

Practice of Epidemiology

Utility of Spatial Point-Pattern Analysis Using Residential and Workplace Geospatial Information to Localize Potential Outbreak Sources

Jonathan L. Chua, Lee Ching Ng, Vernon J. Lee, Marcus E. H. Ong, Er Luen Lim, Hoon Chin Steven Lim, Chee Kheong Ooi, Arif Tyebally, Eillyne Seow, and Mark I-Cheng Chen*

* Correspondence to Dr. Mark I-Cheng Chen, National Centre for Infectious Disease, 16 Jalan Tan Tock Seng, Singapore 308442, Republic of Singapore (e-mail: mark_ic_chen@ncid.sg).

Initially submitted May 8, 2018; accepted for publication December 21, 2018.

Identifying the source of an outbreak facilitates its control. Spatial methods are not optimally used in outbreak investigation, due to a mix of the complexities involved (e.g., methods requiring additional parameter selection), imperfect performance, and lack of confidence in existing options. We simulated 30 mock outbreaks and compared 5 simple methods that do not require parameter selection but could select between mock cases' residential and workplace addresses to localize the source. Each category of site had a unique spatial distribution; residential and workplace address were visually and statistically clustered around the residential neighborhood and city center sites respectively, suggesting that the value of workplace addresses is tied to the location where an outbreak might originate. A modification to centographic statistics that we propose—the center of minimum geometric distance with address selection—was able to localize the mock outbreak source to within a 500 m radius in almost all instances when using workplace in combination with residential addresses. In the sensitivity analysis, when given sufficient workplace data, the method performed well in various scenarios with only 10 cases. It was also successful when applied to past outbreaks, except for a multisite outbreak from a common food supplier.

algorithms; food-borne diseases; geographic information systems; infectious disease outbreaks; source localization; source of outbreak; spatial analysis; workplace

Abbreviation: CI, confidence interval.

The nature of transmission for some infectious diseases influences their spatial epidemiology, but several challenges exist in optimally using spatial data to investigate infectious disease outbreaks. First, information on the spatial relationship between the source of infection and infected cases is often incomplete (1). Other than residential addresses (often routinely captured in administrative health-care data), additional spatial information is less readily available. In outbreaks arising from a food outlet where customers reside nearby, residential address might be most relevant. However, more meals are bought outside the home (2–5), and in food establishments serving concentrations of working populations, workplace addresses might be more relevant. Workplace addresses have indeed been obtained for and shown to be useful in studies of pollution (6), dengue (7)

and Zika (8) epidemiology as well as, more broadly, some syndromic surveillance systems (9).

A second challenge is how to simultaneously integrate multiple sources of address information. While visualization plots are often used when communicating findings from outbreak investigations (10), meaningful combination of 2 or more sources of spatial information to localize the origin of an outbreak has not been addressed. Moreover, in complex spatial techniques, unknown parameter values often need to be estimated using additional data sources or to be subjectively selected by the user. For instance, with Kulldorff's spatial scan statistic underlying population density as well as radius and shape of the scanning window (11) are key parameter selection choices that can significantly influence results (12). Without data from a range

of real outbreaks with both residential and workplace addresses as well as the outbreak source, it is challenging to establish the validity of any of these models and parameter choices.

We aimed to address some of the above challenges. First, we have proposed a data collection methodology for generating mock outbreaks to validate spatial epidemiologic methods. We then used survey data to study factors associated with the distance from the source to residential and workplace addresses, and to validate a simple algorithm we propose that meaningfully combines residential and workplace (and potentially other) addresses to help identify the source of an outbreak by localizing it to a particular vicinity, yet does not require additional parameter selection for implementation. Finally, we validated the algorithm on a group of real food-borne and vector-borne outbreaks where a presumptive geospatial source was identified.

METHODS

Study design for mock outbreaks

Singapore is a dense urban city where consumption of prepared food outside the home is common (13). We designed a study to generate mock outbreaks with data on the geographical dispersion of residential and workplace addresses of potentially affected individuals, focusing on how a food-borne disease outbreak might present should a food establishment in a given location become the source.

A street-intercept survey (14, 15), where participants completed a survey within sight of but (to avoid conflicts with commercial interests) outside of a major entry and exit point of shopping malls, was performed. We used shopping malls as a proxy for food establishments because food establishments are commonly located within and/or in the immediate vicinity of malls.

To ensure adequate representation across Singapore, we surveyed 30 sites, 10 each from 3 broad categories of target locations. We randomly selected 10 train stations outside the central district and then identified 2 categories of target locations for each station. First, for residential hub sites, we randomly selected a mall within 300 m of the station and then identified a corresponding survey site. Secondly, we randomly selected a residential neighborhood served by those

selected stations (at least 500 m from the nearest train station); not all had a mall in their immediate vicinity, but all had clusters of food establishments, which we then used as the target location. Third, for city center sites, we randomly selected 5 train stations in the central district and then randomly selected 2 malls within a 300-m radius of each station for a combined total of 10 sites.

Survey instrument, conduct, and sample size

The National University of Singapore Institutional Review Board approved the study (NUS-13-303). At the chosen locations, trained interviewers approached every tenth person crossing their path. Those giving verbal consent filled out a short questionnaire (approximately 5 minutes long; see Web Appendix 1, available at <https://academic.oup.com/aje>).

We estimated that a sample size of 100 would be required for sufficient numbers in a subsequent sensitivity analysis for mock outbreaks up to a maximum size of 50 cases, assuming that approximately 60% consumed prepared food within the immediate vicinity of recruitment sites. Participants from city center sites would more likely have a workplace address, and we therefore doubled the sample sizes to 200 participants for residential neighborhood and residential hub sites to ensure adequate workplace addresses.

Questionnaires were administered around mealtimes (lunch: 12 PM to 4 PM; dinner: 4 PM to 8 PM) on both weekdays and weekends (with a quarter each for every day of week and meal-time combination).

Statistical analysis and localization algorithms

Residential distance and workplace distance were defined respectively as the Euclidean distance between a participants' residential or workplace address to the survey site. Hierarchical linear modeling (HLM) with a random intercept was used to assess whether residential distance and workplace distance were associated with the category of site, age, sex, time of day, and day of week, as well as whether participants consumed prepared (ready-to-eat) food purchased from food establishments

Table 1. Description of Centographic Methods and Their Modifications to Estimate the Source of an Outbreak, Used in an Analysis of Data from Singapore, 2014–2015

Number	Method Name	Description of Method Calculation
1	Median center	Median <i>x</i> and median <i>y</i> coordinate values
2a	Center of minimum arithmetic distance	Selects the grid ^a point with the minimum sum of arithmetic distance to all coordinate points
2b	Center of minimum arithmetic distance with address selection	When both address types are available, the address type that is closer to each grid ^a point is selected before calculating the sum of distance for each grid point; the grid point with the minimum sum of arithmetic distance is selected
3a	Center of minimum geometric distance ^b	Selects the grid ^a point with the minimum sum of geometric distance to all coordinate points
3b	Center of minimum geometric distance with address selection ^b	When both address types are available, the address type that is closer to each grid ^a point is selected before calculating the sum of distance for each grid point; the grid point with the minimum sum of geometric distance is selected

^a Grid refers to a standard grid (50 m × 50 m) of Singapore created for all analyses using methods 2a–3b.

^b Geometric distance was calculated using the logarithm to base 10, and minimum distance was set to 10 m to prevent negative or undefined values.

at that site, with the site identifier as the random effect. *P* values less than 0.05 were considered statistically significant.

We next tested how various algorithms for centrogographic statistics could locate the geographical source of an outbreak using residential and workplace addresses. The objective was to use the algorithms to identify a putative center in close proximity to the actual source to allow an efficient search for a potential source of contamination. To localize the putative center, we relied on centrogographic methods and our own modification of these methods, described in Table 1.

To determine the performance of each method, the Euclidean distance between the survey site and putative center was calculated. Residential and workplace addresses do not change frequently, but people frequently eat at different locations yet would not travel too far to consume food. Hence, we used 500 m (surrogate for a 5-minute walk) as the benchmark for performance.

We performed sensitivity analyses to assess performance in outbreak scenarios involving fewer cases. At each survey site, we randomly sampled individuals without replacement. For each sample size simulated (50, 25, and 10), we performed 1,000 iterations and then assessed the distribution of distances from the putative centers to the survey sites.

Analysis of real outbreaks

The algorithms were further validated using data from past outbreaks where we had some certainty about the geospatial source of the outbreak. Because workplace addresses were unavailable for the 2 food-borne disease outbreaks (one has been published (16)), we also tested the algorithms on data from 6 vector-borne disease outbreaks (5 chikungunya virus and 1 Zika virus) where workplace addresses are routinely collected. Although vectors are mobile (unlike food establishments), the range of *Aedes aegypti* and *Aedes albopictus* (vectors for dengue, Zika, and chikungunya) are fairly limited (17). Methods 2 and 3 were tested for their ability to estimate a putative center from outbreak cases, and this was then referenced against the presumptive outbreak source. For the food-borne and vector-

borne disease outbreaks, respectively, this was the contaminated food establishment and the center of all points where vectors positive for the outbreak virus were found. Table 2 (with Web Figures 1–9) provides additional information on these outbreaks.

Finally, in an outbreak, some cases whose clinical presentation matches the outbreak case definition are actually unrelated to the outbreak. We tested the algorithms' sensitivity to such unrelated cases. These might be a small fraction of a total outbreak data set, given that investigations typically case-find around a suspected geographical source. However, unrelated cases might have a disproportionate influence early in an outbreak when the number of outbreak-associated cases is small. We generated random samples of 3 real outbreak cases, and we then mixed these with up to 3 individuals (with their associated residential and workplace addresses) randomly sampled from our entire mock outbreak data. The distance between the estimated putative center and the presumptive source was then estimated using methods 2 and 3 and compared graphically.

R, version 3.3.2 (R Foundation for Statistical Computing, Vienna, Austria), was used for all statistical analysis (18); the code is provided in Web Appendix 2.

RESULTS

A total of 5,012 participants completed the survey (according to site: 2,008 from residential neighborhoods; 2,001 from residential hubs; and 1,003 from city center sites). Median age was 26 (range, 16–85) years; 47% were female; 46% were employed; and 61% consumed prepared food from that site (Table 3). Sex distribution was similar across the categories of sites, but age distributions differed (e.g., those older than 20 years and up to 35 years comprised one-third from residential neighborhoods but almost one-half from city center sites).

Spatial distribution of addresses

Web Figure 10 shows 1 representative each for residential neighborhood, residential hub, and city center sites. For residential

Table 2. Description and Performance of Methods on Food-Borne and Vector-Borne Outbreaks in Singapore, 2013–2018

Serial No.	Outbreak	Site Category ^a	All Cases ^b								First 14 Days ^b							
			No. of Cases ^c		β		N ^c	N _w ^d	N _{ws} ^e	β								
			N _w ^d	N _{ws} ^e	Method 2a	Method 2b				Method 3a	Method 3b	Method 2a	Method 2b	Method 3a	Method 3b			
1	A: GE 1 ^f	C	121		0.111													
2	B: GE 2 ^f	A	15		1.524													
3	C: ChikV 1	B	13	9	5	0.436	0.386	0.436	0.436	12	8	4	0.386	0.386	0.436	0.436		
4	D: ChikV 2	C	52	45	30	0.146	0.215	0.336	0.090	8	7	3	0.405	0.381	0.363	0.363		
5	E: ChikV 3	C	12	7	3	0.151	0.151	0.151	0.151	4	0	0	0.197	0.197	0.256	0.256		
6	F: ChikV 4	C	42	40	16	0.020	0.064	0.064	0.064	7	6	2	0.064	0.064	0.064	0.064		
7	G: ChikV 5	B	14	7	6	0.070	0.070	0.070	0.070	11	6	5	0.070	0.070	0.070	0.094		
8	H: ZikV	B	43	35	35	0.412	0.412	0.412	0.412	9	9	9	0.412	0.412	0.412	0.412		

Abbreviations: GE, gastroenteritis; ChikV, chikungunya virus; ZikV, Zika virus.

^a Site categories were: A, city center; B, residential hub; C, residential neighborhood.

^b The unit measurement for the β estimates is kilometers.

^c Number of cases included (all cases capture residential address).

^d Number of valid workplace address from included cases.

^e Number of workplace addresses that were selected by the address selection algorithms (2b and 3b) to localize the outbreak source.

^f No temporal or workplace information was available.

Table 3. Demographic Characteristics of Participants ($n = 5,012$) Surveyed in Singapore, 2014–2015

Characteristic	Overall		Residential Neighborhood		Residential Hub		City Center	
	No. of Participants	%	No. of Participants	%	No. of Participants	%	No. of Participants	%
Age								
16–20	1,342	27	484	24	584	29	274	27
21–35	1,893	38	635	32	729	36	529	53
36–50	880	18	394	20	382	19	104	10
51–65	711	14	373	19	262	13	76	8
66–85	186	4	122	6	44	2	20	2
Sex								
Male	2,656	53	1,094	55	1,099	55	463	47
Female	2,315	47	899	45	887	45	529	53
Employed								
Yes	2,304	46	857	43	875	44	572	57
No	2,700	54	1,150	57	1,120	56	430	43
Food exposure ^a								
Yes	3,070	61	1,122	56	1,256	63	692	69
No	1,942	39	886	44	745	37	311	31

^a Individuals who consumed prepared food from a food establishment at the survey site were considered to have food exposure.

Table 4. Hierarchical Linear Regression Analysis, Using Residential Address and Workplace Address, of Simulated Outbreaks in Singapore, 2014–2015

Variable	Residential Address (Model 1)			Workplace Address (Model 2)		
	β^a	95% CI	P Value	β^a	95% CI	P Value
Site						
Residential neighborhood	0	Referent		0	Referent	
Residential hub	1.285	−0.568, 3.138	0.167	0.643	−0.673, 1.960	0.326
City center	7.287	5.422, 9.152	<0.001	−2.856	−4.219, −1.493	<0.001
Sex						
Male	0	Referent		0	Referent	
Female	−0.063	−0.301, 0.174	0.602	−1.237	−1.765, −0.710	<0.001
Age category, years						
16–20	0.399	0.024, 0.775	0.037	−0.382	−1.423, 0.660	0.474
21–35	0.446	0.097, 0.794	0.012	0.304	−0.327, 0.934	0.347
36–50	0	Referent		0	Referent	
51–65	−0.601	−1.017, −0.185	0.005	−1.211	−2.231, −0.190	0.021
66–95	−0.842	−1.523, −0.161	0.016	−2.609	−5.515, 0.297	0.080
Food exposure ^b						
No	0	Referent		0	Referent	
Yes	−0.536	−0.782, −0.290	<0.001	−0.824	−1.382, −0.267	0.004
Time						
Weekday lunch	0	Referent		0	Referent	
Weekday dinner	0.146	−0.183, 0.474	0.386	1.097	0.341, 1.853	0.005
Weekend lunch	−0.518	−0.847, −0.190	0.002	1.743	1.005, 2.481	<0.001
Weekend dinner	−0.226	−0.555, 0.103	0.179	1.568	0.823, 2.312	<0.001

Abbreviation: CI, confidence interval.

^a The unit measurement for the β estimates is kilometers.

^b Individuals who consumed prepared food from a food establishment at the survey site were considered to have food exposure.

neighborhoods, residential addresses were spatially clustered around the survey site, while workplace addresses were dispersed. The inverse was observed for city center sites. Residential hubs had patterns between city center and residential neighborhood sites, with clustering of both residential and workplace addresses around the site.

Hierarchical linear model regression for residential (model 1) and workplace (model 2) addresses confirmed the visual impressions of spatial distribution patterns (Table 4). Relative to residential neighborhoods, workplace addresses were about 3 km closer to city center sites ($\beta = -2.856$, 95% confidence interval

(CI): $-4.219, -1.493$) while residential addresses were about 7 km further away ($\beta = 7.287$, 95% CI: 5.422, 9.152). Compared with residential neighborhoods, participants from residential hubs lived and worked further from the site (but not significantly so). Having consumed prepared food from the site (versus those who did not) was significantly associated with living or working closer to the site by 500 m (model 1: $\beta = -0.536$, 95% CI: $-0.782, -0.290$) and 800 m (model 2: $\beta = -0.824$, 95% CI: $-1.382, -0.267$), respectively. Participants also visited sites on average 500 m (model 1: $\beta = -0.518$, 95% CI: $-0.847, -0.190$) and 200 m (model 1: $\beta = -0.226$, 95% CI: $-0.555, 0.103$) closer to their home during lunch and dinner on the weekends, respectively, and on average 2 km further away from their workplace over the weekend lunch (model 2: $\beta = 1.743$, 95% CI: 0.823, 2.312) than during weekday lunch. Older participants, on average, lived and worked closer to sites. Those aged 50–65 years visited sites approximately 600 m closer to where they lived (model 1: $\beta = -0.601$, 95% CI: $-1.017, -0.185$) and approximately 1 km closer to where they worked (model 2: $\beta = -1.211$, 95% CI: $-2.231, -0.190$). Similarly, those older than 65 years lived 800 m (model 1: $\beta = -0.842$, 95% CI: $-1.523, -0.161$) closer and worked 3 km (model 2: $\beta = -2.609$, 95% CI: $-5.515, 0.297$) closer to sites they visited.

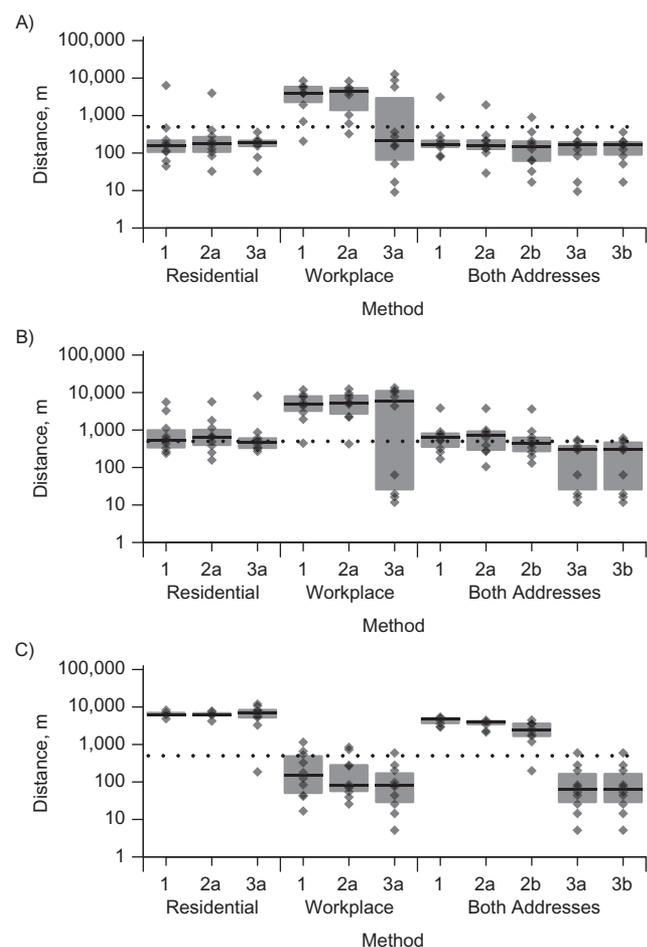


Figure 1. Comparison of 5 methods to localize the source of simulated outbreaks, Singapore, 2014–2015. A) Residential neighborhood ($n = 1,122$); B) residential hub ($n = 1,256$); C) city center ($n = 692$). The analysis was repeated using different spatial information (i.e., residential address, workplace address, or both addresses). The selection algorithm was applied only when both address types were used. The distance measured is the Euclidean distance between the estimated point generated from the various methods and the site of recruitment and therefore proxy source for the outbreak, with a lower value being a more desirable result. Each site's result is indicated by the black diamonds, and the box plot summarizes the result for each method, with the midline on each bar representing the median and the upper and lower extent of the bar depicting the interquartile range for the site. The black dotted horizontal line indicates our chosen benchmark of 500 m, with black diamonds below this line hence representing sites that are localized to a sufficient extent by each method.

Localizing the source of mock outbreaks using combinations of address types

With data from surveyed individuals, the mock outbreak source for residential neighborhoods was localized to within a 500 m radius of the putative center with just residential addresses; method 3a performed best (Figure 1A). Workplace address alone yielded poor results, while combining both addresses performed similarly to using only residential addresses (Figure 1A).

In contrast, using residential addresses alone yielded putative centers relatively far from the city center sites (Figure 1C). However, workplace address localized the majority of sites to within 500 m regardless of the method used (Figure 1C). With combined residential and workplace addresses, methods 3a and 3b had good performance comparable to using only workplace addresses (Figure 1C).

For residential hubs, residential addresses generally performed better than workplace addresses, but only about half the sites were localized to within 500 m, irrespective of the method (Figure 1B). Method 3a localized some sites using just workplace addresses, but the median distance from the putative center to the site was still 6 km (Figure 1B). The best results were from combining residential and workplace addresses using methods 3a and 3b, with 9 of 10 sites localized to within 500 m (Figure 1B).

Impact of the number of mock outbreak cases

Figure 2 presents how method 3b performed when sampling a limited number of individuals for each site. With 50 randomly sampled individuals, residential neighborhood sites were localized to within 500 m in almost all simulations (Figure 2A). With 25 individuals, the results were similar with only a slight decrease in performance at site 9, where approximately 30% of

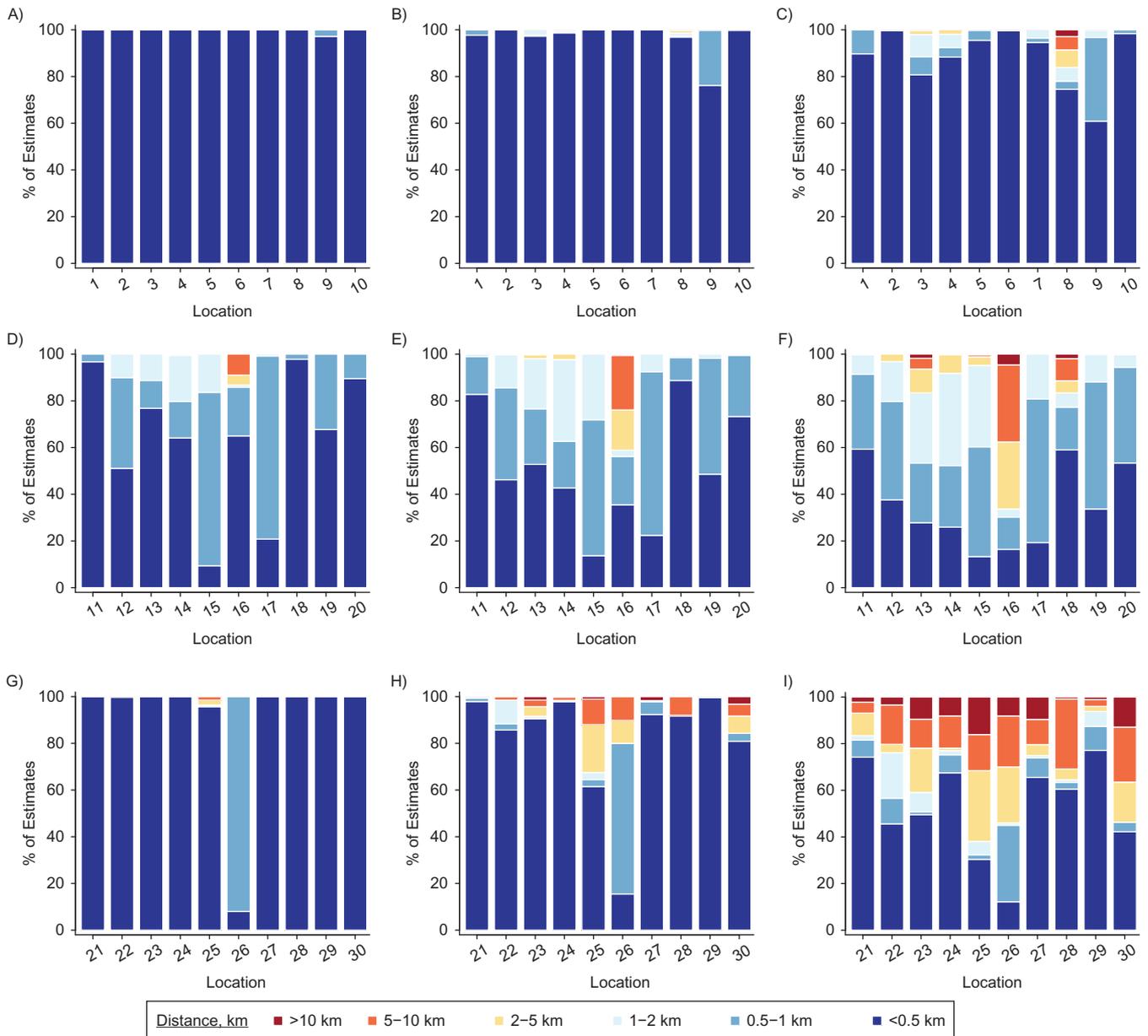


Figure 2. Sensitivity analysis of center of minimum geometric distance with address selection (method 3b) for localizing the source of simulated outbreaks, Singapore, 2014–2015. Rows distinguish between category of site (residential neighborhood (A, B, C); residential hub (D, E, F); city center (G, H, I)) and columns distinguish between number of samples in each iteration (50 samples (A, D, G); 25 samples (B, E, H); 10 samples (C, F, I)). A total of 1,000 iterations were performed for each site numbered from 1 to 30 and repeated with different sample sizes. Each color indicates the proportion of iterations for each site and sample size that was within a range of distance values.

simulations gave a putative center between 500 m and 1 km away (Figure 2B). With 10 individuals, the algorithm still performed well across most sites, except that approximately 15% of the simulations identified a putative center 2 km or more away for site 8 (Figure 2C).

While the algorithm performed worse for residential hubs (Figure 2D–F), with poor ability to localize site 16 in particular, we were able to localize 9 of the 10 sites to within 1 km in >50% of the simulations, even with just 10 individuals.

With 50 individuals at city center sites, we had reasonable results, except for site 26 (>90% of simulations had putative centers ≥ 500 m away, Figure 2G). However, with 25 observations, 10%–40% of simulations had poor localization (>2 km) across several sites. Localization to <500 m was achieved for only approximately 50% of scenarios using data from 10 individuals (Figure 2I). We explored the reasons for the poorer performance by repeating the analysis on all sites for 10 observations while restricting the sample to individuals who gave a

workplace address (Figure 3). Results for city center sites improved noticeably (vs. Figure 2I), with performance better than for residential hubs and closer to that for residential neighborhoods. Methods 3a, 2b, and 2a were inferior to method 3b across most sites at sample sizes of 25 and 10 (Web Figures - 11–13 respectively).

Validating methods 2 and 3 against real outbreaks

We applied the algorithms on past outbreaks to determine the sources of interest—implicated food establishments and locations where chikungunya virus- or Zika virus-positive mosquitoes were trapped—during outbreaks of the respective infections (19). All methods could localize most outbreaks to under 500 m, the exception being 1 food-borne outbreak (Table 2). On limiting the analysis to cases with symptoms in the first 14 days of the outbreak, the accuracy of the methods deteriorated slightly, but estimates remained within 500 m from known sources of interest. Both methods 2 and 3 performed well, with neither clearly superior. A sizeable number of workplace addresses were used by the address selection algorithms 2b and 3b, particularly for outbreaks 4 and 8.

Figure 4 shows the boxplot from 1,000 simulations of 3 cases randomly selected from each outbreak and the effect of unrelated cases on each method. For the food-borne outbreaks (outbreaks 1 and 2), method 3a performed better than method 2a. Moreover, method 3a maintained a similar result even with unrelated cases, except when 3 unrelated cases were added to outbreak 2. For the vector-borne outbreaks, method 2 was marginally better than method 3 in some outbreaks but was prone to poorer performance with unrelated cases, while method 3 returned consistent results even with 3 unrelated cases. Algorithms with address selection performed marginally better than those without (e.g., method 2b performed better than 2a in outbreaks 3, 4, 5 and 7, and method 3b performed better than 3a in outbreak 4).

DISCUSSION

We have described a mock outbreak data-collection method and have used it to demonstrate the relative value of residential and workplace addresses to localize the source of exposure. Then we proposed modified centroid methods that are simple, do not require parameterization, and can combine residential and workplace addresses. The resultant analyses improved on existing methods and highlighted the added value from workplace addresses in localizing outbreaks. This was replicated using data from real outbreaks and corroborates both the utility of our approach for collecting mock outbreak data and the localization algorithms presented. These findings have important implications for geospatial approaches to outbreak investigations and interpretation of surveillance data.

The workplace constitutes a large part of day-time exposure to infectious diseases and often is where people have the greatest density of contact with others. Consequently, it can be a key epidemiologic link between outbreak cases. Our findings provide evidence for this and help justify access by health authorities to sources of such information for surveillance and outbreak investigation.

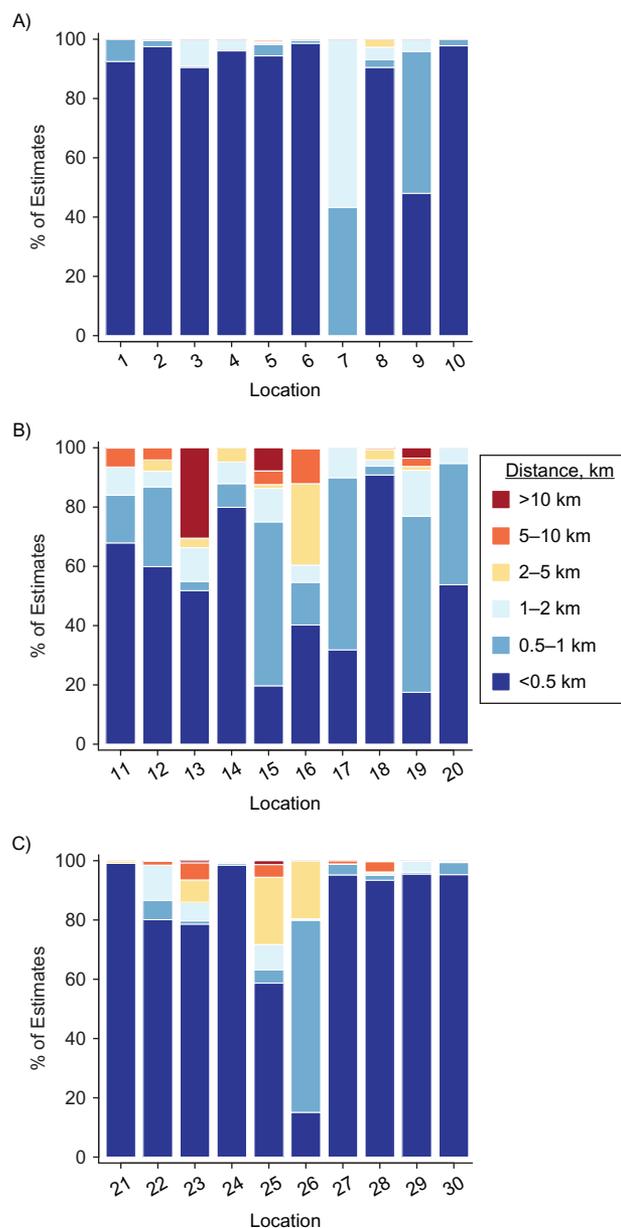


Figure 3. Sensitivity analysis of center of minimum geometric distance with address selection (method 3b) for localizing the source of simulated outbreaks, using only individuals who had a workplace address, Singapore, 2014–2015. The panels distinguish between categories of site: residential neighborhood (A); residential hub (B); city center (C). The analysis is limited to iterations with a sample size of 10. Each color indicates the proportion of iterations for each site that was within a range of distance values.

Next we introduced simple but, to our knowledge, novel modifications to traditional methods for estimation of centroids, and we showed that these modifications improved performance. Our results suggest that our optimal method was robust across a variety of scenarios. In addition, our modified methods were able to handle cases unrelated to an outbreak. Although we had only a limited number of real outbreaks for validation, they were sufficiently varied in their locations, covering all 3

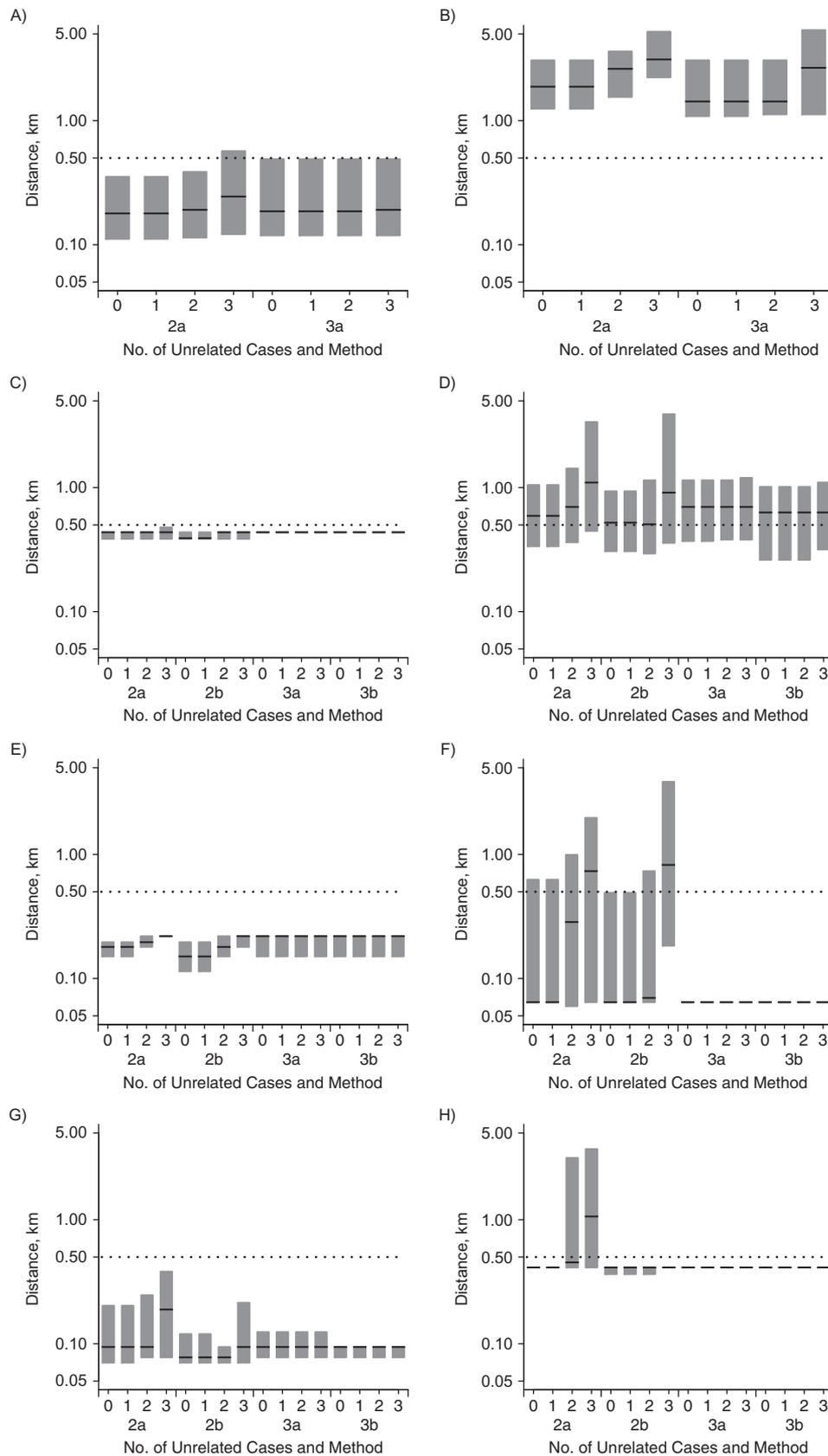


Figure 4. Sensitivity analysis of methods 2 and 3 on 8 real outbreaks, Singapore, 2013–2018. The outbreaks are also described in Table 2, according to serial number: 1 (A), 2 (B), 3 (C), 4 (D), 5 (E), 6 (F), 7 (G), and 8 (H). For each outbreak, 1,000 iterations were performed for each method and combination of number of unrelated cases. The black line is the median distance value between the estimate and the actual sources of interest. The interquartile range is given by the box. Our chosen benchmark of 500 m is indicated by the black horizontal dotted line.

categories of sites surveyed in the mock outbreaks (Table 2). In doing so, we also corroborated our approach to collecting the mock outbreak data for single point source outbreaks, which partially mitigates concerns about limitations such as our choice of preselected sites, response bias, and a lack of a true representation of the actual number of meals distributed across space and time (i.e., day of week and different mealtimes). We believe the approach is simple to execute and can be attempted to provide data for further testing and development of geospatial methods in settings similar to ours, where there are reasonably well-defined areas under surveillance, within which mock outbreak sites can be selected.

In terms of application, we foresee such outbreak localization algorithms used in several scenarios, in some of which this method might be particularly useful. With infections uncommon outside of outbreaks (e.g., sporadic outbreaks of chikungunya and Zika virus infections analyzed here), or where genetic sequencing links cases together, these algorithms could be used at the initial phase of an outbreak to define a case associated with that outbreak, help with recall when interviewing cases, and narrow the search area for potential mosquito breeding. They could also corroborate other information sources (e.g., reports about a specific food establishment causing food-borne disease). It might be of particular value for infections with long incubation periods where food history becomes complicated by recall bias, where this method could help focus on specific geographical clusters of food establishments, and triangulated with other findings to suggest establishments that should be targeted for additional investigation. The results should always be weighed against other findings before making a conclusion and taking action.

There are some limitations to our methods. One relates to the assumption that a singular outbreak arises from a single source at the site of exposure. Accordingly, our method is unable to localize an upstream outbreak source or handle multiple concurrent outbreaks within an area under surveillance. In the outbreak labelled GE2, our method could neither adequately localize at least 1 of 2 preschools that catered food nor the food establishment that provided the food, because there was no discernable spatial relationship between the upstream source and cases. However, had workplace information been collected, it might have been possible to localize the 2 preschools using clustering techniques to first differentiate the 2 locations where individuals were exposed. Further work is thus needed to assess whether spatial clustering techniques, such as those used with syndromic surveillance (20), can synergize well with the algorithms we have presented.

The method also underperforms in locations frequented for reasons unrelated to residence or work. For instance, results for site 16 were poor because it served as a sports hub (the closest residential building being approximately 400 m away), which participants likely frequented mainly for recreational activities. In the case of site 26, this was a transport node serving a local university, and results might have been affected because our survey did not capture student status and where they studied.

While we believe the algorithms presented should be generalizable to other densely populated metropolitan cities, additional validation is required in regions with lower population densities (e.g., rural areas), varying transport systems, and different work (and food) cultures. The algorithm is also dependent on the

availability of other address types. Notably, other government databases (e.g., for tax filing) might contain workplace addresses, but there are privacy concerns in accessing such information. Alternatives include collecting such information at point of notification by health-care providers, or soliciting it directly from cases (e.g., a participatory surveillance approach with patients encouraged to submit such information). Although more difficult to obtain than residential addresses, workplace addresses have been used in spatial studies of vector-borne transmission patterns (7) and in dengue and Zika virus outbreaks in Singapore (8, 21). This suggests that we can obtain and use such information if the justification for doing so exists.

In conclusion, workplace addresses are likely vital to optimizing use of spatial methods in outbreak investigations, particularly in transit hubs and sites frequented by working populations. Performance of spatial methods for outbreak detection and investigation has thus far been mixed (1, 10, 20, 22–25), and it would be interesting to assess the improvement gained with the addition of workplace addresses. Use of the logarithmic distance and address selection approaches presented here potentially helps to synthesize residential with workplace (and additional) types of addresses while reducing the effect of outliers and cases unrelated to the outbreak. Additional validation using both the mock outbreak approach and more data from real outbreaks is recommended.

ACKNOWLEDGMENTS

Author affiliations: Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Republic of Singapore (Jonathan L. Chua, Vernon J. Lee, Mark I-Cheng Chen); Environmental Health Institute, National Environment Agency, Singapore, Republic of Singapore (Lee Ching Ng); Department of Emergency Medicine, Singapore General Hospital, Singapore, Republic of Singapore (Marcus E. H. Ong); Health Services and Systems Research, Duke-NUS Medical School, Singapore, Republic of Singapore (Marcus E. H. Ong); Department of Emergency Medicine, National University Hospital, Singapore, Republic of Singapore (Er Luen Lim); Department of Accident and Emergency, Changi General Hospital, Singapore, Republic of Singapore (Hoon Chin Steven Lim); Department of Emergency Medicine, Tan Tock Seng Hospital, Singapore, Republic of Singapore (Chee Kheong Ooi); Department of Emergency Medicine, KK Women's and Children's Hospital, Singapore, Republic of Singapore (Arif Tyebally); Department of Emergency Medicine, Khoo Teck Puat Hospital, Singapore, Republic of Singapore (Eillyne Seow); and National Centre for Infectious Diseases, Singapore, Republic of Singapore (Mark I-Cheng Chen).

This study was supported by the National Medical Research Council of Singapore (grant NMRC/CIRG/1384/2014). M.I.-C.C. also acknowledges funding from a startup grant from the National University of Singapore (grant R-608-000-068-133).

We thank Tien Wee Siong and Charlene Tow from the Communicable Disease Division, Ministry of Health,

Singapore, for their help to gather the food-borne and vector-borne outbreak case data.

Conflict of interest: none declared.

REFERENCES

- Buscema M, Grossi E, Breda M, et al. Outbreaks source: a new mathematical approach to identify their possible location. *Physica A Stat Mech Appl*. 2009;388(22):4736–4762.
- Jabs J, Devine CM. Time scarcity and food choices: an overview. *Appetite*. 2006;47(2):196–204.
- Kant AK, Graubard BI. Eating out in America, 1987–2000: trends and nutritional correlates. *Prev Med*. 2004;38(2):243–249.
- Kwon YS, Ju SY. Trends in nutrient intakes and consumption while eating-out among Korean adults based on Korea National Health and Nutrition Examination Survey (1998–2012) data. *Nutr Res Pract*. 2014;8(6):670–678.
- Walton K, Kleinman KP, Rifas-Shiman SL, et al. Secular trends in family dinner frequency among adolescents. *BMC Res Notes*. 2016;9:35.
- Lindgren A, Björk J, Stroh E, et al. Adult asthma and traffic exposure at residential address, workplace address, and self-reported daily time outdoor in traffic: a two-stage case-control study. *BMC Public Health*. 2010;10:716.
- Wen TH, Lin MH, Fang CT. Population movement and vector-borne disease transmission: differentiating spatial-temporal diffusion patterns of commuting and noncommuting dengue cases. *Ann Assoc Am Geogr*. 2012;102(5):1026–1037.
- Ho ZJM, Hapuarachchi HC, Barkham T, et al. Outbreak of Zika virus infection in Singapore: an epidemiological, entomological, virological, and clinical analysis. *Lancet Infect Dis*. 2017;17(8):813–821.
- Savory DJ, Cox KL, Emch M, et al. Enhancing spatial detection accuracy for syndromic surveillance with street level incidence data. *Int J Health Geogr*. 2010;9:1.
- Smith CM, Le Comber SC, Fry H, et al. Spatial methods for infectious disease outbreak investigations: systematic literature review. *Euro Surveill*. 2015;20(39):pii=30026.
- Kulldorff M. A spatial scan statistic. *Commun Stat Theory Methods*. 1997;26(6):1481–1496.
- Chen J, Roth RE, Naito AT, et al. Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of US cervical cancer mortality. *Int J Health Geogr*. 2008;7:57.
- Health Promotion Board. Report of the National Nutrition Survey 2010. 2013;9–13. https://www.hpb.gov.sg/docs/default-source/pdf/nns-2010-report.pdf?sfvrsn=18e3f172_2. Accessed August 11, 2017.
- Miller KW, Wilder LB, Stillman FA, et al. The feasibility of a street-intercept survey method in an African-American community. *Am J Public Health*. 1997;87(4):655–658.
- Spencer L, Pagell F, Hallion ME, et al. Applying the transtheoretical model to tobacco cessation and prevention: a review of literature. *Am J Health Promot*. 2002;17(1):7–11.
- Chia G, Ho HJ, Ng CG, et al. An unusual outbreak of rotavirus G8P[8] gastroenteritis in adults in an urban community, Singapore, 2016. *J Clin Virol*. 2018;105:57–63.
- Muir LE, Kay BH. *Aedes aegypti* survival and dispersal estimated by mark-release-recapture in northern Australia. *Am J Trop Med Hyg*. 1998;58(3):277–282.
- R Core Team. *R: A Language and Environment for Statistical Computing*. 2016. <https://www.r-project.org/>.
- Koh WM, Bogich T, Siegel K, et al. The epidemiology of hand, foot and mouth disease in Asia: a systematic review and analysis. *Pediatr Infect Dis J*. 2016;35(10):e285–e300.
- Fritz CE, Schuurman N, Robertson C, et al. A scoping review of spatial cluster analysis techniques for point-event data. *Geospat Health*. 2013;7(2):183–198.
- Ler TS, Ang LW, Yap GSL, et al. Epidemiological characteristics of the 2005 and 2007 dengue epidemics in Singapore—similarities and distinctions. *Western Pac Surveill Response J*. 2011;2(2):24–29.
- Stevenson MD, Rossmo DK, Knell RJ, et al. Geographic profiling as a novel spatial tool for targeting the control of invasive species. *Ecography*. 2012;35(8):704–715.
- Aamodt G, Samuelsen SO, Skrondal A. A simulation study of three methods for detecting disease clusters. *Int J Health Geogr*. 2006;5:15.
- Costa MA, Assunção RM. A fair comparison between the spatial scan and the Besag-Newell Disease clustering tests. *Environ Ecol Stat*. 2005;12(3):301–319.
- Moore DA, Carpenter TE. Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiol Rev*. 1999;21(2):143–161.