was also calculated (Fig. 5). All areas identified as potential clusters in Fig. 4 are outlined in black. Overlapping areas of high RR and high cluster reliability can be interpreted as stable clusters, i.e. these locations are statistically significant high RR areas at the majority of span sizes.

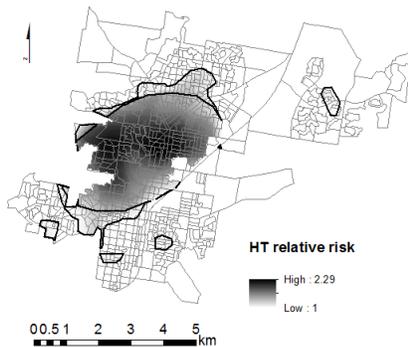


Figure 4. HT mean RR using GAM-SA. Black outlines show potential clusters not identified in Fig. 3.

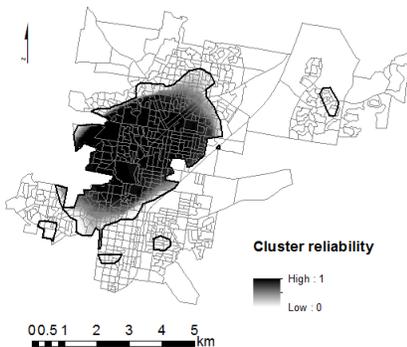


Figure 5. HT cluster reliability GAM-SA. Black outlines show all potential clusters.

For each method only statistically significant ($p < 0.05$) clusters are displayed. Table 1 also shows the RR calculated for each method. SaTScan returns a single RR value, while GAM has continuous values of RR within each cluster – the maximum RR value is shown here.

Table 1. Relative risk and associated p -values

Cluster	RR	p
SaTScan – circular window	2.13	0.00037
SaTScan – elliptical window	2.13	0.001
GAM – optimal span	2.07	< 0.05
GAM – sensitivity analysis	2.29	< 0.05

4. Discussion

The three methods (SaTScan, GAM and GAM-SA) all found a single statistically significant cluster for HT in Albany centered to the north of the main urban center, with all methods agreeing on the approximate location. The cluster detected using GAM was more elliptical in shape, however the maximum RR within the cluster was smaller than that both with SaTScan and GAM-SA. GAM-SA calculated highest RR within the main cluster. This also shows that although 0.6 was selected as the optimal span size across the entire study area using GAM, other span sizes identified higher RR in the cluster area, resulting in a higher mean RR across all span sizes using GAM-SA.

SaTScan is commonly used for disease clustering analysis, but has been criticized for its lack of flexibility in window shape [22]. Circular windows often overstate the true region of disease clustering. Elliptical windows may alleviate this issue by allowing more targeted shapes, but in this case a circular cluster was still identified even when allowing for ellipses. Although the cluster shape identified using GAM was still elliptical, the core shape (darkest areas) identified using GAM-SA was more irregular. Another concern with SaTScan is the influence of the maximum window size [22].

GAM has been found to outperform SaTScan when clusters are irregularly shaped [23]. In addition, GAM returns a continuous surface of RR values, while SaTScan returns a single number within each cluster. This implies RR is homogenous within a cluster, which is unlikely to be the case. Many studies using GAM to detect clusters simply visualize the RR surface, although contours are sometimes included to show which regions are statistically significant. Our method of selecting only statistically significant clusters for both GAM and GAM-SA allows more targeted visualization and analysis.

The single span size used in GAM also means that the RR surface may be overly smoothed, meaning small local clusters may be missed. By applying GAM-SA we can assess the presence and RR values of clusters across all span sizes, highlighting further potential clusters. For example, there are small connected regions to the north and

south of the main cluster that have a lower mean RR (~1.3) and lower reliability (~0.5) but are worth investigating as potential HT clusters. These regions were discarded as being not statistically significant in GAM with optimal span, and were also not included in the SaTScan cluster.

Both SaTScan and GAM take background population at risk into account, in this case population aged over 45. The assumption is that the survey database is spatially representative of the population at risk. In a rural city with self-selected survey participants this is difficult to achieve, however GAM can also be applied to test this assumption [9]. Previous analysis [9] shows that the existing database is largely spatially representative of the over-45 population in Albury, and therefore the HT clusters detected in the current research are reliable. However the analysis could be further extended to take regions of over- and underrepresentation in the survey database into account. Future research will compare disease clusters with spatial biases in the database and investigate methods to correct for this.

The identification of disease clusters is an initial step in spatial analysis of disease mapping. Further research will model HT and other diseases taking social and environmental risk factors into account. This analysis will allow us to examine the drivers of disease at a local spatial level in a rural city, which is a key missing component in the body of research on disease mapping in Australia.

References

- [1] 4364.0.55.001 - National Health Survey: First Results, 2014-15 [Internet]. Abs.gov.au. 2017 [cited 1 April 2017]. Available from: <http://www.abs.gov.au/ausstats/abs@.nsf/mf/4364.0.55.001>
- [2] 4364.0.55.001 - Australian Health Survey: First Results, 2011-12 [Internet]. Abs.gov.au. 2017 [cited 1 April 2017]. Available from: <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4364.0.55.001main+features12011-12>.
- [3] Australia's health 2014 (AIHW) [Internet]. Aihw.gov.au. 2017 [cited 1 April 2017]. Available from: <http://www.aihw.gov.au/publication-detail/?id=60129547205>
- [4] 4364.0.55.002 - Australian Health Survey: Health Service Usage and Health Related Actions, 2011-12 [Internet]. Abs.gov.au. 2017 [cited 1 April 2017]. Available from: <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4364.0.55.002main+features12011-12>.
- [5] Britt H, C Miller G, Henderson J, Bayram C, Harrison C, Valenti L et al. A decade of Australian general practice activity 2004-05 to 2013-14. 1st ed. Sydney: Sydney University Press; 2014
- [6] Foundation T. High blood pressure statistics [Internet]. The Heart Foundation. 2017 [cited 1 April 2017]. Available from: <https://www.heartfoundation.org.au/about-us/what-we-do/heart-disease-in-australia/high-blood-pressure-statistics>
- [7] Ghosh A, Charlton K, Girdo L, Batterham M, McDonald K. Addressing the deficiencies in the evidence-base for primary practice in regional Australia - sentinel practices data sourcing (SPDS) project: a pilot study. BMC Family Practice. 2013; 14(1).
- [8] Perils of place: identifying hotspots of health inequality | Grattan Institute [Internet]. Grattan.edu.au. 2017 [cited 3 April 2017]. Available from: <https://grattan.edu.au/report/perils-of-place-identifying-hotspots-of-health-inequality/>
- [9] Whitsed R, Horta A, Jelinek H. Assessing representativeness of a rural Australian clinical database using a spatial modelling approach. EMBEC'17 & NBC'17. IFMBE Proceedings Series; 2017 (in review).
- [10] Jelinek H, Kolbe C, Wang L, Oxbrow D. Identification of Hypertension and Efficacy of Treatment in a Rural Australian Population. Clinical and Experimental Hypertension. 2008; 30(5):359-366.
- [11] White, Fi, Lexin W, Jelinek, H. Management of Hypertension in Patients with Diabetes Mellitus. Experimental & Clinical Cardiology. 2010; 15(1): 5-8.
- [12] Soljak M, Samarasundera E, Indulkar T, Walford H, Majeed A. Variations in cardiovascular disease under-diagnosis in England: national cross-sectional spatial analysis. BMC Cardiovascular Disorders. 2011;11(1).
- [13] Xu L, Lai D, Fang Y. Spatial analysis of gender variation in the prevalence of hypertension among the middle-aged and elderly population in Zhejiang Province, China. BMC Public Health. 2016;16(1).
- [14] Kulldorff M. A spatial scan statistic. Communications in Statistics - Theory and Methods. 1997;26(6):1481-1496.
- [15] Wang Z, Du Q, Liang S, Nie K, Lin D-N, Chen Y, et al. Analysis of the Spatial Variation of Hospitalization Admissions for Hypertension Disease in Shenzhen, China. International Journal of Environmental Research and Public Health. 2014Mar;11(1):713-33.
- [16] Hastie TJ, Tibshirani RJ. Generalized additive models. London: Chapman & Hall; 1990.
- [17] Australian Bureau of Statistics, Australian Government [Internet]. Abs.gov.au. 2017 [cited 1 April 2017]. Available from: <http://www.abs.gov.au>
- [18] SaTScan - Software for the spatial, temporal, and space-time scan statistics [Internet]. SaTScan - Software for the spatial, temporal, and space-time scan statistics. [cited 1 April 2017]. Available from: <http://www.satscan.org/>
- [19] Aamodt G, Samuelsen S O, Skrondal A. A simulation study of three methods for detecting disease clusters. International Journal of Health Geographics. 2006; 5:15.
- [20] Bai L, Bartell S, Bliss R, Vieira V MapGAM: Mapping smoothed effect estimates from individual-level data. R package, 2016
- [21] The R Project for Statistical Computing [Internet]. R. [cited 1 April 2017]. Available from: <http://www.R-project.org/>
- [22] Chen J, Roth RE, Naito AT, Lengerich EJ, Maceachren AM. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. International Journal of Health Geographics. 2008;7(1):57.
- [23] Young RL, Weinberg J, Vieira V, Ozonoff A, Webster TF. A power comparison of generalized additive models and the spatial scan statistic in a case-control setting. International Journal of Health Geographics. 2010;9(1):37.

Address for correspondence.

A/Prof. Herbert Jelinek
School of Community Health, CSU, PO Box 789, Albury, NSW
2640 Australia
HJelinek@csu.edu.au