

METHODOLOGY

Open Access



Overcoming inefficiencies arising due to the impact of the modifiable areal unit problem on single-aggregation disease maps

Matthew Tuson^{1,2}, Matthew Yap¹, Mei Ruu Kok¹, Bryan Boruff^{3,4}, Kevin Murray⁵, Alistair Vickery¹, Berwin A. Turlach² and David Whyatt^{1*} 

Abstract

Background: In disease mapping, fine-resolution spatial health data are routinely aggregated for various reasons, for example to protect privacy. Usually, such aggregation occurs only once, resulting in 'single-aggregation disease maps' whose representation of the underlying data depends on the chosen set of aggregation units. This dependence is described by the modifiable areal unit problem (MAUP). Despite an extensive literature, in practice, the MAUP is rarely acknowledged, including in disease mapping. Further, despite single-aggregation disease maps being widely relied upon to guide distribution of healthcare resources, potential inefficiencies arising due to the impact of the MAUP on such maps have not previously been investigated.

Results: We introduce the overlay aggregation method (OAM) for disease mapping. This method avoids dependence on any single set of aggregate-level mapping units through incorporating information from many different sets. We characterise OAM as a novel smoothing technique and show how its use results in potentially dramatic improvements in resource allocation efficiency over single-aggregation maps. We demonstrate these findings in a simulation context and through applying OAM to a real-world dataset: ischaemic stroke hospital admissions in Perth, Western Australia, in 2016.

Conclusions: The ongoing, widespread lack of acknowledgement of the MAUP in disease mapping suggests that unawareness of its impact is extensive or that impact is underestimated. Routine implementation of OAM can help avoid resource allocation inefficiencies associated with this phenomenon. Our findings have immediate worldwide implications wherever single-aggregation disease maps are used to guide health policy planning and service delivery.

Keywords: Disease mapping, Modifiable areal unit problem, Single-aggregation disease maps, Zonation-dependence, Resource allocation efficiency

Background

The practice of disease mapping is fundamental to public health [1], with disease maps currently being produced by healthcare organisations worldwide, including the World Health Organisation [2].

A fundamental function of a disease map is to guide geographically-prioritised resource allocation. As such, disease maps have been used to guide the spatial targeting of interventions to address HIV, cholera, Ebola, and malaria [3–11], for example. Disease maps have also been used to guide geographically-targeted responses to the current global COVID-19 pandemic. In Italy, for example, mass COVID-19 testing in the town of Vo proved effective in identifying asymptomatic carriers of the virus, thereby limiting its further spread [12]. Similarly, in

*Correspondence: david.whyatt@uwa.edu.au

¹ Medical School, University of Western Australia, Perth, Australia
Full list of author information is available at the end of the article



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Australia, geographically-localised testing has been carried out in areas with elevated risk of community transmission [13–15].

To evaluate a disease map's utility for this purpose, consideration of its expected 'efficiency' is critical. In this context, efficiency has two key aspects: 'targeting efficiency' and 'logistical efficiency'. The targeting efficiency of a given map can be defined by the percentage of cases that could hypothetically be reached, or 'intervened with' through targeting a certain percentage of the study denominator (usually population). For example, a recent study describing the planned geographically-prioritised distribution of a limited supply of oral cholera vaccines in sub-Saharan Africa [8] proposed that regions with high rates of cholera be prioritised for distribution. On the other hand, the 'logistical efficiency' of a disease map can be represented by the number of distinct geographic 'target regions'. Since both intervention resources and the capacity of authorities to intervene are generally limited, it is essential to balance the targeting and logistical efficiency of a given disease map.

In practice, disease maps are often produced at a single, more-coarse resolution than that at which data were collected or made available (the 'minimal' resolution). This is in order to overcome issues associated with examining fine-resolution data, including: (i) violation of privacy, (ii) infeasible computation, and (iii) the small number problem, which manifests in unstable estimates of disease risk [16]. Since data underlying such maps are essentially an aggregation of minimal-resolution data, they can be termed 'single-aggregation disease maps'. In Australia, for example, population data are obtainable by Statistical Areas Level 1 (SA1; mean population size approximately 400 residents [17]), but disease maps are usually produced by larger areas such as SA2 (between approximately 3,000 and 25,000 residents) or SA3 (between approximately 30,000 and 130,000 residents) [18–20]. This is partly due to extensive ethical clearances limiting access to fine-resolution health data, and partly due to the 'small-number' issues listed above. Similarly, in the US and UK, for example, disease maps are often produced by either county or census ward, despite finer-resolution data being obtainable [21, 22]. As a third example, the previously-cited study of cholera used a grid of 20 × 20 km units in preference to potentially finer-resolution grids, in part due to computational constraints [8].

Unfortunately, single-aggregation disease maps are undermined through being dependent on the chosen set of aggregate-level mapping units. This dependence is described by the modifiable areal unit problem (MAUP). First described in the literature in 1934 [23], and coined in 1979 [24], the MAUP has since been widely

investigated in geography, statistics, and elsewhere. In disease mapping, the MAUP describes how mapped values depend on both the scale of spatial aggregation (the scale aspect) and the position of boundaries between spatial units (the zonation aspect) [25]. These aspects manifest in various misleading phenomena; for example, disease clusters that are relatively small compared to the chosen units, or located where several units intersect, might not be detected.

Despite its extensive literature, in practice, the MAUP is rarely acknowledged, including in disease mapping. This is evidenced by a 2014 review, which noted that the MAUP was recognised in only 1% of papers using spatially aggregated data [26]. More recently, there is little evidence to suggest acknowledgement of the MAUP has increased, with papers containing single-aggregation disease maps but not acknowledging the MAUP constantly being published in top-ranked journals (e.g. see [27–29]). On the other hand, papers that do acknowledge the MAUP often simply demonstrate how different results can be obtained through examining administrative units at different scales (e.g. see [30]). This is consistent with a general tendency of the literature to examine the MAUP's scale aspect, as opposed to its zonation aspect [31]. While some association studies *have* examined the impact of the MAUP's zonation aspect (e.g. see [24, 32, 33]), to our knowledge, no similar study exists in disease mapping. Further, as far as we are aware, no previous study has investigated the impact of the MAUP in the context of efficiency. Such an investigation is critically required, given ubiquitous reliance on single-aggregation disease maps to guide distribution of limited healthcare resources.

Therefore, we introduce the overlay aggregation method (OAM) for disease mapping. Extending a recently suggested approach [33], OAM combines many aggregate-level disease maps to derive a single, minimal-resolution map. Using simulated data, we characterise OAM as both: i) a novel framework to systematically quantify the impact of the MAUP's zonation aspect on single-aggregation disease maps, and ii) a novel smoothing technique that overcomes both this impact and the small-number issues listed previously. We compare OAM to single-aggregation disease maps by their targeting and logistical efficiency, and demonstrate how OAM effectively balances those aspects. We then apply OAM to a real-world dataset: ischaemic stroke hospital admissions (stroke) in Perth, Western Australia (WA), in 2016. Strokes require rapid intervention to avoid irreparable nerve tissue damage [34], necessitating ongoing consideration of patients' access to essential stroke services such as specialist hospital units or ambulance depots.

Understanding the spatial distribution of stroke supports this endeavour through guiding placement of such services.

Methods

Overlay Aggregation Method (OAM)

OAM comprises five steps:

1. Define the minimal units;
2. Set parameter values:
 - a. The number of aggregate-level maps; and
 - b. Their geographical scale of aggregation;
3. Create multiple sets of aggregate-level mapping units, or 'zonations';
4. Create aggregate-level disease maps based on those zonations; and
5. Combine the aggregate-level disease maps to produce a single, minimal-resolution map.

These steps are described in detail below.

Step 1. Define the minimal units

Since its first step involves specifying a set of minimal units, OAM operates in the same way as applied to either truly areal data (e.g. such as that often collected in surveys) or data spatially aggregated from an original point location form. This specification will often be guided by data availability; in Australia, for example, SA1s might often be defined as the minimal units due to health data generally not being obtainable below that resolution. By contrast, if point location data were available, then minimal units substantially smaller than SA1 (or their equivalent in another country) could be defined.

Step 2. Set parameter values

The number of maps and their geographical scale of aggregation together govern the degree of smoothing in OAM's output. We will show how one might choose an appropriate number of maps guided by observation of a minimal degree of change in OAM's output as the number of maps is increased. By contrast, the choice of scale should be guided by the characteristics of a planned intervention (see Sect. [Discussion](#)).

Step 3. Create the aggregate-level zonations

OAM's zonations must completely segment the geographical study area. Following previous authors [32, 33], we show how suitable zonations can be created using the freely-available software AZTool [35, 36]. Briefly, AZTool operates by zoning a set of minimal units according to a target denominator size and potentially other constraints

(e.g. minimum and maximum threshold denominator sizes). To create OAM's zonations, only specification of target and minimum threshold denominator sizes is required.

Step 4. Create the aggregate-level disease maps

Any model deemed appropriate by the user can be fitted to OAM's zonations; OAM in no way advises the choice of model, since its focus is the appropriate *preparation* of available data, rather than its modelling. For the simulated dataset, we created crude rate maps for each of OAM's zonations, while for stroke, we fitted the widely implemented Besag, York & Mollie (BYM) model [37, 38]. The latter modelling was undertaken using the Integrated Nested Laplace Approximation (INLA) approach [39], implemented in the R [40] package INLA (<https://www.r-inla.org>). INLA allows for approximate Bayesian inference in latent Gaussian models. For model formulae and R code, see [41].

Step 5. Combine the aggregate-level disease maps

In this step, minimal unit values are derived through weighting the values of aggregate-level units in which they are comprised (one per zonation), using the formula:

$$v_m = \frac{\sum_{p \in P} \left[\frac{d_m}{d_p} \times v_p \right]}{\sum_{p \in P} \left[\frac{d_m}{d_p} \right]} \quad (1)$$

where p indexes the set P of aggregate-level units comprising minimal unit m ; d_m and d_p are the denominator values for units m and p , respectively; v_p is the value of unit p ; and v_m is the denominator-weighted mean value derived for unit m . For both the simulation and stroke, d_m and d_p are population sizes. For the simulation, v_p is the crude rate of unit p and v_m is the population-weighted mean crude rate of unit m . For stroke, v_p is the estimated relative risk (RR) of unit p and v_m is the population-weighted mean RR of unit m .

Equation (1) is undefined for minimal units with zero denominator values. Therefore, to maintain the smooth appearance of the final map, d_m can be cancelled in Eq. (1) to give:

$$v_m = \frac{\sum_{p \in P} \left[\frac{1}{d_p} \times v_p \right]}{\sum_{p \in P} \left[\frac{1}{d_p} \right]} \quad (2)$$

where all parameters are as defined previously. Equation (2) will output positive values for minimal units with zero denominators and identical values to Eq. (1) otherwise. All OAM results presented below were derived using Eq. (2).

Simulation

A point location dataset of 100 disease cases was generated through specifying a spatially-correlated random field and assigning cases across that field according to a multinomial probability distribution. We suppose that population size is the denominator of interest, with population data available at the resolution of a 100×100 unit grid where each cell has a population size of one. Thus, the total population size is 10,000. We define a 20×20 unit grid to be the set of minimal units; thus, each minimal unit has a population size of 25. Note: population size is defined to be the denominator due to this being the most common situation in practice; however, other denominators, e.g. geographical area, could be used.

We produced three maps of the simulated dataset:

- i) A minimal-resolution map of crude rates;
- ii) A single-aggregation map of crude rates, based on a 5×5 unit grid (population size: 400 per unit); and
- iii) A minimal-resolution map of population-weighted mean crude rates, using OAM.

We used AZTool to create 100 aggregate-level zonations for OAM, based on target and minimum threshold population sizes of 400 and 300, respectively. The former value was chosen in order that the single-aggregation and OAM maps be comparable.

Stroke

Study area

2016 Australian Census SA boundaries for Perth were obtained from the Australian Bureau of Statistics (ABS). Perth was defined to comprise the four Greater Perth 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. The latter were excluded due to Rottnest Island operating primarily as a tourist and day-trip destination and Garden Island housing the Australian Navy's largest fleet base. Sets of minimal and single-aggregation units were defined to be SA1 and SA2, respectively.

Outcome data

Stroke records for Perth in 2016 were extracted from the WA Hospital Morbidity Data Collection (HMDC) [42] and aggregated by SA1. Following [43], we defined stroke admissions to be those with a principal International Statistical Classification of Diseases and Related Health Problems, tenth revision, Australian Modification (ICD-10-AM) [44] diagnosis of I63 (Cerebral infarction), I64 (Stroke, not specified as haemorrhage or infarction), or H34.1 (Central retinal artery occlusion). In total, 3,534 stroke admissions were extracted for WA, of which 11

(0.3%) were excluded due to not having an SA1 of residential address recorded. A further 996 were excluded due to occurring among individuals residing outside of Perth (i.e. in regional WA), and of the remaining admissions, four (0.2%) were excluded due to occurring in SA1s with zero population sizes as reported by the ABS. Thus, a total of 2,523 admissions were available for analysis.

Population data

Population data for the 2016 Australian Census were extracted from the ABS, stratified by SA1.

Analysis

Mirroring the simulation analysis, we produced:

- i) An SA1-resolution map of RRs;
- ii) An SA2-resolution map of RRs; and
- iii) An SA1-resolution map of population-weighted mean RRs, using OAM.

We again used AZTool to create 100 aggregate zonations for OAM, based on target and minimum threshold population sizes of 11,250 and 9,000, respectively. The former value was chosen to approximately match the mean population size of SA2s (11,256). Across all zonations, the median number of SA1s per aggregate unit was 26 (95% quantile interval: 18–33).

Efficiency

For both the simulation and stroke, we compared maps by their:

- i) 'Targeting efficiency curves';
- ii) 'Targeting efficiency maps';
- iii) 'Logistical efficiency curves'; and,
- iv) For exemplar 'target case percentages', 'logistical efficiency maps'.

These outputs are described in detail below.

Targeting efficiency curves

The targeting efficiency curve for a given disease map plots the cumulative percentage of the denominator 'targeted' against the cumulative percentage of cases reached. It is created through 'targeting' the mapping units by their values, in descending order.

Targeting efficiency maps

A targeting efficiency map is a spatial representation of a targeting efficiency curve. In such maps, units are coloured according to the cumulative percentage of cases they collectively contain, following their target order.

Logistical efficiency curves

A logistical efficiency curve plots the number of discontinuous regions requiring targeting to reach different target case percentages (i.e. cumulative percentages of cases). For both the simulation and stroke, discontinuity was determined using rook adjacency.

Logistical efficiency maps

A logistical efficiency map displays the target regions associated with a particular target case percentage. It is essentially a dichotomised version of a targeting efficiency map. We created logistical efficiency maps for the

simulation and stroke based on exemplar target case percentages of 50% and 15%, respectively.

Results

Minimal-resolution analysis of the simulated dataset

Figure 1a maps the simulated disease case locations and the minimal units. The cases are broadly grouped in a vertical band to the right of centre of the study area, with variations in grouping within that band.

Figure 1b shows the minimal-resolution disease map of crude rates.

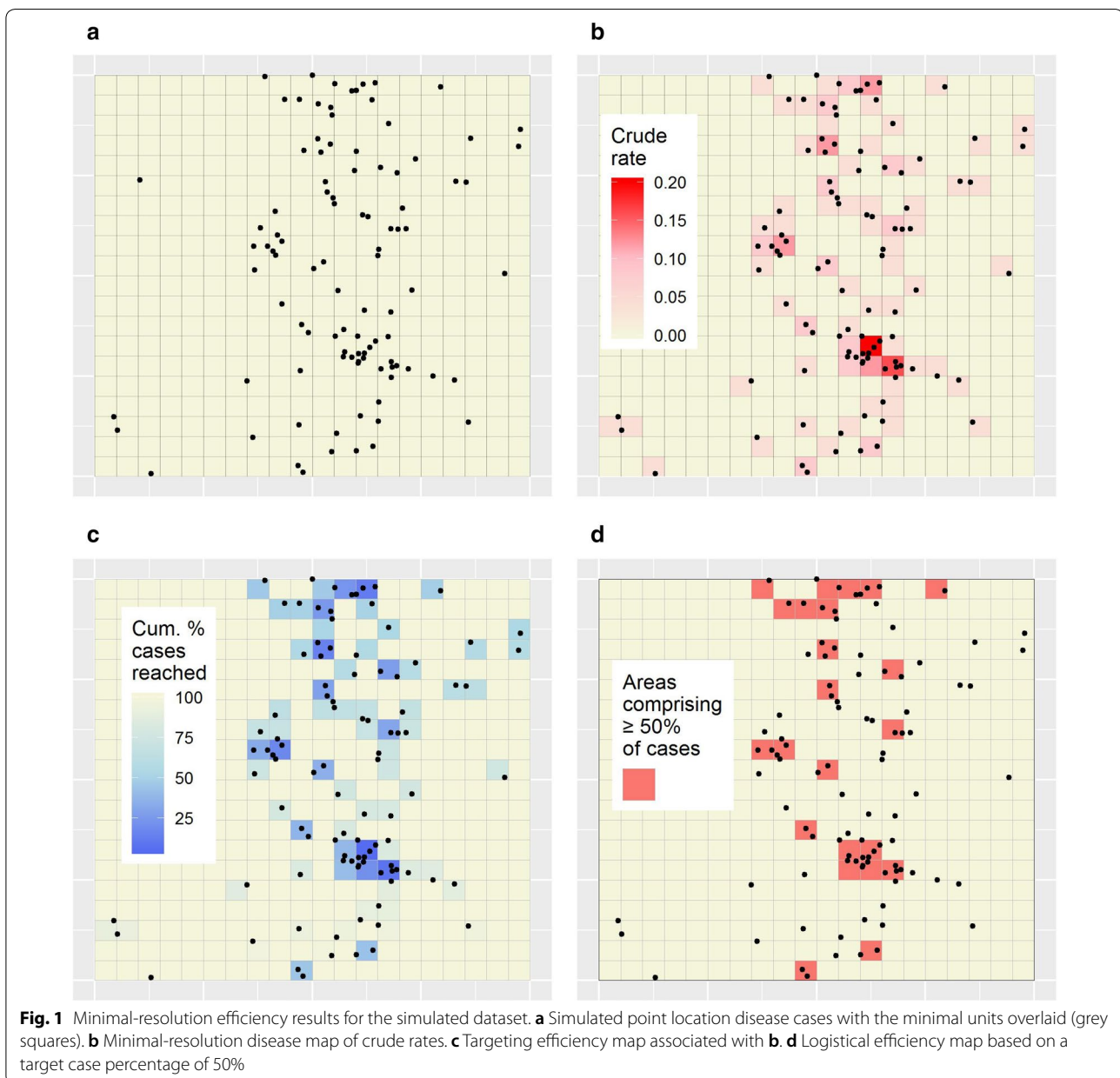


Figure 2a shows the targeting efficiency curve associated with Fig. 1b. This curve indicates that 50% of cases could hypothetically be reached through targeting just 23 minimal units, or 5.8% of the population (Table 1).

Figure 1c shows the targeting efficiency map corresponding to Fig. 2a. In this map, the 23 units noted above are coloured blue-to-light blue.

Figure 2b shows the logistical efficiency curve associated with Fig. 1c. This curve indicates that a maximum of 33 regions require targeting to reach any target case percentage (this was for a target case percentage of 100%).

Finally, Fig. 1d shows the logistical efficiency map associated with the target case percentage of 50%. It shows that the 23 units noted above form 13 discontinuous regions (Table 1). Note: the legend indicates that $\geq 50\%$ of cases are contained within those regions, since, in general, a larger percentage of cases than the specified target might be contained within the regions requiring targeting to reach it. In Fig. 1d the target regions contain exactly 50% of cases, so the \geq symbol is for illustration only.

Single-aggregation analysis of the simulated dataset

Figure 3a maps the simulated disease case locations and the single-aggregation units. Figure 3b shows the single-aggregation map of crude rates; Fig. 2a the single-aggregation targeting efficiency curve; Fig. 3c the targeting efficiency map; Fig. 2b the logistical efficiency curve; and Fig. 3d the logistical efficiency map based on the target case percentage of 50%. Figure 2 demonstrates decreased targeting efficiency but increased logistical efficiency for the single-aggregation strategy compared to the minimal-resolution result. For example, using the single-aggregation strategy, 50% of cases could hypothetically be

Table 1 Exact efficiency data for the simulation

Mapping method	Cum. % of population targeted	Number of target regions
Min	5.8	13
OAM	12.5	3
Agg	20	2

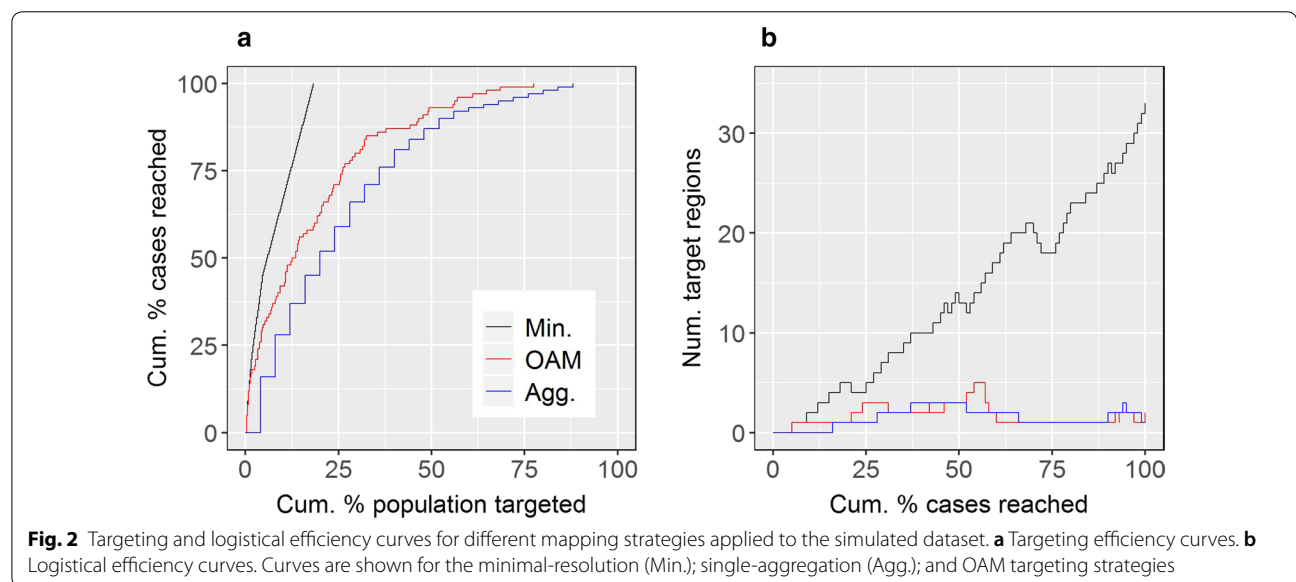
Shown are cumulative percentage of population targeted and number of target regions values associated with a target case percentage of 50%, for the minimal-resolution ('Min.'). single-aggregation ('Agg.') and OAM targeting strategies.

reached through targeting 20% of the population within two discontinuous regions (Table 1).

OAM analysis of the simulated dataset

Clearly, a trade-off exists between the targeting and logistical aspects of efficiency. A minimal-resolution strategy will always maximise targeting efficiency, but at the expense of logistical efficiency. This is the impact of the small number problem. Further, while computational expense is not presently an issue, privacy is likely to be violated by the presentation of rates based on small numbers in Fig. 1b. The single-aggregation strategy overcomes these issues, but at the expense of targeting efficiency. This is the impact of the MAUP. Thus, a targeting strategy is needed that balances the targeting and logistical aspects of efficiency; here, we demonstrate how OAM constitutes such a strategy.

Figures 4a–c show three of OAM’s 100 zonations, and Figs. 4d–f crude rate maps based on those zonations. Figure 4g shows the minimal-resolution map of population-weighted mean crude rates produced using OAM.



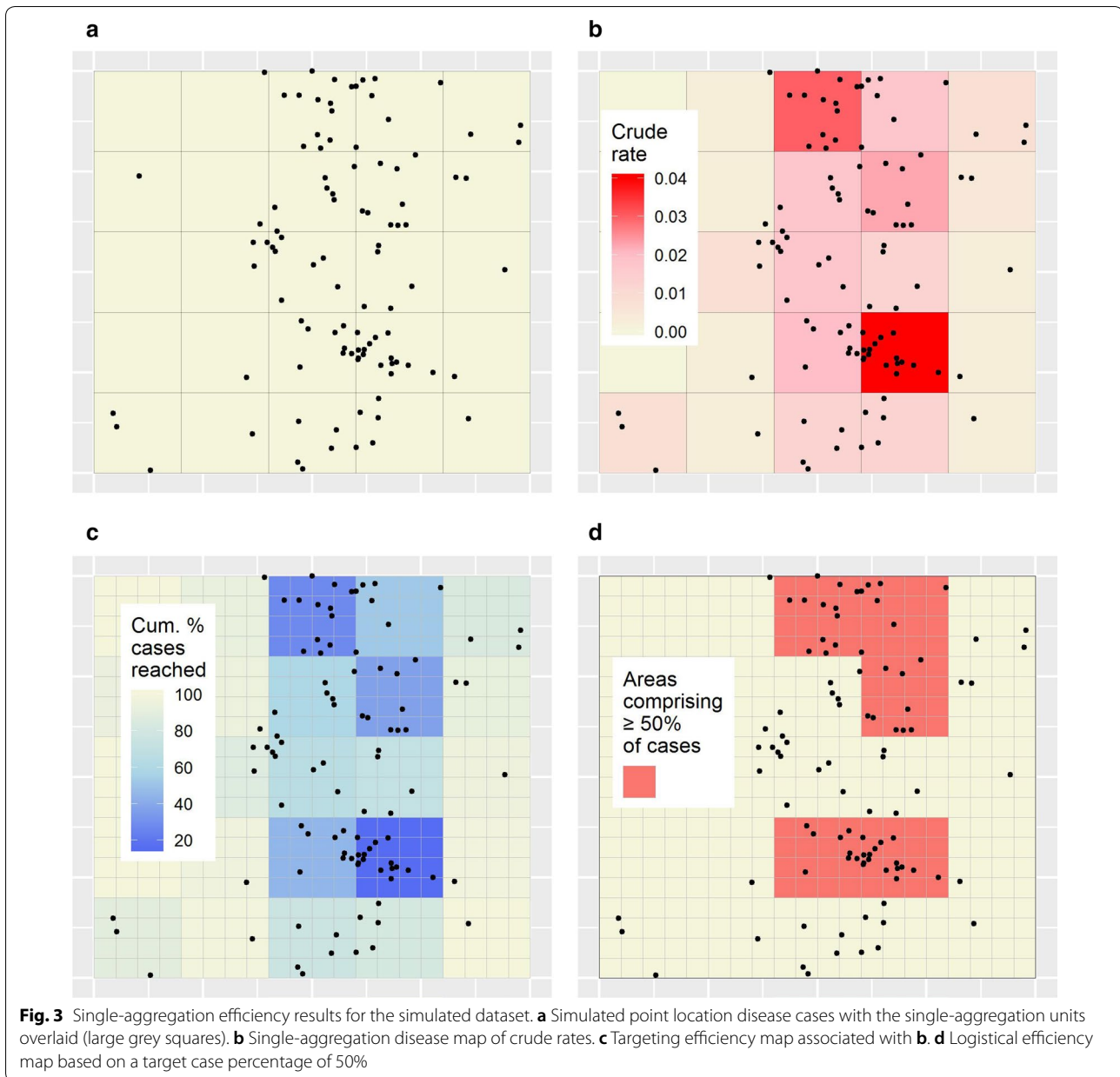


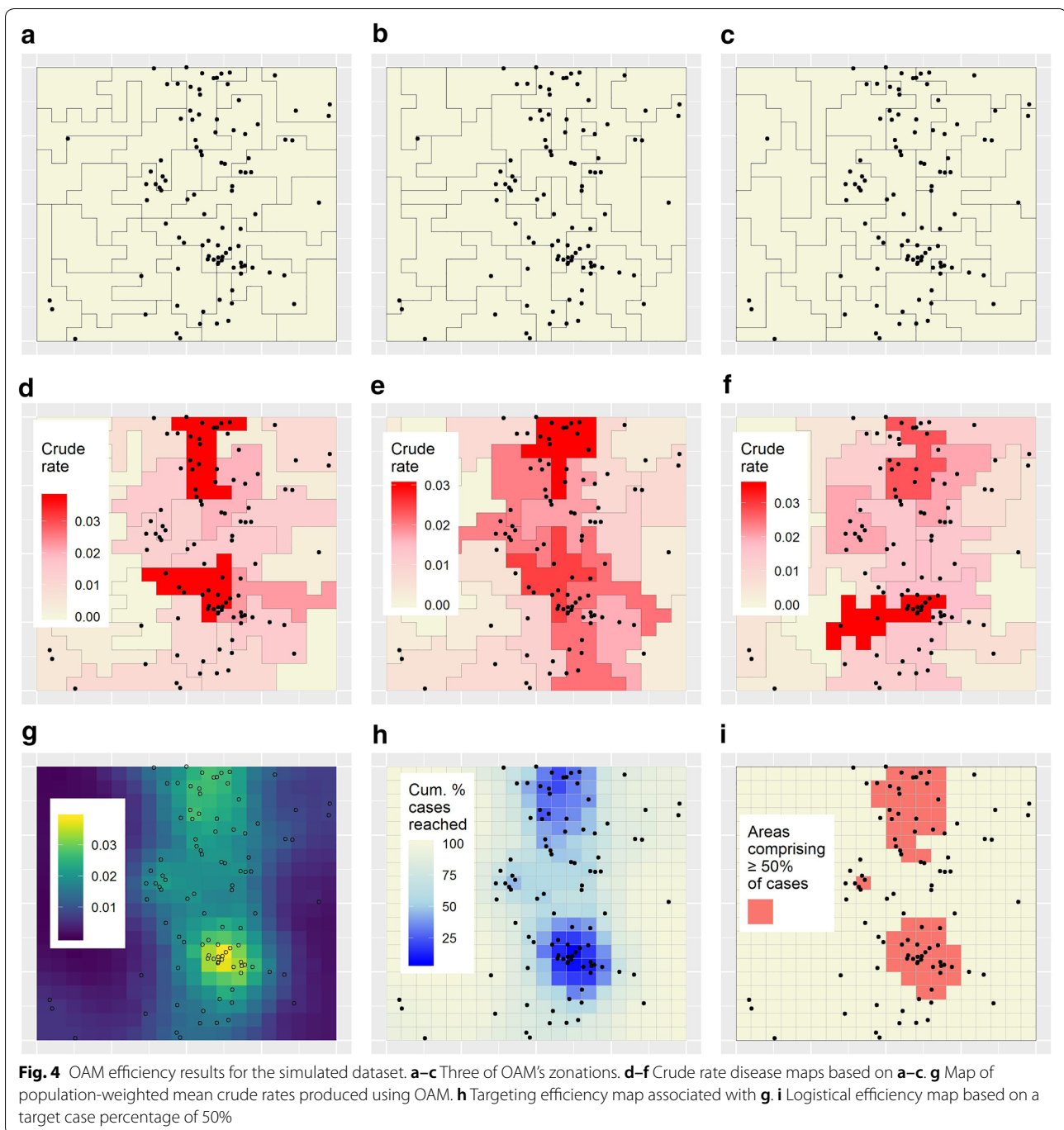
Figure 2a shows the targeting efficiency curve associated with Fig. 4g. The curve lies between the minimal-resolution and single-aggregation curves, illustrating how OAM compromises between those strategies with respect to targeting efficiency. For example, using OAM, 50% of cases could hypothetically be reached through targeting 12.5% of the population (Table 1).

Figure 4h and Fig. 2b show OAM's targeting efficiency map and logistical efficiency curve, respectively. The latter indicates comparable or somewhat decreased logistical efficiency for OAM as compared to the single-aggregation strategy. For example, using OAM, three

discontiguous target regions contain the 12.5% of the population noted above (Table 1). These regions are mapped in Fig. 4i.

Sensitivity to key parameter values

To investigate OAM's sensitivity to the number and set of zonations, we created nine additional sets of 100 zonations using AZTool and applied OAM to each. This analysis is described in Additional file 1: 'Additional results' (section 'Sensitivity to key parameter values'). Minimal differences in efficiency were observed between zonation sets and as the number of zonations



approached 100, suggesting that the choice of zonation set is relatively immaterial and that 100 zonations is sufficient to mask idiosyncrasies of particular zonations, at least for the simulated dataset examined here.

OAM as a smoothing technique

Figure 4g has the appearance of being a smooth, albeit relatively grainy surface. This effect derives from the

effective creation of spatial smoothing kernels within OAM, characterised by distance decay in the sense that, in the zoning process, nearby units are generally grouped more often than those more distant. Additional file 1: 'Additional results' (section 'Effective smoothing kernels') illustrates this phenomenon, showing effective smoothing kernels created for two minimal units. Thus, OAM can be characterised as a smoothing technique, with the

degree of smoothing governed by the target denominator size and the number of zonations. This is demonstrated in Additional file 1: 'Additional results' (section 'OAM as a smoothing technique'), where we apply OAM to the simulated dataset based on target population sizes of 100, 200, 400, and 800.

Comparison to existing smoothing techniques

Characterisation of OAM as a smoothing technique suggests comparison to existing smoothing techniques. As an example, we derived a smoothed map of the simulated dataset using a kernel-smoothed spatial density technique implemented in the R package *sparr* [41, 45]. Specifically, we used the *bivariate.density* function in that package. Details of this analysis are provided in Additional file 1: 'Additional results' (section 'Comparison to existing smoothing techniques'). At two different resolutions, maps produced using the *bivariate.density* function and their associated efficiency were markedly similar to those of OAM.

Global and local zonation dependence

OAM's incorporation of multiple single-aggregation disease maps uniquely facilitates investigation of the impact of the MAUP's zonation aspect on such maps. To illustrate, in each of OAM's component maps, we classify as hotspots those units whose lower crude rate confidence bounds exceed the overall rate of 0.01. Confidence bounds for each unit were derived using the *pois.exact* function implemented in the R package *epitools* [46], based on a confidence level of 64%. However, both the method used to derive confidence bounds and the choice of confidence level are relatively immaterial, since they do not affect the map ultimately produced using OAM or its associated efficiency (see Additional file 1: 'Additional results' (section 'Justification of the zonation-dependence confidence level')). As a rule of thumb, we suggest choosing a confidence interval such that the hotspot counts derived below range between zero and 100, i.e. the full dynamic range for that quantity.

Figures 5a–c show hotspots classified based on Figs. 4d–f. Clearly, different hotspots are detected in each map, demonstrating how single-aggregation maps are unreliable for planning purposes. Following [34], but implementing a variation of Eq. (5) in that paper, such differences can be quantified between all pairs of zonations by the average probability that a minimal unit appearing in a hotspot in any given zonation will *not* appear similarly in another. This value measures the global 'zonation-dependence' of single-aggregation hotspots. To calculate it, consider that, for minimal units appearing in hotspots in a given zonation z_i , the probability of *not* appearing in a hotspot in a new, different zonation z_j can be calculated as:

$$p_{j|i} = \frac{\text{num. min. units in hotspots in zonation } z_i \text{ but not zonation } z_j}{\text{num. min. units in hotspots in zonation } z_i} \quad (3)$$

Then, for any given minimal unit appearing in a hotspot in a particular zonation z_i , the average probability $p_{\cdot|i}$ of that unit not appearing in a hotspot in another, different zonation may be calculated as the mean of $p_{j|i}$ over all $j \neq i$. The mean of values $p_{j|i}$ may then be presented with its associated uncertainty represented by the 2.5% and 97.5% quantiles of the distribution of all values $p_{\cdot|i}$. Using Eq. (3), we calculated an average zonation-dependence probability of 0.41 (95% quantile interval: 0.34–0.5) for the simulated dataset. This suggests that hotspots classified in single-aggregation disease maps of that dataset are moderately zonation-dependent.

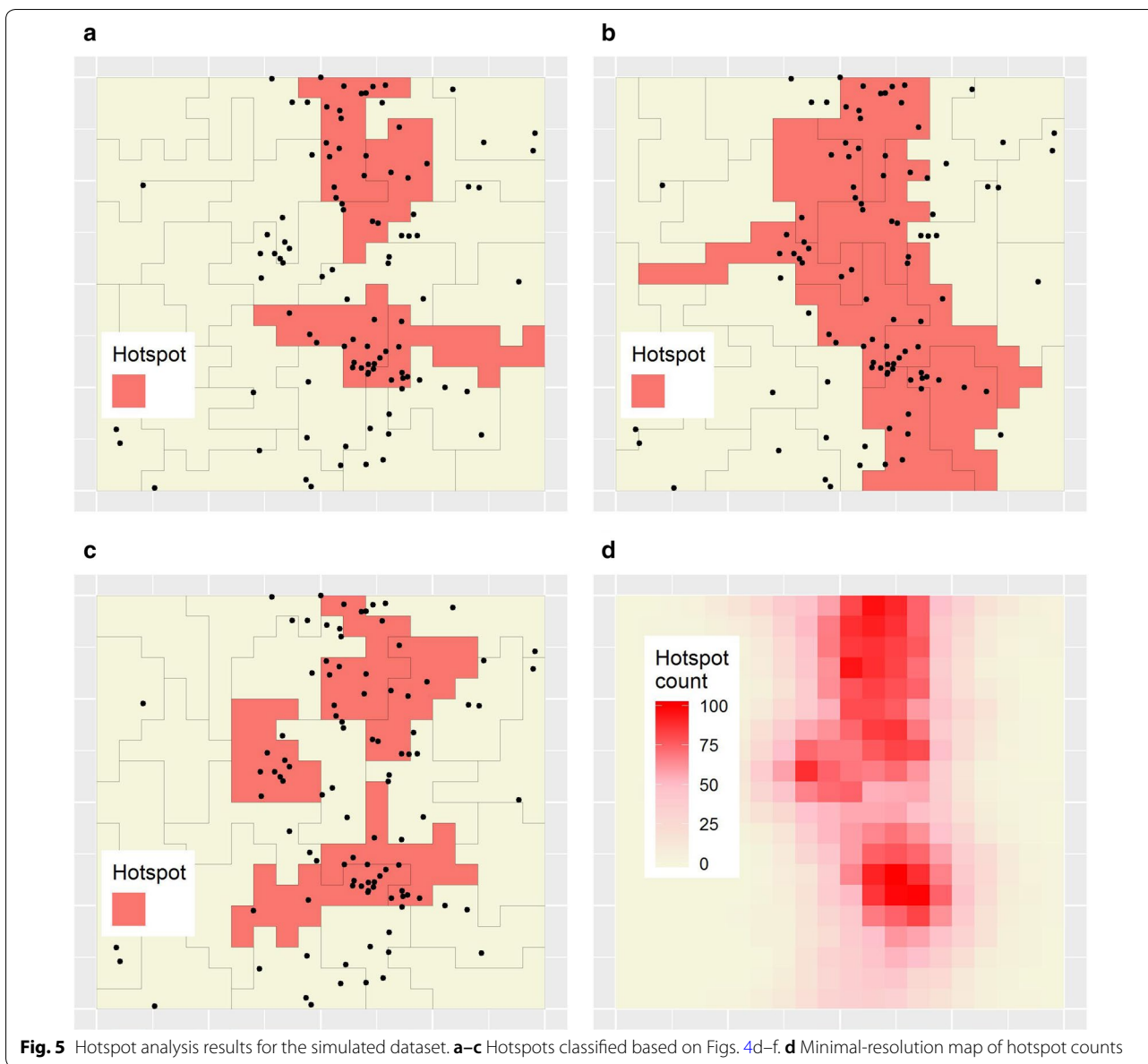
To characterise local zonation dependence, again following [34], we derived a minimal-resolution 'hotspot count'. This value records the number of zonations in which a given minimal unit appears in a hotspot. It is calculated as:

$$HSC_m = \sum_{p \in P} [HS_p] \quad (4)$$

where parameters m , p , and P are as defined previously; HS_p is an indicator variable recording whether or not unit p was a hotspot in its zonation; and HSC_m is the hotspot count derived for unit m . Figure 5d shows the hotspot count map for the simulated dataset.

The classification as hotspots (or non-hotspots) of minimal units with hotspot counts equal to zero or 100 is *zonation independent*, i.e. independent of the aggregate-level zonation used. By contrast, the classification of minimal units with hotspot counts between one and 99 is, to varying degrees, *zonation dependent*. Exploring this, we define 'zonation-dependent negatives' (ZDNs) to be minimal units classified as hotspots in most (i.e. ≥ 80), but not all zonations. Using this threshold, there were 26 ZDNs. Similarly, we define 'zonation-dependent positives' (ZDPs) to be minimal units appearing in hotspots in at least one, but relatively few (i.e. ≤ 20) zonations. Using this threshold, there were 139 ZDPs. Both ZDPs and ZDNs should disturb policymakers, since they effectively represent relatively zonation-independent hotspot or non-hotspot regions that might not be classified as such due to idiosyncrasies of particular single-aggregation disease maps. Illustrating the potential extent of this problem, 41.3% of minimal units were classified as either ZDPs or ZDNs.

Note: the ≥ 80 and ≤ 20 thresholds used above are arbitrary. Depending on the situation, different thresholds might be appropriate; for example, classification of both ZDPs and ZDNs using a threshold of 50 would result in a discussion of regions classified as hotspots (or



non-hotspots) in the *majority* of zonations, rather than in *most*.

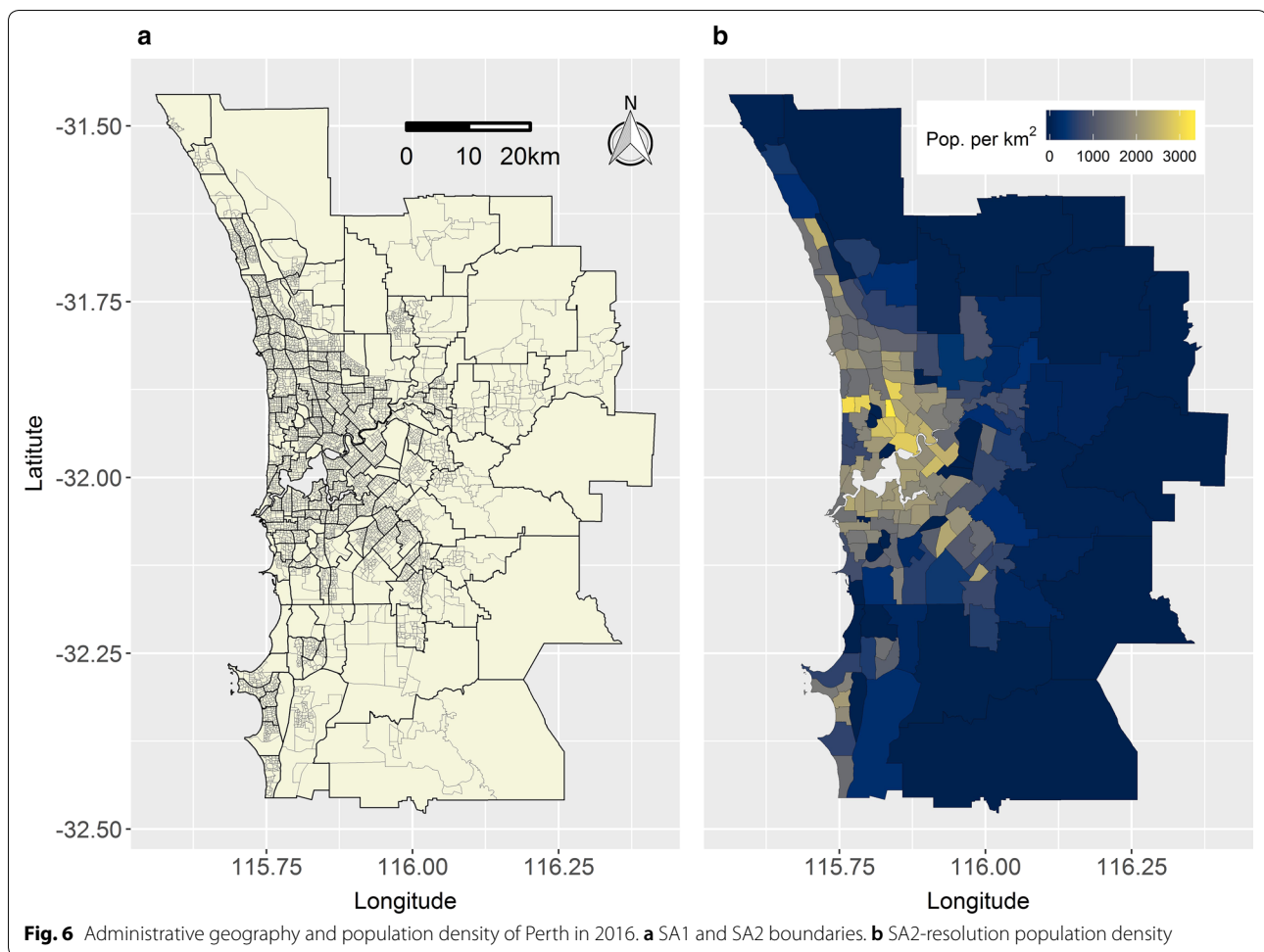
Stroke

Figure 6a maps the 4,248 SA1s and 164 SA2s in Perth, and Fig. 6b Perth’s 2016 population density by SA2. Perth’s population, which in 2016 was 1.85 million, straddles the Swan and Canning Rivers inland and sprawls north to south along the coastline.

SA1 and SA2 maps of stroke RR are not shown in order to maintain a manageable figure list. However, Fig. 7 shows the map of population-weighted mean RRs produced using OAM. Note: values in this map have been

perturbed by a small amount to further protect privacy. Several regions with high weighted RRs are observed, in particular two regions located on the southern coastline and inland in the south-east.

Figures 8a-b show the targeting and logistical efficiency curves for the SA1, SA2, and OAM targeting strategies. Consistent with previous results, the SA1 strategy had optimal targeting efficiency but poor logistical efficiency; for example, 15% of cases could hypothetically be reached through targeting 1.4% of the population within 63 discontinuous regions (Table 2). By contrast, the SA2 strategy was characterised by decreased targeting efficiency but increased logistical efficiency compared to



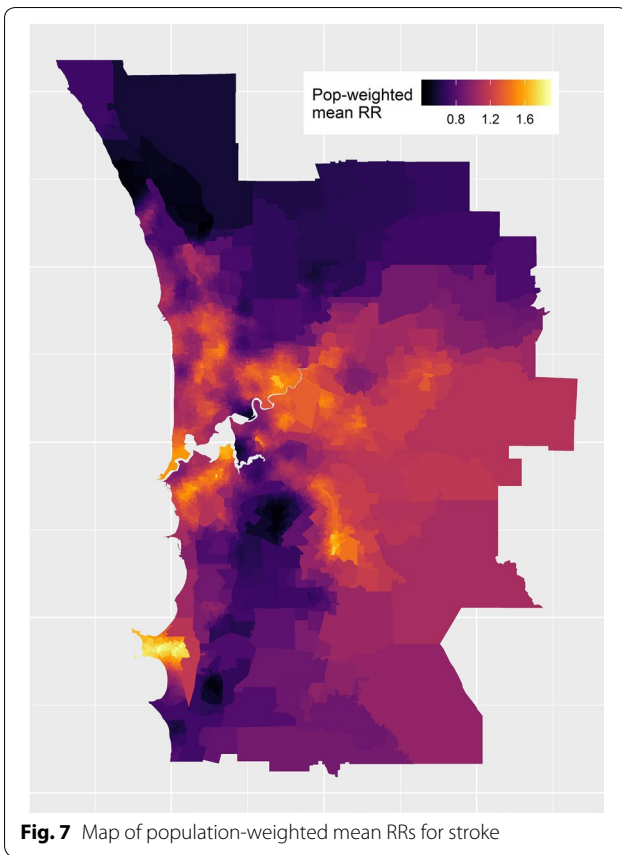
the SA1 result (Fig. 8). For example, 15% of cases could be hypothetically be reached through targeting 8.6% of the population within 11 discontinuous regions (Table 2). Finally, OAM balanced targeting and logistical efficiency (Fig. 8); using OAM, to hypothetically reach 15% of cases, 5.9% of the population would need to be targeted within 15 discontinuous regions (Table 2). Figures 9a–c show the logistical efficiency maps corresponding to the target case percentage of 15%, for the three strategies.

Figures 10a–c show hotspots classified among SA2s and in two of OAM's zonations. Here, following the rule of thumb proposed previously, hotspots were defined to be those areas whose lower 81% credible interval bounds exceeded 1. As for the simulated dataset, different hotspots are detected in each map in Figs. 10a–c, reinforcing the previous conclusion that single-aggregation disease maps are unreliable for planning purposes. We calculated a global probability of zonation-dependence of 0.52 (95% quantile interval: 0.48–0.56); this suggests that, on average, a hotspot in a given single-aggregation disease map

of stroke will, as often as not, *not* be similarly classified in a different map of the same data.

Figure 10d maps the SA1 hotspot counts for stroke. Applying the same thresholds used previously (i.e. ≥ 80 and ≤ 20), we classified 133 stroke ZDNs and 1,823 stroke ZDPs. Together, these regions comprised almost half (46%) of the entire study area.

Finally, to explore the impact on efficiency of implementing a complex model versus no model within OAM, Additional file 1: 'Additional results' (section 'Crude rate analysis of stroke') describes a crude rate analysis of stroke. As might be expected, regardless of the targeting strategy (i.e. SA1, SA2, or OAM), the crude rate analysis was characterised by increased targeting efficiency but decreased logistical efficiency compared to the BYM model analysis. This is due to the additional smoothing of the underlying data that occurs when using the BYM model.



with high risk of stroke, in particular two regions located on Perth’s southern coastline and inland in the south-east. In planning interventions, these regions might be considered alongside information such as proximity to specialist stroke units or ambulance depots. An example intervention might be the deployment of mobile stroke units, such as has occurred in some cities in recent years [47]. However, it is not within scope for this paper to advocate or evaluate the feasibility of any particular intervention.

Perhaps most critically, we have demonstrated how the classification of hotspots (or non-hotspots) in single-aggregation disease maps may depend on the particular aggregate-level zonation used. Therefore, no result in any single-aggregation map is reliable. OAM constitutes a novel framework to quantify this phenomenon and describe its local effects. However, we note the extent of the problem is likely to vary by case. Where disease clusters are larger than the chosen mapping units, agreement between single-aggregation maps based on different, but similarly sized units is likely to be high. An example of this might be the mapping of under-5 mortality in sub-Saharan Africa using a grid of 5×5 km units [48]. In that study, regions with high mortality rates were generally large compared to the chosen units; therefore, it is likely that different configurations of the chosen grid would have identified approximately similar high-rate regions. By contrast, where disease clusters are smaller than the chosen units, agreement between single-aggregation maps based on alternative configurations of those units is likely to be lower. An example of this might be the previously-cited study of cholera in sub-Saharan Africa [8], where many of the classified hotspots were relatively

Discussion

We have introduced OAM and demonstrated its utility in efficiently mapping the geographical distribution of stroke in Perth in 2016. OAM identified several regions

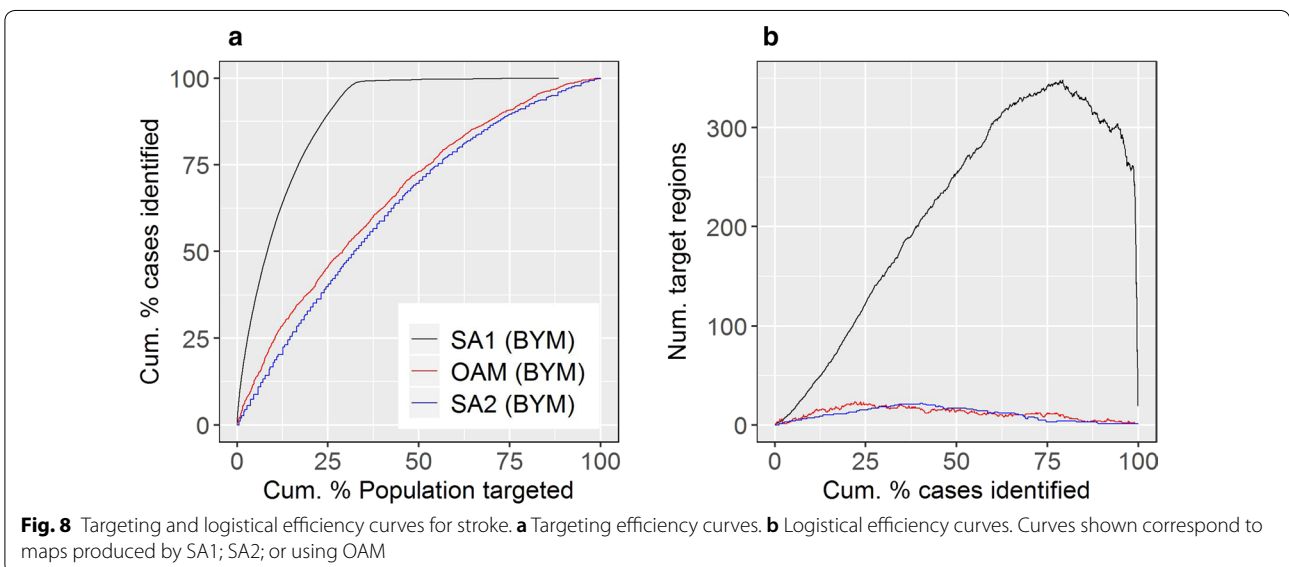


Table 2 Exact efficiency data for stroke

Mapping method	Cum. % of population targeted	Number of target regions
SA1	1.4	63
OAM	5.9	15
SA2	8.6	11

Shown are cumulative percentage of population targeted and number of target regions values associated with a target case percentage of 15%, for the SA1, SA2, and OAM targeting strategies.

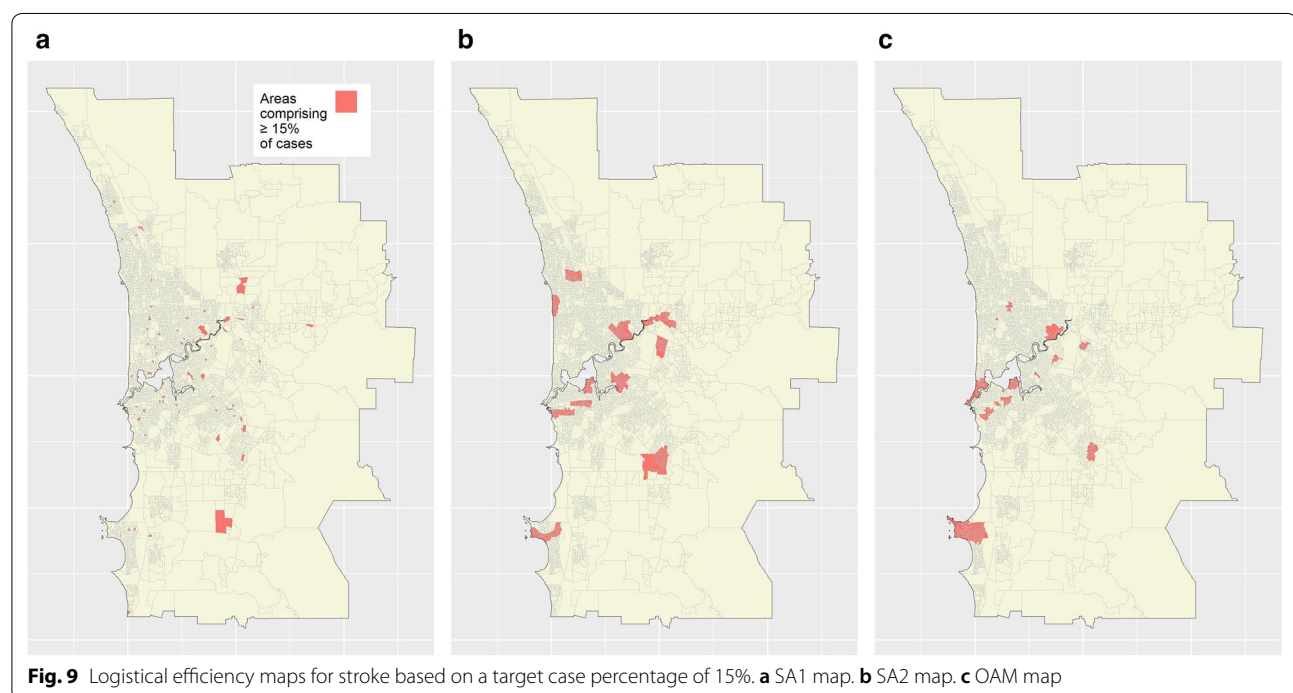
punctate. Therefore, it is likely that different configurations of the grid of 20×20 km units would have identified quite different hotspots. Future work should investigate potential differences in zonation dependence between datasets, as alluded to above.

Further, we have demonstrated how the impact of the MAUP on single-aggregation disease maps manifests in substantially reduced targeting efficiency due to prioritisation of logistical efficiency. This should disturb policy-makers, given the generally scarce nature of healthcare resources. OAM's second function as a smoothing technique represents a solution, with maps produced using OAM balancing the two aspects of efficiency. However, in demonstrating this, we have shown how OAM's output is strikingly similar to that of at least one existing smoothing technique applied to minimal-resolution data. Therefore, beyond the fact that such techniques have not previously considered efficiency as defined here, below

we outline the key features differentiating them from OAM.

Most notably, OAM facilitates derivation of smoothed, minimal-resolution estimates without directly modelling minimal-resolution data. This is achieved through models being fitted only to aggregate-level data, a generally less prohibitive task than would be their direct application to minimal-resolution data. Thus, more complex models than could otherwise be fitted can be applied within OAM; for example models including covariates or fully Bayesian models. To our knowledge, this feature is unique in the literature, with the exception of certain interpolation techniques that allow for derivation of estimates below the resolution at which data are available (e.g. see [49]). However, unlike OAM, such techniques are potentially undermined by the ecological fallacy, a problem related to, but distinct from, the MAUP.

Where both OAM and an alternative smoothing technique could be implemented, it might still be more advantageous to implement OAM, for these reasons: first, OAM uniquely accounts for edge effects. In contrast to existing smoothing techniques, which account for edge effects through implementing various corrections (e.g. see [45]), OAM does so through its effective smoothing kernels naturally adapting to the study area boundary and other borders. Second, OAM could be extended to model spatiotemporal data. This might involve simply examining its zonations at multiple time points. By contrast, not all existing smoothing techniques can model spatiotemporal data. Third, OAM stabilises population



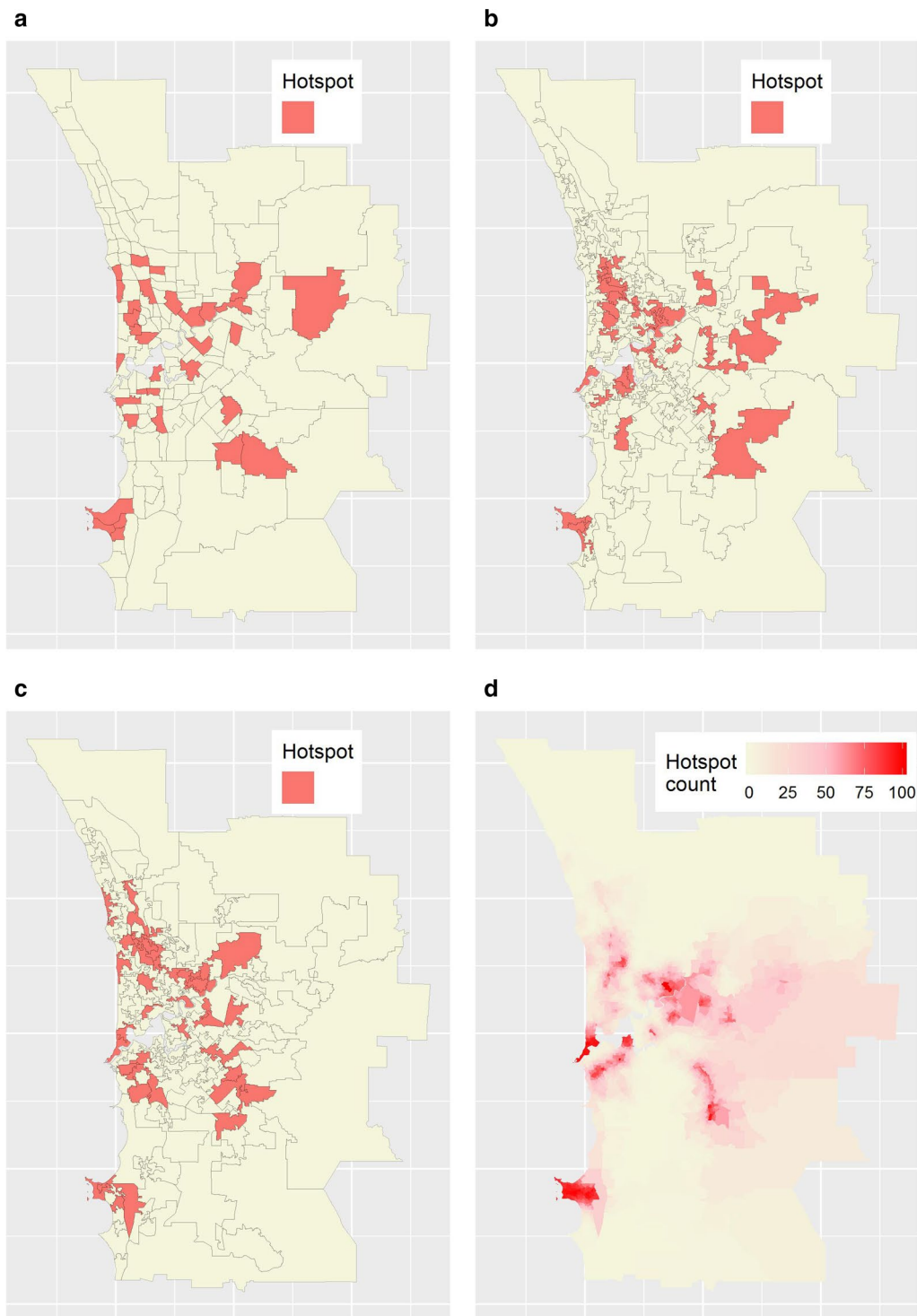


Fig. 10 Hotspot analysis results for stroke. **a** SA2 hotspots. **b, c** Hotspots based on two of OAM's zonations. **d** SA1-resolution hotspot counts

sizes between spatial units. Additional file 1: 'Additional results' (section 'Stabilisation of population sizes') demonstrates this, showing the distribution of population

sizes among SA1s, SA2s, and units within OAM's zonations in the stroke analysis. In this attribute, OAM is unique from most existing smoothing techniques, with

the exception of some kernel-smoothing approaches that implement adaptive bandwidths (e.g. see [50]). However, those techniques are often limited in their ability to additionally examine covariates or spatiotemporal data. Further, it has been noted that they might be computationally expensive even for moderately-sized datasets [51]. Finally, OAM protects patient privacy, since the process of combining information from multiple single-aggregation disease maps constitutes a geographical encryption key. To back-calculate data input into OAM would require: (i) extracting the values displayed in a given map; (ii) reversing their perturbation; and (iii) identifying a set of zonations and distribution of cases which, when modelled, exactly reproduce the unperturbed values. These steps constitute a complex deconvolution problem, almost certainly without a unique solution. Accordingly, some minimal-resolution values could be reported, since this would negligibly reduce the problem's complexity.

Importantly, we note that the impact of the MAUP described, quantified and overcome by OAM is that of its zonation aspect arising due to aggregation beyond the minimal resolution. OAM does not overcome the corresponding impact of the MAUP's scale aspect; rather, we have simply demonstrated that an increased scale of aggregation within OAM results in increased smoothing. This is unavoidable, and also characterises the output of other smoothing techniques. In addition, OAM does not assess the impact of the MAUP as manifesting due to aggregation by the minimal unit in the first place. That impact is also unavoidable when producing a disease map, since to do so the study area must be discretised in some way. Within OAM, this occurs when data are aggregated by the minimal units. By comparison, kernel-smoothing algorithms usually display information by a fine-resolution grid. As another example, spatial point process models often approximate a continuous spatial intensity surface at a fine, grid-based resolution, or in a more complex manner through applying basis functions across a triangulated mesh [42]. As a third possibility, discretisation might occur as a final step, i.e. post-modelling, purely for the purpose of producing a map.

It could be argued that the examination of multiple different zonations within OAM ignores the inherent importance or meaningfulness of pre-defined administrative areas. This view is correct in certain aspects; for example, funding to address disease is often distributed based on local government areas, health districts, or hospital catchments. Further, certain interventions are best targeted by administrative boundaries. For example, at the time of writing, some regions in Victoria, Australia had recently re-entered 'lockdown' to address a second wave of COVID-19 [15]. The impending lockdown was

communicated to residents through reference to post-code boundaries, and it is likely that any other approach would have caused confusion. However, this reality should not preclude the situation where interventions to address disease are targeted to regions potentially overlapping different administrative units. Since intervention resources are invariably scarce, as a general rule, disease should be mapped as it is distributed, and funding or other intervention measures adapted to this distribution to most effectively and efficiently address it, not the other way around.

The scale of aggregation within OAM parallels the choice of a smoothing bandwidth in some other smoothing techniques. There is a rich literature devoted to evaluating various data-based or theoretical techniques of bandwidth selection for such methods (e.g. see [52]). Future work might investigate application of these techniques within OAM. In the meantime, we suggest users choose an appropriate scale of aggregation guided by the characteristics of a planned intervention. For example, a relatively fine scale might be appropriate when planning a highly localised intervention, since this will result in delineation of increasingly punctate and discontinuous target regions. However, such an approach will potentially be limited by computational constraints and the relative rareness of the disease being examined (rare conditions will generally require a greater degree of aggregation).

There is potential to extend OAM to incorporate statistical measures of uncertainty (e.g. confidence intervals) for its map output. Since this functionality is not currently included, OAM cannot be used to statistically classify hotspots, such as is often desired in practice to characterise inequality between geographical regions. This was intentional, since our focus has been to demonstrate OAM's utility in delineating target regions characterised by their efficiency, and compare this efficiency to that of single-aggregation maps. While future work might extend OAM to enable classification of hotspots, it would not be sensible to quantify uncertainty in OAM's efficiency output, since, by definition, that output is directly associated with a given observed dataset.

While AZTool is a useful, free software, it is not an essential component of OAM; OAM's zonations could be created using any one of several available zonation tools, many of them automated. Such tools have become increasingly available in recent years, though they are not always free. Alternatively, custom computer code could be written. Future work might investigate the potential impact of implementing OAM using zonations created using different software.

Our presentation of OAM in the first instance as applied to a point location dataset generalises our

findings beyond the specific set of administrative units in the stroke application. Essentially, it is unimportant whether the minimal and aggregate-level units are polygons (e.g. administrative units) or grid cells. For the simulation, the minimal units were grid cells while the aggregate-level units were polygons. By contrast, for stroke, both the minimal and aggregate-level units were administrative units. As a third possibility, only aggregate-level information might be available, but for multiple zonations. In that case, a minimal-resolution value could still be derived using OAM, simply through implementing its fourth and fifth steps. Future work might investigate the exact way in which such a procedure might be implemented.

As a final point, we consider another disease mapping method commonly applied to both point location and areal disease data: spatial scans [53]. As an example, we consider the popular method SaTScan [54, 55]. Within SaTScan, scanning windows of different sizes that are usually circular are evaluated across a given geographical study area, with the aim of identifying one or more 'likely' clusters. Our interest is to compare such output to that of OAM in the context of efficiency. As such, we note that spatial scans differ from the smoothing techniques described above in that they cannot output minimal-resolution values; rather, values associated with each window are output. Thus, it is not clear how such output might be used to calculate targeting and logistical efficiency. Key to doing so would be effective management of the overlapping windows. Future work might usefully investigate different strategies to achieve this.

Conclusions

In conclusion, we have demonstrated how pronounced inefficiencies could result from reliance on single-aggregation disease maps to guide distribution of healthcare resources. Routine implementation of OAM can help prevent this. We hope that this paper will be a catalyst for increased awareness and acknowledgement of the MAUP in disease mapping and beyond. As a first step, studies examining areal data, both within and beyond the context of disease mapping, should routinely acknowledge the MAUP as a limiting feature whenever it is impossible to explicitly address its impact.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12942-020-00236-y>.

Additional file 1. Additional results.

Abbreviations

ABS: Australian Bureau of Statistics; HMDC: Hospital Morbidity Data Collection; MAUP: Modifiable areal unit problem; OAM: Overlay aggregation method; SA1-4: Statistical Areas Levels 1-4; WA: Western Australia; ZDP: Zonation-dependent positive; ZDN: Zonation-dependent negative.

Acknowledgements

The authors acknowledge the WA Data Linkage Branch, the WA Data Collections and their custodians, in particular the Hospital Morbidity Data Collection. We also acknowledge 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.

Authors' contributions

DW and MT conceived the study and formulated its design. MY and MRK contributed to the study conception and its design. MT performed all analyses, wrote the initial draft, and edited subsequent drafts. KM, BB, AV, DW, and BT provided formal analysis. All authors read and approved the final manuscript.

Funding

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

Availability of data and materials

The authors declare that data supporting the findings of this study are available within the paper. Population data for Perth are freely available online. Hospitalisation data for Perth are confidential and available by application to the Department of Health Western Australia, Data Linkage Branch.

Ethics approval and consent to participate

The study was approved by the Western Australia Department of Health Human Research Ethics Committee.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Medical School, University of Western Australia, Perth, Australia. ² Department of Mathematics and Statistics, University of Western Australia, Perth, Australia. ³ UWA School of Agriculture and Environment, University of Western Australia, Perth, Australia. ⁴ Department of Geography, University of Western Australia, Perth, Australia. ⁵ School of Population and Global Health, University of Western Australia, Perth, Australia.

Received: 15 April 2020 Accepted: 21 September 2020

Published online: 03 October 2020

References

1. Waller LA, Gotway CA. Spatial clustering of health events: regional count data. In: Shewhart WA, Wilks SS, editors. Applied spatial statistics for public health data. Hoboken: Wiley; 2004. p. 204.
2. World Health Organisation (WHO). Global Health Observatory (GHO) Map Gallery. 2019. https://www.who.int/gho/map_gallery/en/. Accessed 18 June 2019.
3. Kedziora DJ, Stuart RM, Pearson J, Latypov A, Dierst-Davies R, Duda M, Avaliani N, Wilson DP, Kerr CC. Optimal allocation of HIV resources among geographical regions. *BMC Public Health*. 2019;19(1):1–5.
4. Anderson SJ, Ghys PD, Ombam R, Hallett TB. HIV prevention where it is needed most: comparison of strategies for the geographical allocation of interventions. *J Int AIDS Soc*. 2017;20(4):e25020.
5. Anderson SJ, Cherutich P, Kilonzo N, Cremin I, Fecht D, Kimanga D, Harper M, Masha RL, Ngongo PB, Maina W, Dybul M. Maximising the effect of combination HIV prevention through prioritisation of

- the people and places in greatest need: a modelling study. *Lancet*. 2014;384(9939):249–56.
6. Meyer-Rath G, McGillen JB, Cuadros DF, Hallett TB, Bhatt S, Wabiri N, Tanser F, Rehle T. Targeting the right interventions to the right people and places: the role of geospatial analysis in HIV program planning. *AIDS*. 2018;32(8):957.
 7. Joint United Nations Programme on HIV/AIDS (UNAIDS). Location, location: connecting people faster to HIV services. 2013 https://www.unaids.org/sites/default/files/media_asset/LocationLocation_en.pdf. Accessed 18 June 2019.
 8. Lessler J, Moore SM, Luquero FJ, McKay HS, Grais R, Henkens M, Mengel M, Dunoyer J, M'bangombe M, Lee EC, Djingarey MH. Mapping the burden of cholera in sub-Saharan Africa and implications for control: an analysis of data across geographical scales. *Lancet*. 2018;391(10133):1908–15.
 9. Lee EC, Azman AS, Kaminsky J, Moore SM, McKay HS, Lessler J. The projected impact of geographic targeting of oral cholera vaccination in sub-Saharan Africa: a modeling study. *PLoS Med*. 2019;16(12):e1003003.
 10. Strategic Advisory Group of Experts (SAGE) on Immunization. SAGE Interim Recommendations on Vaccination against Ebola Virus Disease (EVD). 2019. https://www.who.int/immunization/interim_ebola_recommendations_feb_2019.pdf. Accessed 18 June 2019.
 11. Bousema T, Griffin JT, Sauerwein RW, Smith DL, Churcher TS, Takken W, Ghani A, Drakeley C, Gosling R. Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS Med*. 2012;9(1):e1001165.
 12. Crisanti A. and Cassone A. In: The Guardian. In one Italian town, we showed mass testing could eradicate the coronavirus. 2020. <https://www.theguardian.com/commentisfree/2020/mar/20/eradicated-coronavirus-mass-testing-covid-19-italy-vo>. Accessed 25 March 2020.
 13. Government of Western Australia Department of Health. New heat map highlights COVID-19 hotspots. 2020. <https://ww2.health.wa.gov.au/Media-releases/2020/New-heat-map-highlights-COVID19-hotspots>. Accessed 3 April 2020.
 14. Health Protection NSW. NSW COVID-19 case statistics—up to 8pm 26 July 2020. 2020. <https://www.health.nsw.gov.au/Infectious/covid-19/Pages/stats-nsw.aspx>. Accessed 27 July 2020.
 15. Victorian Government. COVID-19 suburban testing blitz: Hotspot intervention plan. 2020. https://www.dhhs.vic.gov.au/sites/default/files/documents/202006/2001628_COVID-19_COVID-19%2520suburban%2520testing%2520blitz.pdf. Accessed 28 June 2020.
 16. Kennedy S. The small number problem and the accuracy of spatial databases. In: Goodchild MF, Gopal S, editors. *The accuracy of spatial databases*. New York: Taylor & Francis; 1989. p. 134–140.
 17. Australian Bureau of Statistics. Australian Statistical Geography Standard (ASGS): volume 1—main structure and greater capital city statistical areas, July 2016; 2016. Report number: 1270.0.55.001. <https://www.abs.gov.au/ausstats/abs@.nsf/mf/1270.0.55.001>. Accessed 15 May 2019.
 18. Cramb SM, Mengersen KL, Baade PD. Atlas of cancer in Queensland: geographical variation in incidence and survival, 1998–2007. Brisbane: Viertel Centre for Research in Cancer Control, Cancer Council Queensland; 2011.
 19. Australian Institute of Health and Welfare (AIHW). The 2018 Australian atlas of healthcare variation. 2018. <https://www.safetyandquality.gov.au/our-work/healthcare-variation/atlas-2018>. Accessed 18 June 2019.
 20. Australian Institute of Health and Welfare (AIHW). AIHW data by geography. 2019. <https://www.aihw.gov.au/about-our-data/aihw-data-by-geography>. Accessed 18 June 2019.
 21. Centers for Chronic Disease Prevention (CDC). Division for heart disease and stroke prevention. Chronic disease GIS exchange map gallery. 2019. <https://www.cdc.gov/dhdsp/maps/gisx/mapgallery?Sort=%3A%3Aasc>. Accessed 18 June 2019.
 22. Public Health England. Public Health Dashboard: Increasing the transparency of public health. 2019. <https://fingertips.phe.org.uk/topic/public-health-dashboard/map-with-data>. Accessed 18 June 2019.
 23. Gehlke CE, Biehl K. Certain effects of grouping upon the size of the correlation coefficient in census tract material. *J Am Stat Assoc*. 1934;29(185A):169–70.
 24. Openshaw S, Taylor PJ. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In: Wrigley N, editor. *Statistical applications in the spatial sciences*, vol. 127. London: Pion; 1979. p. 127–144 (chapter 5).
 25. Openshaw S. *Concepts and Techniques in Modern Geography No.38: the modifiable areal unit problem*. Norwich: Geo Books; 1984.
 26. Manley D. Scale, aggregation, and the modifiable areal unit problem. In: Fischer MM, Nijkamp P, editors. *The handbook of regional science*. Berlin: Springer; 2014. p. 1157–1171 (Chapter 59).
 27. Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*. 2020;582:389–94.
 28. Drew DA, Nguyen LH, Steves CJ, Menni C, Freydin M, Varsavsky T, Sudre CH, Cardoso MJ, Ourselin S, Wolf J, Spector TD. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science*. 2020;368:1362.
 29. Reiner RC Jr, Hay SI, Wiens KE, Deshpande A, Baumann MM, Lindstedt PA, Blacker BF, Troeger CE, Earl L, Munro SB, Abate D. Mapping geographical inequalities in childhood diarrhoeal morbidity and mortality in low-income and middle-income countries, 2000–17: analysis for the Global Burden of Disease Study 2017. *Lancet*. 2020;395:1779.
 30. Xu P, Huang H, Dong N. The modifiable areal unit problem in traffic safety: basic issue, potential solutions and future research. *J Traff Transp Eng*. 2018;5(1):73–82.
 31. Goodchild MF. Scale in GIS: an overview. *Geomorphology*. 2011;130:5–9.
 32. Burden S, Steel D. Empirical zoning distributions for small area data. *Geogr Anal*. 2016;48(4):373–90.
 33. Tuson M, Yap M, Kok MR, Murray K, Turlach B, Whyatt D. Incorporating geography into a new generalized theoretical and statistical framework addressing the modifiable areal unit problem. *Int J Health Geogr*. 2019;18(1):6.
 34. Fassbender K, Grotta JC, Walter S, Grunwald IQ, Ragoschke-Schumm A, Saver JL. Mobile stroke units for prehospital thrombolysis, triage, and beyond: benefits and challenges. *Lancet Neurol*. 2017;16(3):227–37.
 35. Martin D. Extending the automated zoning procedure to reconcile incompatible zoning systems. *Int J Geogr Inf Sci*. 2003;17(2):181–96.
 36. Cockings S, Harfoot A, Martin D, Hornby D. Maintaining existing zoning systems using automated zone-design techniques: methods for creating the 2011 Census output geographies for England and Wales. *Environ Plan A*. 2011;43(10):2399–418.
 37. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math*. 1991;43(1):1–20.
 38. Konstantinoudis G, Schuhmacher D, Rue H, Spycher BD. Discrete versus continuous domain models for disease mapping. *Spat Spatiotemporal Epidemiol*. 2020;32:100319.
 39. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *J R Stat Soc Ser B*. 2009; 71(2): 319–392. www.r-inla.org.
 40. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. <https://www.R-project.org/>.
 41. Moraga P. Small area disease risk estimation and visualization using R. *R J*. 2018;10(1):495–506.
 42. Holman CD, Bass AJ, Rouse IL, Hobbs MS. Population-based linkage of health records in Western Australia: development of a health services research linked database. *Aust N Z J Public Health*. 1999;23(5):453–9.
 43. Kokotailo RA, Hill MD. Coding of stroke and stroke risk factors using international classification of diseases, revisions 9 and 10. *Stroke*. 2005;36(8):1776–811.
 44. National Centre for Classification. The international statistical classification of diseases and related health problems, tenth revision, Australian modification. 7th ed. Lidcombe: National Centre for Classification; 2010.
 45. Davies TM, Marshall JC, Hazelton ML. Tutorial on kernel estimation of continuous spatial and spatiotemporal relative risk. *Stat Med*. 2018;37(7):1191–221.
 46. Tomas J. Aragon. *epitools: epidemiology tools*. R package version 0.5-10.1.2020. <https://CRAN.R-project.org/package=epitools>.
 47. Calderon VJ, Kasturiarachi BM, Lin E, Bansal V, Zaidat OO. Review of the mobile stroke unit experience worldwide. *Interv Neurol*. 2018;7(6):347–58.
 48. Golding N, Burstein R, Longbottom J, Browne AJ, Fullman N, Osgood-Zimmerman A, Earl L, Bhatt S, Cameron E, Casey DC, Dwyer-Lindgren L. Mapping under-5 and neonatal mortality in Africa, 2000–15: a baseline analysis for the Sustainable Development Goals. *Lancet*. 2017;390(10108):2171–82.
 49. Goovaerts P. Geostatistical analysis of disease data: accounting for spatial support and population density in the isopleth mapping of cancer

- mortality risk using area-to-point Poisson kriging. *Int J Health Geogr*. 2006;5(1):52.
50. Carlos HA, Shi X, Sargent J, Tanski S, Berke EM. Density estimation and adaptive bandwidths: a primer for public health practitioners. *Int J Health Geogr*. 2010;9(1):1–8.
 51. Davies TM. sparr: spatial and spatiotemporal relative risk. R package version 2.2-13-16.9.2018. <https://cran.r-project.org/web/packages/sparr>.
 52. Lemke D, Mattauch V, Heidinger O, Pebesma E, Hense HW. Comparing adaptive and fixed bandwidth-based kernel density estimates in spatial cancer epidemiology. *International journal of health geographics*. 2015;14(1):15.
 53. Waller L. Point process models and methods in spatial epidemiology in "Handbook of Spatial Statistics". Gelfand AE, Diggle PJ, Fuentes M, & Guttorp P, eds. New York: Taylor & Francis 2010.
 54. Kulldorff M. A spatial scan statistic. *Commun Stat Theor Methods*. 1997;26(6):1481–96.
 55. Kulldorff M, International Management Services, Inc. SaTScan v. 7.0: Software for the spatial and space-time scan statistics. Bethesda: National Cancer Institute; 2006.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

