pagepress

Geospatial Health 2023; volume 18:1144

Investigating local variation in disease rates within high-rate regions identified using smoothing

Matthew Tuson,[1,2] Matthew Yap,[2] David Whyatt[2]

[1]School of Physics, Mathematics and Computing, University of Western Australia, Perth, Western Australia; [2]Medical School, University of Western Australia, Perth, Western Australia, Australia

Correspondence: Matthew Tuson, School of Physics, Mathematics and Computing, University of Western Australia, 35 Stirling Highway, Crawley 6009, Perth, Western Australia, Australia.
Tel.: +61424472410. E-mail: matthew.tuson@uwa.edu.au

Key words: ecological bias; mental health services; geographic mapping; urban health services; health services needs and demand.

Acknowledgements: the Authors acknowledge the WA Data Linkage Branch, the WA Data Collections and their custodians, and use of AZTool (v1.0.3; 25/8/11; https://www.geodata.soton.ac. uk/software/AZTool). AZTool is copyright David Martin, Samantha Cockings and the University of Southampton.

Ethics approval and consent to participate: the study was approved by the Western Australia Department of Health Human Research Ethics Committee under the Project Reference Number RGS0000002799, on 17th June 2019. Requirement for patient consent was waived by the Committee due to the infeasibility of obtaining such consent and the low risk of harm and identification. All methods were performed in accordance with the guidelines and regulations set by the Committee and the Declaration of Helsinki.

Consent for publication: not applicable.

Availability of data and materials: restrictions apply to the availability of the health data examined in this study. These data were obtained from the Department of Health, Western Australia, and are available from the Department by application. The population data used are publicly available from the Australian Bureau of Statistics (ABS).

Contributions: MY conceived the study design and performed all analyses. DW secured funding. MT provided statistical and coding expertise. All authors were involved in the planning, writing, and editing of the manuscript.

Conflict of interest: the Authors declare no competing interests. The funding body had no role in the design of the study; the analysis or interpretation of the data; the writing of the manuscript; or the decision to publish the results. The funding body was also the collector and provider of the health data; these data are routinely collected; research is a secondary purpose of that collection.

Funding: the study was funded by the Department of Health, Western Australia.

Received for publication: 12 August 2022.
Revision received: 20 January 2023.
Accepted for publication: 20 January 2023.

Geospatial Health 2023; 18:1144
doi:10.4081/gh.2023.1144

Abstract

Exploratory disease maps are designed to identify risk factors of disease and guide appropriate responses to disease and help-seeking behaviour. However, when produced using aggregate-level administrative units, as is standard practice, disease maps may mislead users due to the Modifiable Areal Unit Problem (MAUP). Smoothed maps of fine-resolution data mitigate the MAUP but may still obscure spatial patterns and features. To investigate these issues, we mapped rates of Mental Health-Related Emergency Department (MHED) presentations in Perth, Western Australia, in 2018/19 using Australian Bureau of Statistics (ABS) Statistical Areas Level 2 (SA2) boundaries and a recent spatial smoothing technique: the Overlay Aggregation Method (OAM). Then, we investigated local variation in rates within high-rate regions delineated using both approaches. The SA2- and OAM-based maps identified two and five high-rate regions, respectively, with the latter not conforming to SA2 boundaries. Meanwhile, both sets of high-rate regions were found to comprise a select number of localised areas with exceptionally high rates. These results demonstrate how, due to the MAUP, disease maps that are produced using aggregate-level administrative units are unreliable as a basis for delineating geographic regions of interest for targeted interventions. Instead, reliance on such maps to guide responses may compromise the efficient and equitable delivery of healthcare. Detailed investigation of local variation in rates within high-rate regions identified using both administrative units and smoothing is required to improve hypothesis generation and the design of healthcare responses.

Introduction

The spatial distribution of disease and its correlates is regularly investigated by government health departments and by researchers. Often, the purpose of this investigation is to inform spatial interventions that address disease burden. For example, disease maps can help facilitate efficient targeting of finite healthcare resources to high-risk populations (Lessler *et al.*, 2018; Tuson *et al.*, 2020).

Often, point-level (*e.g.*, individual-level) data are available as the starting point for analysis. However, for the analysis of rates, a denominator (usually population) needs to be defined, based on a set of areal units. In Australia, examples of such units include administrative units available under the Australian Statistical Geography Standard (ASGS) – *e.g.*, Statistical Areas Level 1 (SA1s) (ABS, 2016a), which each comprise approximately 430 residents, or 200 dwellings. Comparable areas to SA1s in other countries include census Output Areas in the UK and Census

Blocks in the US. SA1-level population sizes and case counts are often small enough that their reporting compromises privacy. Consequently, analysts and custodians cannot freely share them, e.g., with the scientific community or the public. Additionally, geographic units with small population sizes often have statistically unstable rates, which are unreliable for decision-making (Werner & Strosnider, 2020). Finally, analysing SA1s independently of one another prevents identification of spatially contiguous, high-rate regions of disease. Such regions are often of interest to decision-makers, since they represent the opportunity to intervene with 'at risk' populations at scales larger than the finest resolution at which data are collected or made available (the 'minimal' resolution).

For these reasons, higher-scale administrative units are often used to produce disease maps. In the example of the ASGS, such units might be SA2s, which each contain, on average, approximately ten thousand residents (ABS, 2016a). However, such aggregation exacerbates a common geographical problem: the Modifiable Areal Unit Problem (MAUP) (Openshaw & Taylor, 1979; Tuson *et al.*, 2019). This problem is two-fold in nature: conclusions drawn can differ when results are based on data aggregated spatially by either i) different scales, or ii) different geographical boundaries at a given scale. Examples include aggregation of SA1s by either SA2s or SA3s (the "scale", or aggregation aspect) and gerrymandering of US political votes (Hodge *et al.*, 2010; the boundary, or "zonation" aspect). Furthermore, aggregation by higher-scale administrative units has the effect of smoothing underlying features within the boundaries of the chosen units, and diluting features that straddle the units' borders. Thus, features of interest are easily obscured when aggregating. These issues are amplified when the units used are far larger than necessary, e.g., Statistical Areas Level 3 (SA3s) in Australia and other, similar units in other jurisdictions (*e.g.,* see: ACSQHC, 2017; AIHW, 2007; Center for the Evaluative Clinical Sciences, 1996; Kamel Boulos & Geraghty, 2020; Mooney & Juhász, 2020); in such cases, policy makers and health service planners may be misled by 'zonation-dependent' findings that are based on such units (Tuson *et al.*, 2020). To address these issues, minimal-resolution (e.g., SA1-level in Australia) data may be smoothed using a variety of available techniques, *e.g.*, locally weighted average and empirical Bayes approaches (Waller & Gotway, 2004; Tuson *et al.*, 2020). Smoothing protects privacy, stabilises rates, preserves general features of interest that might otherwise be obscured when using higher-scale administrative units, and avoids dependence on any single set of such units (Devine *et al.*, 1994; Talbot *et al.,* 2000). However, an inherent assumption underlying smoothing is that neighbouring areas share similar characteristics, e.g., disease rates (Tobler, 1970); i.e., that they are spatially auto-correlated. This assumption reflects W. R. Tobler's First Law of Geography: "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). Despite the risk of this Law inducing the ecological fallacy, it is rarely challenged.

To investigate these issues, in this paper, we: i) exemplify the ability of smoothing to address 'single-aggregation' issues (*i.e.,* the MAUP and the obscuration of fine-resolution features) associated with an SA2-resolution map of Mental Health-Related Emergency Department (MHED) presentations in metropolitan Perth, Western Australia (WA), in 2018/19; and ii) test the assumption that nearby areas share similar characteristics, through investigating local variation in rates within high-rate regions identified using both SA2s and smoothing.

Materials and Methods

Data sources

2016 Australian Census Statistical Area (SA) boundaries for Perth were obtained from the Australian Bureau of Statistics (ABS) (ABS, 2016a). Perth was defined to comprise the five Greater Perth Statistical Areas Level 4 (SA4s): 'Perth – Inner', 'Perth – South East', Perth – South West', 'Perth – North East', and 'Perth – North West', excluding two single-SA1 islands: Rottnest Island and Garden Island, which operate primarily as a tourist and day-trip destination and a naval base, respectively.

ABS resident population data for the 2011 and 2016 Australian Censuses, stratified by SA1, were obtained via the ABS TableBuilder tool (ABS, 2016b). These data were standardised to 2016 SA1 boundaries using geographical correspondence ratios obtained from the ABS (ABS, 2016a), and SA1-level population values for the 2018/19 Australian financial year (July 1 2018 to June 30 2019) were derived via linear extrapolation.

MHED presentations for 2018/19 were extracted from the WA Emergency Department Data Collection (EDDC; Holman *et al.*, 1999). These were defined to be presentations with either: i) an ICD-10-AM (International Classification of Diseases, Tenth Revision, Australian Modification; National Centre for Classification in Health, 2010) diagnosis code beginning with "F" (Mental and behavioural disorders) or ii) a Major Diagnostic Category (IHPA, 2019) (primarily used in regional WA) of 19 (Mental diseases and disorders) or 20 (Alcohol/drug use and alcohol/drug induced organic mental disorders). ED data for Saint John of God ED, Murdoch were not available; these typically represent < 1% of presentations to Perth EDs each year (St John of God Health Care, 2016). All other presentations meeting the above diagnostic definition within WA were included, if their patient SA1s of residence fell within the study area.

Defining instability for individual units

Following Werner and Strosnider (2020), the Residual Standard Error (RSE) for each SA1 and SA2 was calculated to represent statistical stability (or instability), with RSEs < 30% considered to be stable. Specifically, the RSE for each unit i (*i.e.,* each SA1 and SA2) was calculated as:

$$RSE_i = \frac{1}{\sqrt{e_i}} \times 100$$

where e_i is the expected number of presentations for unit i, which itself was calculated as:

$$e_i = \frac{\sum_i O_i}{\sum_i n_i} \times n_i$$

where O_i and n_i are the presentation count and population size for unit i, respectively.

Confidence intervals

Byar's approximation (Breslow & Day, 1987) was used to calculate confidence intervals for non-zero rates.

Suppression of data to protect privacy

For the purposes of display (*i.e.*, in the figures), rates for units with either a presentation count or a population size between one and five, inclusive, were suppressed to protect privacy. The rates for such units were recalculated based on a presentation count of five.

Smoothing

The Overlay Aggregation Method (OAM) (Tuson *et al.*, 2020) was used to smooth the SA1-resolution data. Briefly, this method involves: i) creating multiple, aggregate-level zonations of a given geographic study area, by repeatedly aggregating a set of fine-resolution, or 'minimal-level', 'building block' spatial units defining that geography (here, SA1s) into larger, contiguous polygons, or 'zones'; ii) independently producing disease maps based on these zonations; and iii) combining, or overlaying these maps to produce a single, minimal-resolution map (Tuson *et al.*, 2020).

For use in OAM, ten zonations of Perth were created based on a target population size of 2,000 residents. In step ii above, the aggregate-scale disease maps represented crude rates of MHED presentations for each aggregate unit. In step iii, the SA1-level values calculated were population-weighted mean crude rates.

Identifying high-rate regions

For both the SA2- and OAM-produced maps, the sets of high-rate regions were defined to be the sets of areas exhibiting maximum 'targeting efficiency' (Tuson *et al.*, 2020), based on an exemplar target presentation percentage of 5%; *i.e.*, the sets of areas with the smallest population size that also contained at least 5% of MHED presentations. These areas were delineated through ranking the relevant geographic areas (*i.e.*, SA2s and SA1s, respectively) by their rates, from highest to lowest, then identifying the sets of highest-ranked units that contained at least a cumulative 5% of presentations. This approach, which has been used to identify high-rate regions previously (Lessler *et al.*, 2018; Tuson *et al.*, 2020), was applied before excluding units with unstable rates.

Software

R version 4.0.2 was used for all analyses (R Core Team, 2020). The publicly available software AZTool (Cockings *et al.*, 2011; Martin, 2003) was used to create the zonations used within OAM.

Results

25,863 MHED presentations occurred among a population of approximately 1.95 million people residing within the study area of Perth in 2018/19. Based on these values and the defined RSE threshold of 30%, a minimum population size of 840 in any unit was identified as being required for statistical stability. Only 177 out of 4,248 total SA1s (4.2% of SA1s) and 142 out of 164 total SA2s (87% of SA2s) had populations greater than this threshold. The 22 SA2s with unstable rates included all 19 SA2s whose rates required suppression to protect privacy.

Figure 1a maps the 142 SA2s that had stable rates, and Figure 1b maps the set of two SA2s that were maximally 'targeting efficient', based on the specified target of 5% of presentations. The

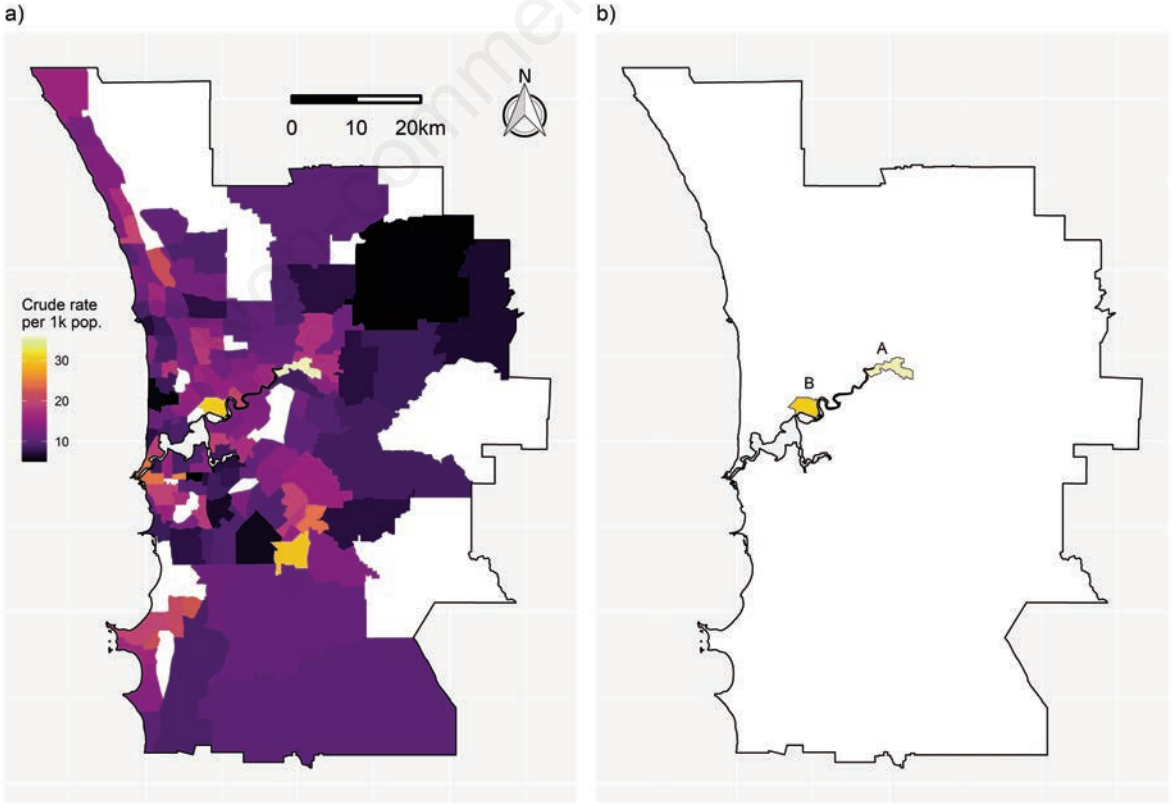

Figure 1. a) Crude, SA2-level rates of MHED presentations across Perth; b) High-rate SA2s classified based on the specified target presentation percentage of 5%.

latter SA2s, which are labelled 'A' and 'B' in Figure 1b, comprised 5.7% of all MHED presentations and 2.3% of the population; thus, their targeting efficiency can be expressed as 2.5% of MHED presentations for every 1% of the population 'targeted'.

Figure 2 shows rates and associated 95% confidence intervals for SA1s comprised within the SA2s identified in Figure 1b. Specifically, panels a) and b) in Figure 2 correspond to SA2s A and B in Figure 1b. Within each panel in Figure 2, the SA1s are ordered from west to east by geographic centroid and are coloured red unless they required suppression to protect privacy, in which case they are coloured blue. Horizontal red lines indicate the population rates for each SA2 (37.1 and 31.9 per 1,000 population, respective-

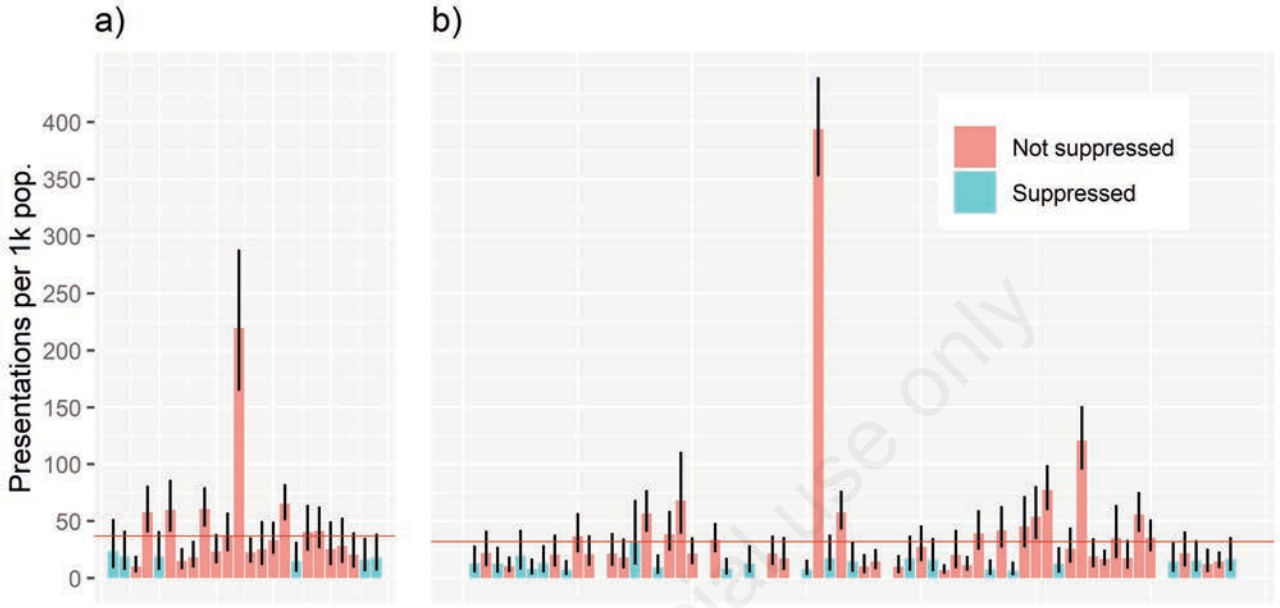

Figure 2. Crude rates of SA1s comprised within high-rate SA2s identified in Figure 1b. A single SA1 with a population less than 6 is not displayed in Figure 2b due to the y-axis being truncated.

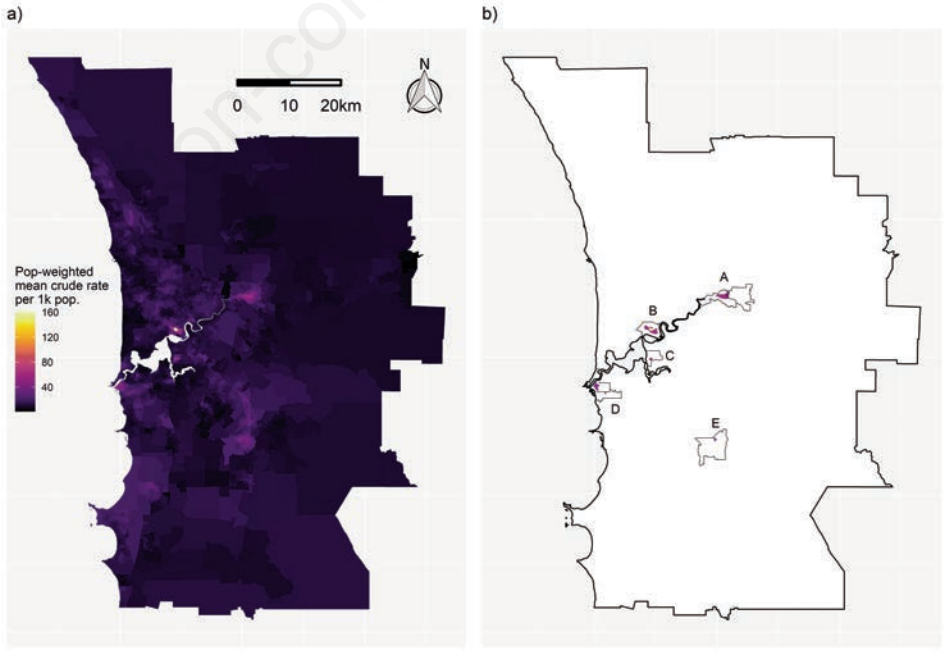

Figure 3. a) Smoothed, SA1-resolution rates computed using OAM; b) High-rate regions identified using OAM, based on the specified target presentation percentage of 5%, with the boundaries of encompassing SA2s overlaid.

ly). Within SA2 A in Figure 1b, one SA1 (out of 24) had an exceptionally high rate based on an arbitrarily chosen threshold of 100 presentations per 1,000 residents. Within SA2 B, two SA1s (out of 67) had rates greater than that threshold. Thus, while the SA2s in Figure 1b were classified as being 'high-rate', only two SA1s at most within each region had exceptionally high rates.

Figure 3a shows the smoothed map produced using OAM, and Figure 3b shows a filtered version of that map with the set of areas containing 5.1% of MHED presentations delineated. The latter areas, labelled 'A' to 'E', respectively, in Figure 3b, were highly localised. Further, some, *e.g.*, region C, did not overlap the high-rate SA2s (Figure 1b). In total, five distinct high-rate regions were identified. As depicted in Figure 4, region A straddled the border of two neighbouring SA2s; region B was located near the centre of an SA2; regions C and E were small in size and located at the borders of two different SA2s; and region D comprised two non-contiguous but nearby areas within neighbouring SA2s. Together, these regions contained 0.8% of the population; thus, their targeting efficiency can be expressed as 6.4% of MHED presentations for every 1% of the population targeted; *i.e.,* more than double the corresponding value for the set of high-rate SA2s (2.5%).

Figure 5 shows crude rates and confidence intervals for SA1s comprised within the SA2s encompassing the high-rate regions identified using OAM. Its panels are labelled 'a' to 'e', to match the labels of the latter regions. In each panel, the bars are coloured either: green, to represent SA1s that were classified as 'high-rate' using OAM; red, to represent SA1s that were not classified as such; or blue, to represent SA1s whose rates required suppression. None of the green-coloured SA1s had rates that required suppression. Horizontal red lines indicate the population rates across SA1s in each panel. In each of Figures 5a-c and 5e, only a single SA1 had an exceptionally high rate based on the threshold defined previously. In some cases, these SA1s impacted upon those around them; in Figure 5b, for example, one SA1 had a rate that was approximately ten times higher than the population rate for its panel (394 versus 32 presentations per 1,000 population), and, by virtue of their proximity to this SA1, several nearby SA1s, despite having relatively low un-smoothed rates, were classified as being 'high-rate'. Similar results were observed in Figure 5a and 5c. By contrast, OAM sometimes did not detect all SA1s with exceptionally high rates in a panel (*e.g.*, the SA1 marked with an asterisk in Figure 5d). Finally, in Figure 5e, the smoothing induced by OAM had relatively little effect in terms of classifying SA1s as 'high-rate' based on their proximity to the one with an exceptionally high rate. To determine if the rates of exceptionally high-rate SA1s identified in Figure 5 were transient, corresponding rates for all SA1s in that figure for the previous two Australian financial years (*i.e.*, 2016/17 and 2017/18) were examined. As shown in the Supplementary Material, the SA1s with exceptionally high rates in Figures 5a-c (*i.e.*, in 2018/19) also had exceptionally high rates in the two previous years. By contrast, three of the SA1s that had exceptionally high rates in 2018/19 in Figure 5d did not have correspondingly high rates in the other years. The same was true for the exceptionally high-rate SA1 in Figure 5e. Interestingly, the SA1 marked with an asterisk in Figure 5d, which, as noted, was *not* classified as being 'high-rate' in 2018/19 despite having an excep-

Figure 4. Magnified versions of high-rate regions identified using OAM (Figure 3b). Black lines indicate the boundaries of SA2s encompassing those regions.

tionally high rate in that year, also had an exceptionally high rate in the two previous years.

If only the populations in the three SA1s with both exceptionally and persistently high rates in Figures 5a-c were targeted with scale-appropriate interventions, their collective targeting efficiency could be expressed as 20.8% of MHED presentations for each percentage of population targeted (calculation: 1.6% of MHED presentations divided by 0.08% of the population in 2018/19).

Discussion

This study has described: i) limitations associated with the use of single-aggregation administrative units to delineate sets of geographic target regions for interventions; ii) how smoothing can be applied to mitigate those limitations; and iii) an investigation of local variation in rates within high-rate regions delineated using both aggregate-level administrative units and smoothing. Below

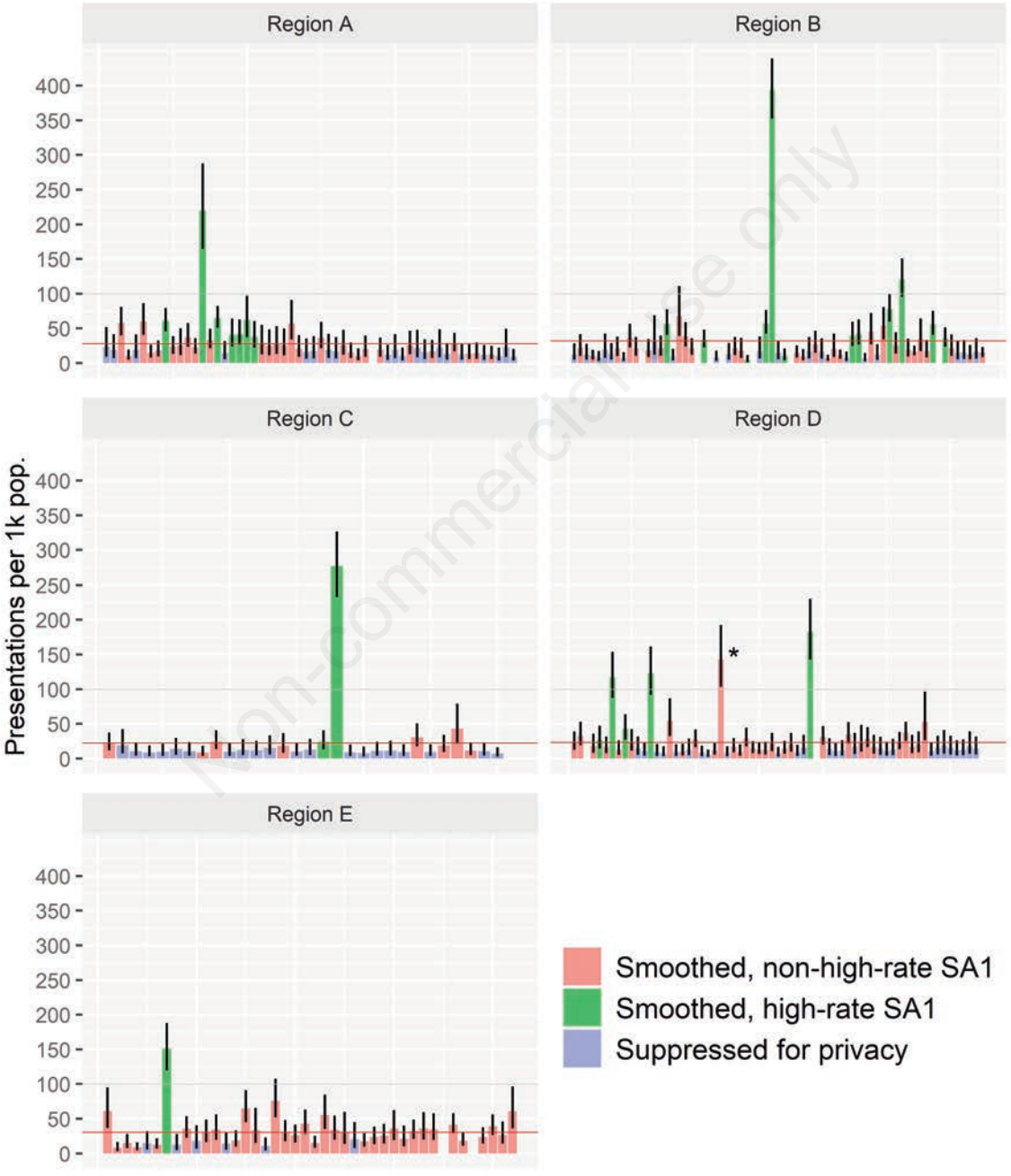

Figure 5. Magnified versions of high-rate regions identified using OAM (Figure 3b). Black lines indicate the boundaries of SA2s encompassing those regions.

we expand upon each of these issues. As noted above, a key focus of this paper has been to reiterate that maps produced using SA2s (and similar units in other jurisdictions, e.g., ZIP codes in the US) may be misleading (Tuson *et al.*, 2020). Population sizes vary widely amongst SA2s (Tuson *et al.*, 2020); consequently, in the present analysis, many SA2s had statistically unstable rates, despite MHED presentations being relatively common. Thus, the use of SA2s to guide the targeting of interventions such as new clinics or outreach services to SA2-sized populations may lead to either i) over-servicing of populations that *don't* have high need, or ii) non-consideration of small, specific neighbourhoods that *do* have high need. As noted, these problems are not specific to SA2s; they impact whenever targeting of interventions is undertaken guided by single-aggregation maps. Further, they will be amplified when units that are larger than SA2s, e.g., SA3s, which have previously been used to examine healthcare variation and reductions in potentially preventable hospitalisations in Australia (ACSQHC, 2017; Queensland Clinical Senate, 2018), are used. Exemplifying the impact of these issues, the targeting efficiency of the high-rate SA2s identified in the present study was less than half that of the OAM-identified regions.

As alluded to above, this paper has demonstrated how smoothing can be utilised to mitigate single-aggregation issues. However, it has also shown some limitations of smoothing. For example, several SA1s nearby those with exceptionally high rates were classified as being 'high-rate' only after smoothing, and some SA1s with exceptionally and persistently high rates that were proximal to high-rate regions were not classified accordingly. These observations illustrate the importance of further investigation of local variation in rates within high-rate regions identified using smoothing. In the present study, undertaking such an investigation, a select number of SA1s with exceptionally high rates that persisted over time were revealed. These regions represent regions of interest for localised, geographically targeted MHED interventions. Their further investigation revealed a variety of plausible risk factors, including: a high density of short-term accommodation; aged care facilities; a small residential area surrounded by industrial zones and liquor stores; public housing; and reports of antisocial behaviour (data not shown). These factors were readily apparent when examining relevant street maps. Interestingly, they were idiosyncratic, rather than systemic; for example, not all SA1s containing liquor stores or aged care facilities had exceptionally high rates. Such observations allow for nuanced generation of hypotheses; however, before causal inference can be made, further investigation is required into features that may have contributed to the high presentation rates, e.g., spatial, or socio-demographic factors (*e.g.*, age, ethnicity, morbidity), and combinations of these features. Timely identification of such features may lead to improved planning of scale-appropriate healthcare responses. It is worth noting that smoothing techniques other than OAM may be used to achieve similar results to those presented here. Indeed, OAM has been shown to produce similar results to those of at least one other smoother in a grid-based simulation (Tuson *et al.*, 2020). Here, we used OAM for several reasons: first, for its intuitive handling of edge effects, which is important when dealing with irregularly-shaped geographical features such as Perth's rivers; second, for its population-adaptive smoothing, which ensures statistical stability of computed rates and facilitates detection of features at scales that are relevant to planned interventions (Tuson *et al.*, 2020); and third, for the unique way in which it ensures protection of privacy, through assigning convolutions of values from many aggregations to each minimal unit (Tuson *et al.*, 2020). The latter feature prevents back-calculation of minimal unit values, and pre-empts application of perturbation techniques, which might otherwise be required to protect privacy (Wieland *et al.*, 2008). However, these features are not unique to OAM. For example, population-adaptive smoothing can also be undertaken using some kernel density methods (e.g., see Carlos *et al.* 2020). Whichever smoothing technique is used, it should be applied, and its performance appropriately examined, in whatever way it is usually used in practice. This might involve calculating statistics such as Moran's I (Moran, 1950), for example, to measure the degree of autocorrelation that remains in the model's residuals. The choice of smoother should also consider whether global or local smoothing is appropriate; here, we have exemplified the application of OAM, which is a local smoother; however, in other contexts, global smoothing, e.g., as achieved through application of Empirical Bayes techniques (Clayton & Kaldor, 1987), might be preferred.

While this paper examines MHED presentations, the methodology employed could also be applied to other conditions and outcomes both in health and beyond, e.g., in ecology, criminology, forestry and mining. However, in each application, the process of identifying features of interest is the same: first, choose an appropriate smoother. If using OAM, this will involve choosing an appropriate target population (or other denominator) size based on the desired degree of granularity while accounting for statistical stability. This choice is related to the choices of kernel and bandwidth that are required when applying kernel smoothers. Typically, the choice of denominator/bandwidth will depend, at least in part, on the relative rarity of the outcome being examined; here, for OAM, since MHED presentations are relatively common, a target population size of 2,000 residents was used. However, for less common conditions, a larger target population size might be appropriate. For example, a previous study examining inpatient admissions for stroke in Perth used a target population size of 10,000 residents (Tuson *et al.*, 2020). One way to determine an appropriate target might be to use the RSE as a guide. Importantly, if a small denominator/bandwidth is used, the relative influence of the rates of individual fine-resolution units (here, SA1s) on the classification as either 'high rate' or not of surrounding units will increase. Second, define a threshold percentage of events to identify high-rate regions. In this study, a threshold of 5% of MHED presentations was used, to limit the number of geographical features of interest for further investigation. And finally, visualise street map, satellite, or other sources of data to generate hypotheses regarding why particular areas might have exceptionally high rates.

MHED presentations are frequently used as markers of mental disease and ill-health for populations. However, in addition to disease, such presentations are influenced by cultural, educational, and proximity factors. Thus, while this study cannot describe the true geographic distribution of mental illness or disease, MHED presentations arguably represent the most complete population level data available (as opposed to self-report or survey data, for example, Althubaiti, 2016; Latkin *et al.*, 2017). It is also possible that different MHED diagnosis groups have different spatial distributions; this was not investigated here due to the focus being on the comparison of smoothed and single-aggregation maps, and on the investigation of localised variation in rates within high-rate regions identified in those maps.

In evaluating differences in rates between areas, confidence intervals were reported but formal significance testing was not undertaken. Instead, rates were examined over time, and, where

patterns persisted, specific SA1s were classified as being areas of interest. This approach was taken partly to pre-empt application of a multiple testing adjustment, e.g., in order to control the false discovery rate (Catelan & Biggeri, 2010), since such adjustments could proceed in a variety of ways, e.g., within each region or year of data or across all years combined, and partly in order not to detract from the paper's primary focus (see above).

Some of the parameters specified in this paper, e.g., the choice of 30% for the RSE threshold in classifying statistical stability, and the number of zonations created for OAM, could vary. As noted previously, the former value was chosen guided by previous research (Werner & Strosnider, 2020); however, other thresholds could be used. A sensitivity analysis was not conducted; however, given that many of the units examined had extremely small population sizes, it is expected that varying the threshold would minimally affect the conclusions drawn. Similarly, it is expected that increasing the number of zonations created would minimally impact results; here, the value 10 was used, based on: i) previous studies (*e.g.* see Cockings & Martin, 2005), and ii) preliminary analyses, which showed minimal variation when using additional zonations (*data not shown*).

Conclusions

Disease maps are widely produced to improve the understanding of the spatial distribution of disease or demand for health services, and thereby to inform efforts to prevent and respond to spatial areas of 'excess'. However, measurement of disease in space must be undertaken in a manner that allows for fulfilment of that purpose. Unfortunately, the current standard practice of using a single set of aggregate-level administrative units to map disease consistently fails this fundamental objective. Addressing this, this study has shown that appropriate smoothing of fine-resolution data can reveal general features of interest while maintaining the desirable qualities of privacy and stability that are characteristic of higher-scale aggregations. However, it has also shown that smoothing can still compromise the description of important fine-resolution features. Therefore, to identify such features, detailed investigation of high-rate regions identified using smoothing is required. Such investigation can be undertaken using the approach described here. Employing this approach, in the present study, several neighbourhoods with exceptionally and persistently high presentation rates were revealed. These observations could help social and health service planners prepare clinical or policy responses that are tailored to the unique and individual needs of the identified high-risk neighbourhoods.

References

Althubaiti A. 2016. Information bias in health research: Definition, pitfalls, and adjustment methods. J Multidiscip Healthc 9:211-217.

Australian Bureau of Statistics (ABS). 2016a. Australian Statistical Geography Standard (ASGS) Volume 1 – main structure and greater capital city statistical area. Catalogue number 1270.0.55.001, ABS, Canberra, ACT, Australia.

Australian Bureau of Statistics (ABS). 2016b. Census of population and housing – Counting persons, place of usual residence (SA1), TableBuilder. Findings based on use of ABS TableBuilder Data.

Australian Commission of Safety and Quality in Health Care (ACSQHC). 2017. Australian atlas of healthcare variation series. Available from: https://www.safetyandquality.gov.au/publications-and-resources/australian-atlas-healthcare-variation-series

Australian Institute of Health and Welfare (AIHW). 2007. Atlas of avoidable hospitalisations in Australia: Ambulatory care-sensitive conditions. Available from: https://www.aihw.gov.au/reports/hospitals/atlas-avoidable-hospitalisations-australia/contents/table-of-contents

Breslow NE, Day NE. 1987. Statistical methods in cancer research. Volume II—The design and analysis of cohort studies. IARC Sci Publ. 82:1-406.

Carlos HA, Shi X, Sargent J, Tanski S, Berke EM. 2010. Density estimation and adaptive bandwidths: A primer for public health practitioners. Int J Health Geogr 9:1–8.

Catelan D, Biggeri A. 2010. Multiple testing in disease mapping and descriptive epidemiology. Geospat Health 4:219-229.

Center for the Evaluative Clinical Sciences, Dartmouth Medical School. 1996. The Dartmouth atlas of health care. Chicago, Ill.: American Hospital Publishing. Available from: https://www.dartmouthatlas.org/

Clayton D, Kaldor J, 1987. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 43:671–81.

Cockings S, Harfoot A, Martin D, Hornby D. 2011. Maintaining existing zoning systems using automated zone-design techniques: methods for creating the 2011 Census output geographies for England and Wales. Environ Plan A 43:2399-418.

Cockings S, Martin D. 2005. Zone design for environment and health studies using pre-aggregated data. Soc Sci Med 60:2729-2742.

Devine OJ, Louis TA, Halloran ME. 1994. Empirical Bayes methods for stabilizing incidence rates before mapping. Epidemiology 5:622-630.

Hodge JK, Marshall E, Patterson G. 2010. Gerrymandering and convexity. College Math J 41:312-324.

Holman CAJ, Bass AJ, Rouse IL, Hobbs MST. 1999. Population-based linkage of health records in Western Australia: development of a health services research linked database. Aust N Z J Public Health 23:453-9.

Independent Hospital Pricing Authority (IHPA). 2019. Australian Refined Diagnosis Related Groups (AR-DRG) Version 10.0. Available from: https://www.aihw.gov.au/reports/hospitals/ar-drg-data-cubes/contents/data-cubes

Kamel Boulos MN, Geraghty EM. 2020. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: How 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. Int J Health Geogr 19:8.

Latkin CA, Edwards C, Davey-Rothwell MA, Tobin KE. 2017. The relationship between social desirability bias and self-reports of health, substance use, and social network factors among urban substance users in Baltimore, Maryland. Addict Behav 73:133-6.

Lessler J, Moore SM, Luquero FJ, McKay HS, Grais R, Henkens M, Mengel M, Dunoyer J, M'bangombe M, Lee EC, Djingarey MH, Sudre B, Bompangue D, Fraser RSM, Abubakar A, Perea

W, Legros D, Azman AS. 2018. Mapping the burden of cholera in sub-Saharan Africa and implications for control: An analysis of data across geographical scales. Lancet 391:1908-15.

Martin D. Extending the automated zoning procedure to reconcile incompatible zoning systems. 2003. Int J Geogr Inf Sci 17:181-96.

Mooney P, Juhász L. 2020. Mapping COVID-19: How web-based maps contribute to the infodemic. Dialogues in Hum Geogr 10:265-270.

Moran PAP. 1950. Notes on continuous stochastic phenomena. Biometrika 37:17–23.

Openshaw S, Taylor PJ. 1979. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. Chapter 5. In: Wrigley N, editor. Statistical applications in the spatial sciences, vol. 127. London: Pion; p. 127–44.

Queensland Clinical Senate. 2018. Dare to compare: reducing unwarranted variation for potentially preventable hospitalisations. Brisbane: Queensland Health. Available from: https://clinicalexcellence.qld.gov.au/priority-areas/clinician-engagement/queensland-clinical-senate/meetings/dare-compare-reducing

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/

St John of God Health Care. 2016. Annual Report 2015–16. Melbourne: St John of God Health Care. Available from: https://annualreport2016.sjog.org.au/

Talbot TO, Kulldorf M, Forand SP, Haley VB. 2000. Evaluation of spatial filters to create smoothed maps of health data. Stat Med 19:2399-2408.

The International statistical classification of diseases and related health problems, tenth revision, Australian modification (ICD-10-AM). 2010. 7th ed. ed. Lidcombe, NSW: Lidcombe, NSW: National Centre for Classification in Health. Available from: https://meteor.aihw.gov.au/content/391301

Tobler, WR. 1970. A computer movie simulating urban growth in the Detroit region. Economic Geography, Supplement: Proceedings. International Geographical Union. Commission on Quantitative Methods 46:234-240.

Tuson M, Yap M, Kok MR, Boruff B, Murray K, Vickery A, Turlach BA, Whyatt D. 2020. Overcoming inefficiencies arising due to the impact of the modifiable areal unit problem on single-ggregation disease maps. Int J Health Geogr 19:40.

Tuson M, Yap M, Kok MR, Murray K, Turlach B, Whyatt D. 2019. Incorporating geography into a new generalized theoretical and statistical framework addressing the modifiable areal unit problem. Int J Health Geogr 18:6.

Waller LA, Gotway CA. 2004. Chapter 4: Visualizing Spatial Data. In Balding DJ, Cressie NAC, Fisher NI, Johnstone IM, Kadane JB, Molenberghs G, Ryan LM, Scott DW, Smith AFM, Tengels JL, eds. Applied spatial statistics for public health data. John Wiley & Sons, Inc., Hoboken, New Jersey. p. 86-104.

Werner AK, Strosnider HM. 2020. Developing a surveillance system of sub-county data: Finding suitable population thresholds for geographic aggregations. Spat Spatiotemporal Epidemiol 33:100339.

Wieland SC, Cassa CA, Mandl KD, Berger B. 2008. Revealing the spatial distribution of a disease while preserving privacy. Proc Natl Acad Sci USA 105:17608-17613.

Online supplementary material:

Figure S1. Rates and 95% confidence intervals for the 2017/18 Australian financial year, for SA1s comprised within SA2s encompassing high-rate regions identified using OAM in 2018/19 (see Figure 5, main text).

Figure S2. Rates and 95% confidence intervals for the 2016/17 Australian financial year, for SA1s comprised within SA2s encompassing high-rate regions identified using OAM in2018/19 (see Figure 5, main text).