Feasibility Study for Identifying Suicide Clusters Using Real-time Coronial Data

Report prepared by

Leo Roberts

Angela Clapperton

Matthew Spittal

Centre for Mental Health

The University of Melbourne

Contact for queries

Matthew Spittal, m.spittal@unimelb.edu.au

Acknowledgement

The research contained in this report would not be possible without the Coroners Court of Victoria who initiated discussions with us about the problem of identifying suicide clusters, collected and coded the necessary real-time Victorian Suicide Register data and provided the University of Melbourne access to these data for this project. We also acknowledge funding for this research from the Department of Health Victoria and the Australian Institute of Health and Welfare.

Note

The following report contains information about the collection and reporting of suicide in Victoria. The authors acknowledge those bereaved by suicide and recognise that behind these numbers are people with families and friends who are deeply affected by suicide. Some people may find the content of this report confronting or distressing. If you are affected in this way, please contact Beyond Blue, 1300 22 4636, www.beyondblue.org.au or Lifeline, 13 11 13, www.lifeline.org.au.

Background

Suicide is a major societal problem. In Australia, there were 3,144 deaths by suicide in 2021 with suicide being the leading cause of death amongst young people (18-24 years).[1] The impact of suicides is particularly devastating for families, friends and communities when they occur in clusters. A suicide cluster is defined as a group of suicides that occur closer together in time and space than expected given the underlying suicide rate.[2] Clusters are of significant concern: (a) because they typically occur among young people and Indigenous people, groups in which news of a suicide can spread rapidly and widely; and (b) due to the potential for each suicide within a cluster to lead to further suicides in the community.[2,3]

There is strong evidence supporting the safety and effectiveness of many suicide prevention interventions,[4] including interventions for responding to suicide clusters.[5-7] But the potential for these interventions to save lives and avert the consequences of suicide for communities is limited because our current monitoring and dissemination methods are not nimble enough to identify when suicides clusters are forming in real time.[8] This stems from two key problems.

The first problem is a time lag in the reporting of suicides. The main data source that we, and other Australian researchers, have used previously to detect suicide clusters is the National Coronial Information System (NCIS). However, relying on closed cases (cases where a coroner has made a determination) means there is typically a delay of 12-18 months between the death of a person by suicide and the data becoming available in NCIS. This severely hampers the ability of current cluster detection methods to find emerging clusters and means any response is delivered too late to be useful.

The second problem is that the location of suicides and the location of the deceased residential address are usually reported at a relatively high level, such as by postcode or local government area. This creates several follow-on issues that limit the utility of the data. First, high-level aggregations are often too imprecise to detect the true location of clusters and inform targeted prevention efforts. Second, upon identifying any clusters, the precise locations of death remain unknown, offering little information to those developing prevention and postvention strategies. Finally, methods of tracking the number of suicides over time often focus on a specific geographic boundary, but since related suicides can occur outside a boundary, the full picture may not be visible to those responding to clusters.

The Coroners Court of Victoria (CCV) has been aware of these issues in preparing their regular reports for the Suicide Prevention and Response Office in the Victorian Department of Health. In early 2021, CCV approached us to see if it was possible to use their Victorian Suicide Register (VSR)[9] to overcome the problems of timeliness and data aggregation to detect suicide clusters in real time. Together, we proposed a project to the Australian Institute of Health and Welfare (AIHW) that sought to test the feasibility of using this data to undertake regular monitoring of suicide clusters in Victoria. This report summarises that work.

Project aims

Building on the work of CCV in developing a real-time suicide register and our previous work on suicide cluster detection using various data sources, our overarching aim was to test the feasibility of using the VSR to undertake regular monitoring of suicide clusters in Victoria. The key research question was: *Can modern cluster detection methods be used on real-time data with precise geocoordinates to monitor the emergence of suicide clusters?* To answer this question, the project was organised in 6 stages. These were:

1. Obtain ethical approval.
2. Develop a secure data platform for storing and analysing the VSR data.
3. Obtain additional data on the population sizes at various geographic levels (mesh blocks, statistical area 1, etc), shapefiles for these same geographic areas, undertake data cleaning and data alignment.
4. Undertake space-time cluster detection analysis to identify reference clusters.
5. Simulate real-time cluster detection analysis.
6. Report on the results.

In summarising our research, we have grouped stages 1 to 3 together and presented stages 4 and 5 in their own sections.

Stages 1 to 3: General methods

Ethical approval, agreement to share data, data storage and security

The University of Melbourne (UoM) Ethics Committee approved the project on 10th August 2021 (reference: 2021-21928-20627-3). Judge John Cain (the State Coroner) signed an order for the release of VSR data for the purpose of this project on 5th October 2021 and an information sharing agreement between the Court Services Victoria (acting on behalf of CCV) and the University of Melbourne was executed on the 28th September 2021.

In consultation with the UoM's Research Computing Services, a highly secure system for data storage and analysis was established for us that (a) housed the data safely; (b) met the intensive computer requirements for cluster detection analysis and (c) had potential for long-term use. Under this arrangement (available for the life of the project), raw data from the CCV was transferred to the research data management platform Mediaflux, located in the UoM's secure data centre. Named UoM Researchers (Leo Roberts, Angela Clapperton and Matthew Spittal) can access the data from

Mediaflux via two virtual desktop platforms, both suitable for further storage and analysis. These are (1) Researcher Desktop (which offers Linux and Windows interfaces) and (2) Research Server (which offers more computing power on a Linux interface). To enter either virtual environment, named researchers must first use their UoM login credentials (with two factor authentication) to enter the university network when on campus or enter the university's Virtual Private Network when off campus.

Suicide Data

CCV provided us with two data extracts from the VSR for the purpose of this study. Both datasets contained the following variables: a case number identifier, date of suicide, age, sex, incident location latitude, incident location longitude, incident suburb, incident location description (e.g., usual residence or non-residential location), residential location latitude, residential location longitude, residential suburb and the method of suicide. The first dataset (provided January 2022) contained this information for all Victorian suicides from January 1, 2008, to December 31, 2021. The second dataset (provided July 2022) contained updated information for suicide from January 1, 2008, to June 30, 2022, and is the basis of all analyses reported here. This second dataset held information for 9155 deaths, although 208 deaths had no associated residential address (essential for our chosen analysis – see below) and were excluded from all analyses. The resultant file had information for 8957 deaths (2252 females, 6705 males, mean age = 45 years). Residential latitude and longitude were used exclusively for cluster detection analysis, as opposed to incident latitude and longitude. This choice reflected the goal to use real-time data to identify communities in crisis, rather than to identify locations where numerous suicides occurred (sometimes called suicide hotspots).

Additional data

We explored the feasibility of identifying clusters at three relatively precise Australian Bureau of Statistics (ABS) geographic classification levels: mesh blocks (MB; usually 30 to 60 dwellings), statistical area level 1 (SA1, average population of around 400 people) and statistical area level 2 (SA2, average population of around 10,000 people). Searching for clusters at these three levels constrains any identified clusters to be either a complete single geography (e.g., a single SA2) or a collection of neighbouring geographies (e.g., a group of contiguous SA2s), depending on the size of the cluster. Under this approach, identified clusters could never, for example, cut across an SA2. Detecting MB-, SA1- and SA2-level clusters requires both population data for each area and corresponding shapefiles that specify the geographical location of each area (i.e., information about the location and population is needed for every relevant MB, SA1 and SA2). Accordingly, usual resident populations from the 2016 census for all Victorian MBs, SA1s and SA2s were extracted from ABS table builder tool. ABS shapefiles (2016 census) for the same geographies were downloaded from the ABS website. We used 2016 Census data because this snapshot approximated the mid-point of the suicide data and therefore was the best estimate of population and area location available.

Cluster detection analysis

Cluster detection analysis was performed using the opensource software, SaTScan, which detects unexpectedly large groups of observations using a method called the *scan statistic*. The scan statistic investigates the presence of clusters by exhaustively moving a window of variable dimensions across a geographic space, while recording, for each variation of the window, the observed and expected number of events inside and outside the window. Depending on the nature of analysis, each window can capture observed and expected counts in an amount of time (a temporal analysis), in a spatial area (a spatial analysis) or both (a space-time analysis). While spatial only analyses were originally considered, space-time analyses became our sole focus, with the goal of identifying both the location and duration of clusters. Space-time cluster exploration was undertaken using a retrospective discrete Poission model (one of several options in SaTScan), recommended in situations where case numbers and a background population at risk are available (e.g., census data). Under the discrete Poisson model, the number of cases in each location is Poisson-distributed, according to the population at risk. Given the null hypothesis of equal disease risk, the expected number of cases in each area is proportional to the person-years in that area (i.e., the number of people in the area multiplied by the number of years at risk).

SaTScan's method of investigating the presence of a space-time cluster can be visualised as moving a virtual cylinder across a map where the base of the cylinder represents the area under investigation and the height corresponds to time. The height and radius of the cylinder is varied within prespecified limits, gradually increasing from the specified minimum dimensions until maximum limits are reached. In our case, SaTScan calculated, for each cylinder instance, the observed and expected number of suicides *inside* the cylinder and the observed and expected numbers of suicides *outside* the cylinder. The expected number of suicides was determined by scaling the number of suicides per person-year across the whole dataset to the much smaller number of person-years inside the cylinder.

Using the observed and expected metrics, SaTScan calculates a log likelihood ratio (the test statistic) that measures the risk of suicide inside the cylinder relative to the outside, which is then maximised across all cylinders. The cylinder with the maximum test statistic is the candidate cluster least likely to have occurred by chance (i.e., the most likely cluster). Monte Carlo simulation (the default SaTScan method) is then used to test the significance of the most likely cluster and other high-likelihood clusters, with the following two steps. First, the maximum test statistic is determined from

999 randomly generated datasets (drawn from across the entire space-time landscape provided to SaTScan). Second, the original test statistic of the candidate cluster is ranked within the full set of 1000 maximum test-statistics, with the p-value corresponding to the rank (e.g., p = 50/1000 = 0.05). Technical description of space-time cluster detection using the scan statistic can be found in the SaTScan user manual[10].

Given our long-term goal of developing a suicide cluster surveillance system, it was advantageous to establish a process whereby SaTScan could be run remotely and automatically from a more generic statistical software. This was achieved using the opensource statistical software, R, via the *rsatscan* package[11] specifically written to execute SaTScan analyses inside R. Following the initial setup work, this approach allowed us to perform all data cleaning, data organisation, analysis and reporting inside the same software and allowed automation of large parts of the process from data-cleaning to mapping/plotting.

SaTScan file setup

To run the retrospective space-time discrete Poisson model, three files are needed: a case file, a coordinates file and a population file. The case file must contain the number of observed events at a given location within a given time period. For us, this meant aggregating the number of suicides by geography (MB, SA1 or SA2) and time period (day, month or year). As an example, the resultant case file might show the number of suicides in each Victorian SA2 for each month, for all the SA2/month combinations where any suicides occurred. To determine the geography associated with each suicide, MB, SA1 and SA2 shape files were linked to the residential latitude and longitude of each event with a spatial join via the R package, ‘sf’[12]. Essentially, this process determined which MB, SA1 and SA2 polygon surrounded each residential suicide location. Shape files and residential coordinates were standardised using the widely used WGS 84 latitude/longitude system (EPSG: 4326).

The coordinates file requires a latitude and longitude location for each geographical area (e.g., a single position for each SA2). For SA2s and SA1s, we defined this location as the population-weighted centroid, where the populations of constituent MBs were used as the weighting factor. The population weighting process involves determining the geometric centre of each constituent MB (i.e., the latitude and longitude of the midpoint of each polygon), then, for all SA2s and SA1s, calculating a population-weighted average of these latitude and longitude values using MB populations. For the MB level analyses, the geographic location assigned to each MB was the geometric centroid (rather than the population-weighted centroid), since populations are not available below the MB level. In the case of the discrete Poisson model, the choice between using population weighted centroids or geometric centroids as reference locations had no meaningful impact on the findings (only the number of cases within each geographical area can affect that) but does offer some advantages if mapping cluster circles.

The population file requires a population value for each area. As mentioned, we used ABS usual resident populations from the 2016 Census as the estimate for each MB, SA1 and SA2 in Victoria.

Stage 4: Reference Cluster Detection

Our first analytic step was to apply the scan statistic to the full dataset provided by the CCV (January 2008 to July 2022) and search for historical space-time clusters (referred to as reference clusters). In doing so, we aimed to assess if the scan statistic was able to uncover any statistically significant clusters (its first test), and if so, whether or not the identified clusters made sense to those at CCV who regularly monitor suicide events in the state (its second test). Conducting this initial analysis also allowed us to refine our understanding and specification of the SaTScan tool.

Method – Reference Cluster Detection

Before running any SaTScan model, several parameters that affect that computation of the scan statistic must first be specified. Table 1 outlines and explains the key parameters, the selections we made, and the associated reasoning. Notably, we deliberately varied two parameters in Table 1 (maximum cluster radius and maximum cluster duration), rather than choosing a single option. In practice, this meant repeating the MB, SA1 and SA2 models several times, adjusting the relevant parameters on each occasion. We adopted this computationally expensive approach with the ambition of converging on significant clusters with several models (i.e., robust clusters), rather than with an isolated model. Varying both parameters also clarified their impact on cluster detection.

Modelling was separately conducted using (a) suicides counts of people of all ages (the all-ages analysis), with the whole population used to determine expected counts and (b) using suicide counts of people under 25 years (the under 25 analysis), with the under 25 population used to determine expected counts. Given the additional variation of geographical level (3 levels), maximum cluster radius (4 levels) and maximum cluster duration (3 levels), 72 models were run in total (36 for all ages; 36 for under 25-year-olds). We classified clusters as possible clusters if their p value was $\leq$ 0.01. We implemented this liberal alpha value with the view that occasional false positives would have little downside in a suicide surveillance system (e.g., the ramification might be additional file review in the first instance).

Table 1. SaTScan parameters and reasons for selection

Parameter	Parameter explanation	Option used	Reason
Window shape	SaTScan can progressively scan circular or ellipsoid windows. Ellipsoid windows can be helpful when exploring clusters that are unlikely to be captured by a circle (e.g., along coastlines or other boundaries) and are potentially well suited to certain hotspot (incident) analyses.	Circular windows	Since suicide residences were distributed across the state and do not reliably sit alongside borders, circle clusters were considered a good default option. Future research could consider the potential benefits of using ellipsoid windows. Note that by running SaTScan models at the MB, SA1 or SA2 level, clusters are forced to the shape of those geographies and are not actually circular. SaTScan still theoretically explores circular windows, but in effect, if the circle captures any part of a geography, all cases in that geography will make up the observed count, as opposed to those strictly within the circle. SaTScan will output the mid-point of the cluster and the radius, so a circle could be drawn on a map, but the most accurate depiction of the cluster would be the outline of the geographies involved in the cluster.
Scan areas with high or low rates	Using high rates searches for clusters that are characterised by a high number of events relative to the number expected (as opposed to exploring areas with a low number of events relative to the number expected)	High rates	To identify areas with unusually high suicide rates.

Time aggregation level	Models can be specified at the day, month, or year level. For example, a day-level model explores the number of cases in each area each day. Clusters emerging from a day-level model have a minimum duration of one day and identify precise temporal clusters in day units. Clusters emerging from a month-level model have a minimum size of one month and identify clusters in units of months.	Month	Models were initially tested at the day level (e.g., using the number of cases in each SA2 each day). However, no significant clusters were found across numerous models explored. Month-level models remained as our best option to detect clusters in real-time.
Maximum cluster radius	A limit can be placed on the size of the circle of identifiable clusters.	20km, 50km, 100km	We explored three maximum cluster radius options, for one, to better understand the impact on cluster statistics and second, to use a variety of models to converge on probable clusters. We also explored a 200km maximum, but this did not identify an additional cluster.
Maximum cluster duration	A limit can be placed on the duration of identifiable clusters.	1 month, 3 months, 6 months, 12 months	As above, we applied a variety of maximum cluster lengths to understand the statistical impact and converge on likely clusters.

Results - Reference Cluster Detection

Table 2 summarises the details of the six significant clusters detected in the all-ages analysis. Table 3 contains details of the three significant clusters detected in the under-25 analysis. Both tables report the geographic unit of analysis, cluster duration (in months), the number of composite geographical units in the cluster (e.g., the number of MBs that make up the cluster), basic cluster metrics (observed count, expected count and p value), the estimated population living in the area in which the cluster was detected, and the number of models that detected the cluster as significant (i.e., out of 36 run).

Both tables demonstrate that the scan statistic can be used to identify significant clusters in a jurisdiction like Victoria (~700 suicides per year). Discussions with CCV and the Suicide Prevention and Response Office about these cluster locations on the 8^{th} and 10^{th} of July 2022 reinforced the viability of our approach, with most significant clusters mapping to the times and places they were concerned about.

Table 2: Possible clusters for all age groups

Cluster	Geography Level	Duration of Cluster (Months)	Number of Areas in Cluster	Observed	Expected	p value	Population of Cluster Area	Number of Models Identifying Cluster
1	MB	4	121	8	0.27	0.04	7800	6
2	MB	10	34	8	0.23	0.02	2700	3
3	MB	1	7	nr	0	>0.01	>150	12
4	SA1	4	31	9	0.47	0.08	13500	6
5	SA1	9	16	9	0.43	0.04	5600	3
6	SA2	10	3	17	3.44	0.07	39500	3

Note: nr = not reported due to small cell sizes. Where this occurs, the p-value and the population size have been rounded up to preserve the anonymity of the cluster's location.

Table 3: Possible clusters for all people aged < 25 years

Cluster Name (Central SA2 Name)	Geography Level	Duration of Cluster (Months)	Number of Areas in Cluster	Observed	Expected	p value	Population of Cluster Area	Number of Models Identifying Cluster
1	SA1	1	18	nr	0.01	>0.10	1850	3
2	SA2	1	4	nr	0.04	>0.01	10200	8
3	SA2	2	6	5	0.11	0.01	14950	6

Note: nr = not reported due to small cell sizes. Where this occurs, the p-value and the population size have been rounded up to preserve the anonymity of the cluster's location.

Stage 5: Simulation of Real-time Cluster Detection

Having established that the scan statistic could detect meaningful space-time suicide clusters in Victoria, we next assessed our ability to identify a cluster in real time. This involved a computationally expensive retrospective analysis, in which we ran the scan statistic over progressively more recent windows of data. By imagining a situation where (a) the CCV would provide us with a new month of data every month and (b) where we would execute the scan statistic upon receiving that data, we could discover the month of data that lead to the statistical recognition of a given cluster (i.e., the cluster detection date). By comparing the duration of an identified cluster (i.e., when the cluster finished) to the cluster detection date, we could assess the gap and estimate how quickly we could have identified the cluster if operating as an active suicide surveillance system. In other words, could we detect the cluster months after it finished? As it finished? Or when it was an earlier form of an even bigger cluster?

Method – Simulation of Real-time Cluster Detection

Our approach was to run a series of models on moving two-year windows of suicide data, adding one new month while dropping the oldest month with each model run. At the conclusion of modelling, we collated information about all significant clusters found, especially noting the window of data that *first* revealed each cluster (significant clusters were usually found repeatedly with sequential models). As mentioned, this allowed us to estimate when clusters would have been identified if the CCV had provided us with a new month of data, every month. Noting that SaTScan calculates expected counts using the base rate of events across the entire data set, it was important to provide SaTScan with enough data to compute valid expected counts and provide relatively up-to-date base rate information to compute relevant expected counts. Two-year windows (containing around 1300 suicide events each) were deemed sufficient to meet both requirements.

To clarify the simulation approach, the first window modelled was January 2015 to December 2016 (inclusive, i.e., 24 months of data), with the following model assessing suicide counts between February 2015 and January 2017. This process was repeated until the final model examined suicide counts between July 2020 and June 2022. If, for example, the same cluster was significant in the May 2020 - April 2022 model, the June 2020 - May 2022 model and the July 2020 - June 2022 model, we determined that, hypothetically, we would have first identified that cluster at roughly the end of April 2022. For simulated models, we applied the same parameters as found in Table 1 except we fixed the maximum radius to 100km and the maximum duration to 12 months. This adjustment was made to minimise computing time (given the very large number of models being run) while capturing as many clusters as possible. The simulation was performed once again at the MB, SA2 and SA2 levels, separated for people of all ages and for people under 25. In total, this resulted in the execution and analysis of 402 models (201 for each age group).

Results – Simulation of Real-time Cluster Detection

All-ages simulation. Table 4 summarises the details of earliest detection of each significant cluster identified across the all-ages simulation. In total, 16 distinct significant clusters were identified, although on occasions, technically distinct clusters could be practically mapped to a single cluster (e.g., Cluster 3). The duplication occurred because separate MB, SA1 and SA2 models homed in on the same set of suicides or because a near-by space-time window also turned out to be a significant cluster. In either case, a slightly different cluster boundary was drawn even if it was capturing essentially the same group of events.

Table 4: Significant clusters identified in the all-ages simulation. Functionally identical clusters have been grouped (e.g., 3a - 3e), even if they have marginally distinct cluster dimensions.

Cluster	Geography Level	Duration of Cluster (Months)	Number of Areas in Cluster	Observed	Expected	p value	Population of Cluster Area
1a	SA2	7	2	11	1.746	0.094	26200
2a	SA2	11	3	17	4.156	0.097	40000
3a	MB	10	34	8	0.257	0.006	2700
3b	MB	9	101	9	0.497	0.043	6000
3c	MB	7	101	8	0.395	0.1	6000
3d	SA1	9	16	9	0.475	0.015	6000
3e	SA1	7	16	8	0.367	0.023	6000
4a	SA2	8	3	15	3.016	0.04	40000
4b	SA2	10	3	17	3.825	0.041	40000
4c	SA2	9	4	17	3.845	0.033	44000
5a	MB	8	220	10	0.749	0.086	10000
5b	SA1	8	27	10	0.859	0.089	11500
6a	MB	1	7	nr	0.001	>0.01	>150
7a	SA2	5	6	14	2.825	0.08	58600
8a	MB	8	2	nr	0.003	>0.05	>50
8b	SA1	8	1	nr	0.007	>0.10	>100

Note: nr = not reported due to small cell sizes. Where this occurs, the p-value and the population size have been rounded up to preserve the anonymity of the cluster's location.

Figures 1 and 2 display the detection timeline for, respectively, Cluster 3a and Cluster 7a. These figures indicate (a) the suicide history within the geographical area of the cluster since 2015; (b) the specific two-year window of data modelled within which the cluster was detected (green arrow); (c) the duration of the cluster identified (blue arrow); and (d) when the cluster would have (theoretically) been identified given provision of suicide data every month.

Figure 1 indicates that Cluster 3a would have been in identified at the end of month 59, immediately following provision of that month's data. Figure 1 also indicates that before the beginning of the cluster, suicide was uncommon in the area, with no events recorded since the beginning of 2015. While the cluster appeared to be detected when it was effectively over, two further deaths occurred later in the timeline. Speculatively, given timely identification of this cluster, there may have been a chance for intervention in the area, although whether these two deaths could have been averted is an open question. Figure 2 implies that the Cluster 7a would have been identified at the end of month 85, once again immediately following provision of that month's data. While the area exhibits a history of suicide, the identified cluster appears to capture a relatively large group of deaths. Moreover, two further deaths occurred shortly after simulated identification of the cluster, once again highlighting the potential for intervention. Figures 1 and 2 both imply that given a two-year window of Victorian Suicide data, and monthly data updates, significant clusters could be detected as soon as they recognisably become a cluster (within a month).

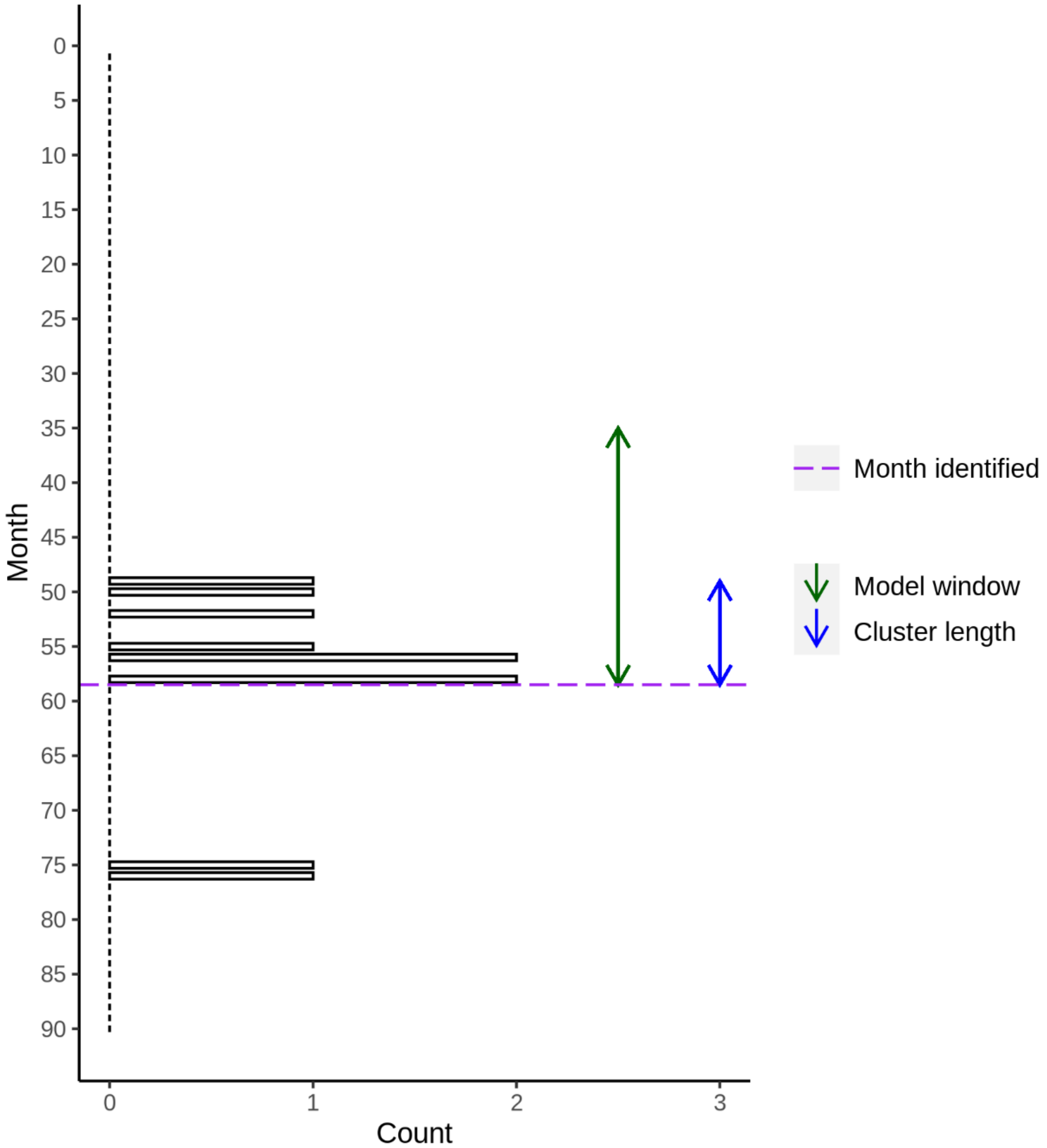

Figure 1. Timeline of cluster detection and suicide events for all-ages cluster 3a (detected with an MB model)

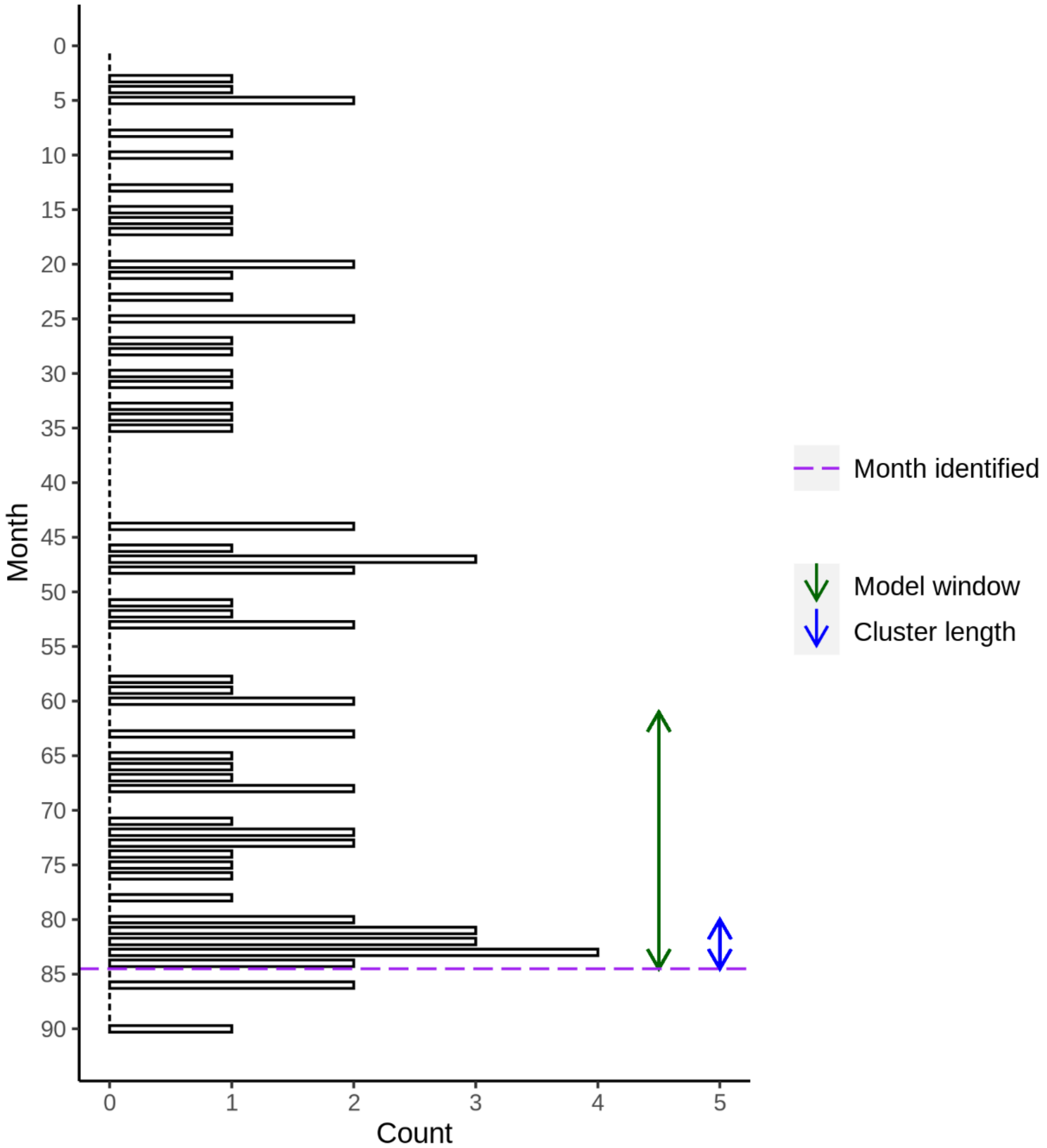

Figure 2. Timeline of cluster detection and suicide events for all-ages cluster 7a (detected with an SA2 model).

Under-25 simulation. Table 5 summarises the details of the earliest detection of each significant cluster identified across the under-25 simulation. In total, eight distinct significant clusters were identified, with duplication again observed in some technically distinct clusters (e.g., Cluster 4). Cluster detection timelines presented in Figures 3 and 4, respectively for Cluster 4c and Cluster 3c. Figure 3 demonstrates little history of suicide inside the cluster boundary until the cluster itself occurred between months 84 and 85. Our modelling indicates that this cluster would have been identified following supply of month 85's data. Likewise, Figure 4 shows little suicide history within the cluster boundary until four deaths occurred in one month (month 59). Our modelling indicates that this cluster would have been identified following supply of data at the end of that month.

Table 5: Significant clusters identified in the under-25 simulation. Functionally identical clusters have been grouped (e.g., 3a – 3c), even if they have marginally distinct cluster dimensions.

Cluster	Geography Level	Duration of Cluster (Months)	Number of Areas in Cluster	Observed	Expected	p value	Population of Cluster Area
1a	SA1	12	100	8	0.536	0.083	1200
2a	SA1	4	6	nr	0.009	>0.01	>600
3a	MB	1	82	nr	0.006	>0.01	>1500
3b	SA1	1	18	nr	0.008	>0.01	>2000
3c	SA2	1	4	nr	0.045	>0.01	>10500
4a	SA1	3	156	6	0.211	0.081	16200
4b	SA1	5	156	7	0.338	0.093	16200
4c	SA2	2	6	5	0.131	0.008	15000

Note: nr = not reported due to small cell sizes. Where this occurs, the p-value and the population size have been rounded up to preserve the anonymity of the cluster's location.

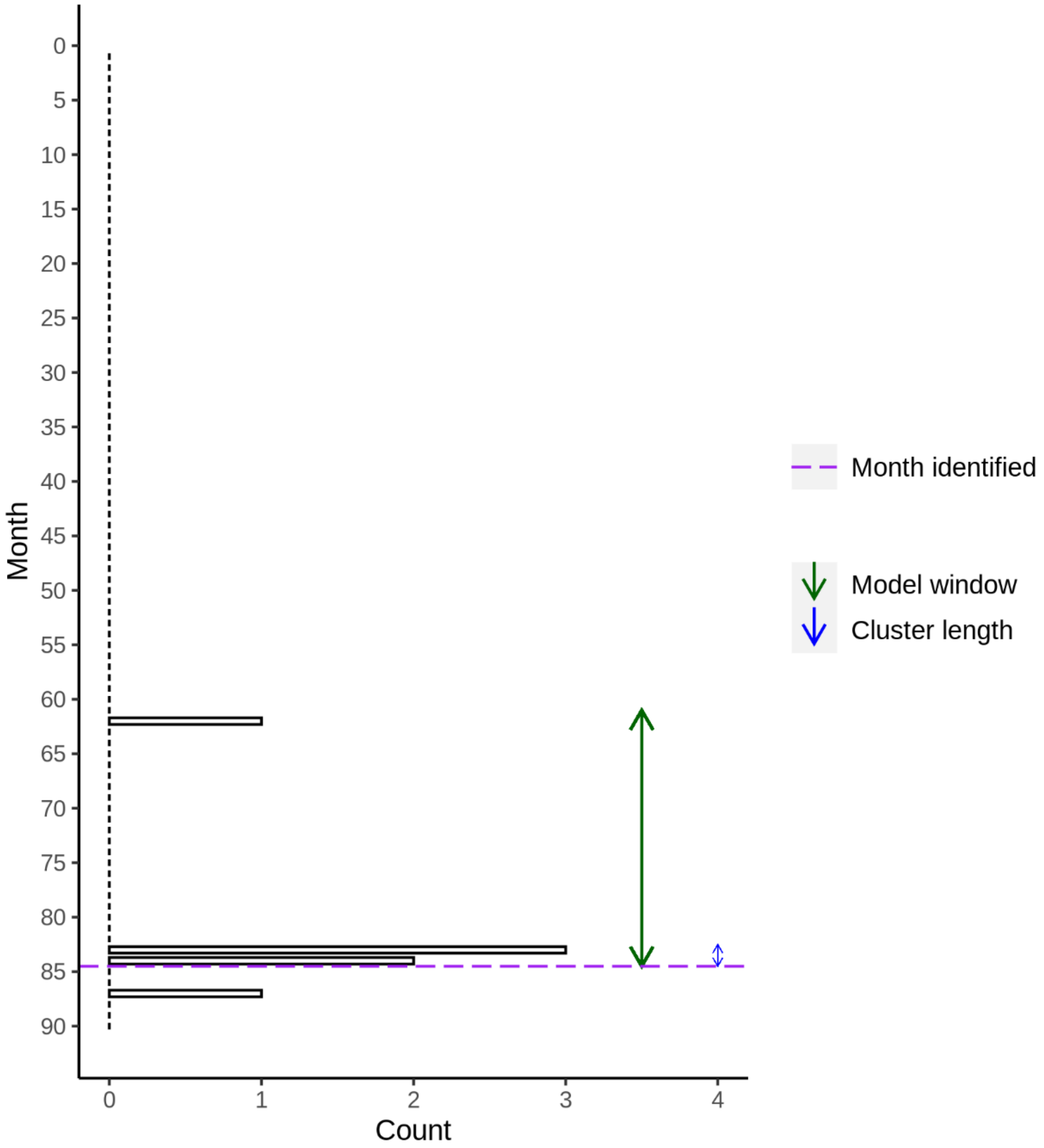

Figure 3. Timeline of cluster detection and suicide events for Cluster 4c (detected with an SA2 model).

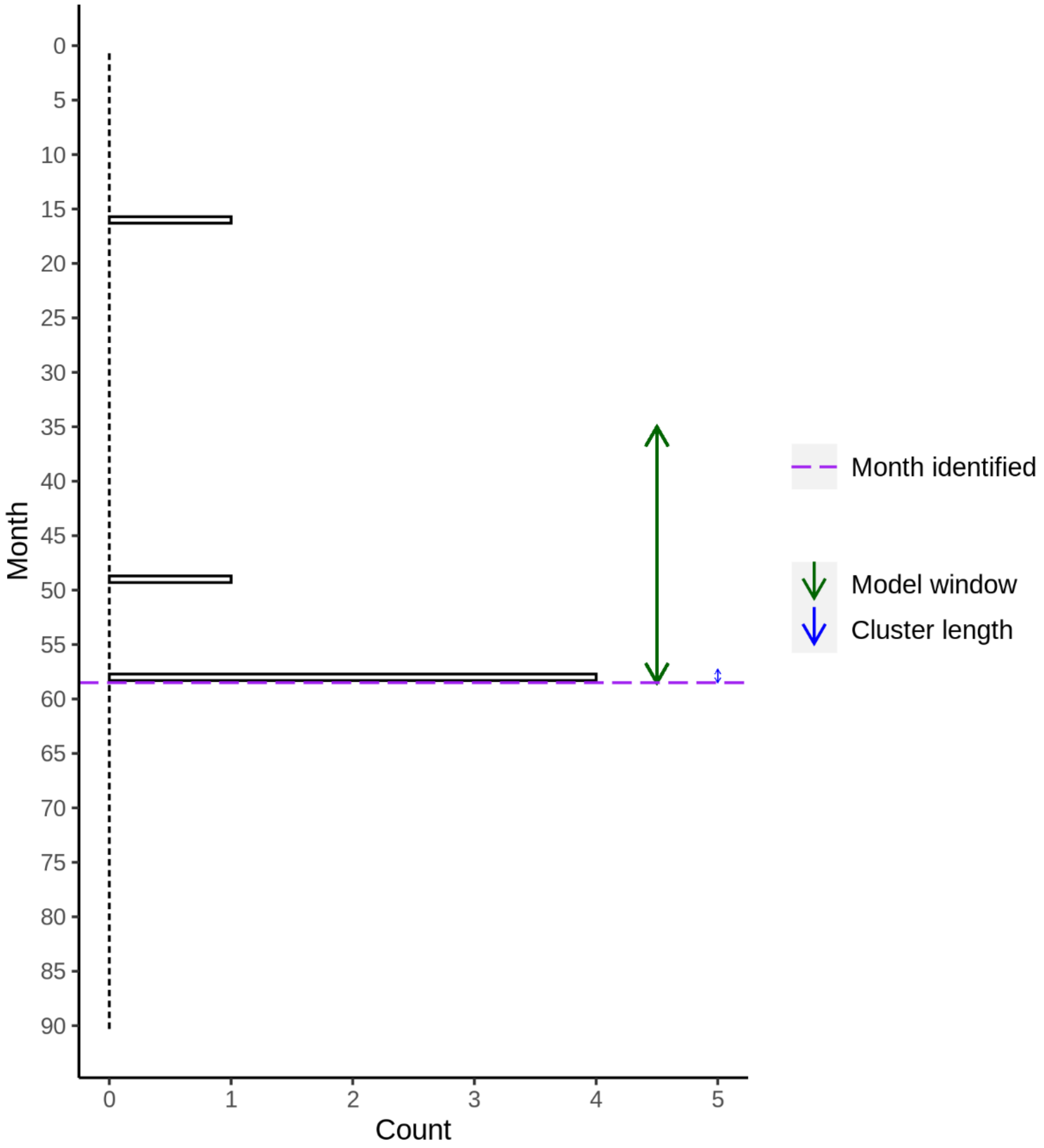

Figure 4. Timeline of cluster detection and suicide events for Cluster 3c (detected with an SA2 model).

Discussion

In this project, we tested the feasibility of a real-time suicide cluster monitoring system using real-time suicide data collated and provided by CCV. The initial steps included obtaining ethical approval, approval to access the VSR data and procuring a secure data platform for data storage and use, all of which were accomplished by late 2021. In the later stages of 2021 and throughout 2022 we carried out extensive modelling of the historical data available to us. Through this process, we refined our analysis, conducted finalised space-time cluster detection analyses (identifying a series of reference clusters) and simulated real-time cluster detection analysis.

The major findings to emerge from this research are that (a) the scan statistic can be applied to VSR data to retrospectively detect significant suicide clusters in Victoria and (b) that monthly provision of suicide data does, in principle, allow us to detect the presence of suicide clusters in near real time. As our illustration of the cluster timeline suggests (Figures 1, 2, 3), this real-time identification of clusters creates new opportunities for community-level interventions to prevent future suicide in the area (demonstrable in our simulation), potentially reducing the growth of clusters.

The implication of these findings is that it is feasible to combine the VSR with the scan statistic to develop a real-time suicide cluster detection system. As noted in the introduction, such a system has two advantages over existing approaches to detect clusters. Most obviously, it is in real time (unlike all previous studies we have been involved in), but equally importantly, it overcomes the failure to detect clusters that cross large geographic boundaries (e.g., local government areas). This approach therefore solves both problems that CCV identified at the outset of the project.

Future directions

Technical refinement of cluster detection. With more funding and time, further background testing of SaTScan-based modelling techniques could take place (or alternative space-time cluster detection approaches). The following possibilities exist for fine tuning the cluster detection analysis:

- The use of ellipse-shaped windows (instead of the circular window setting) could allow improved cluster detection in coastal regions, or along other natural boundaries, but this is yet to be tested.
- The use of the 2021 census to provide more up-to-date population estimates, and thus improve expected suicide count accuracy for the analysis of more current suicide data. Annual population estimates at area levels, if obtainable, would be even better.
- Exploration of "area-free" cluster detection methods. One of the available SaTScan methods, the space-time permutation model, detects clusters without the need to aggregate events by area (e.g., by SA2). The ability to use such an approach (or an alternative) would allow detection of clusters without any boundary constraints. We were able to briefly test the space-time permutation model, but it did not reveal significant clusters. Nevertheless, further investigation of this and alternative area-free methods could lead to a superior cluster-detection approach. This approach has been used in Wales[13] and we are aware of this approach being used more recently in Taiwan.
- Simulating cluster detection with bi-monthly provision of data. SaTScan has few temporal granularity options (day, month, year-level options), of which we found month-level analysis best suited to suicide cluster detection. However, the scan statistic could still be meaningfully run with a fortnightly data update (or even more frequently). Any deaths in the early stages of the month would be accounted for,

which could facilitate even earlier cluster detection (e.g., a cluster might be detected halfway through the month instead of at the end). A bi-monthly simulation could be conducted to assess the likely benefits.

Development and implementation of a suicide cluster surveillance system in Victoria. Given the feasibility of our approach for detecting suicide clusters in real-time using the VSR, and given that this project came out of CCV's desire to have improved monitoring of suicide clusters, the next logical step is to begin working with them and the Suicide Prevention and Response Office to gauge interest in implementing the system. If they were interested in using this system for ongoing monitoring, then we see several avenues for future collaboration. These are:

- Negotiating access to the VSR for monitoring. Our current agreement with CCV is to use the data for the purpose of this research project. Therefore, a new agreement would need to be negotiated if the data were to be used for ongoing monitoring.
- Further developing a suite of reports for CCV, the Suicide Prevention and Response Office and other end-users. The outputs we have developed (the timeseries plots in Figures 1-4 and interactive maps that are not reported here) are a good starting point but likely require further refinement. We could lead a project where we co-design the outputs with data users, implement them so that the monitoring system automatically generates these outputs, and then iteratively refine the reports until users are satisfied.
- Reaching agreement about how the reports should be disseminated. A best-practice approach would be for the University of Melbourne to upload the interactive map and monthly reports to a secure portal. From there, authorised users could log-in to the portal to access reports. Access would be controlled, so for instance, if an end-user

left their role, access could be revoked, and similarly a new authorised user could be given access.

Developing a national suicide cluster monitoring system: The feasibility of detecting suicide clusters in Victoria opens up the possibility of extending the system to other states and territories in Australia. As each state and territory now has their own real-time suicide register, it would theoretically be possible to apply the scan statistic with the settings we have used to these if the registers contain the minimum data needed for such analyses. A broader question concerns the use of other data – specifically data on self-harm – for cluster monitoring. Ambulance data and hospital presentation data could be used for this. For this to be successful, we would need to repeat the analyses done here to identify the right set of parameter settings for these data. It is also an open question as to whether it is better to combine all the data into a single dataset for analysis or develop separate systems for suicide and self-harm (and potentially one system for ambulance data and another for hospital presentation data). We would need to undertake research to resolve these important issues, but we also need to consult with end users to better understand what is most useful for them.

Summary

This work arose because CCV, a data-savvy court that uses sophisticated methods to monitor suicides, recognised they could not adequately monitor suicide clusters using their existing methods. This feasibility study therefore sought to understand if statistical/epidemiological cluster detection methods using real-time data with precise geocoordinates could be used to monitor the emergence of suicide clusters. Our findings show that this approach is highly feasible. Using data from the VSR, we have been able to use the scan statistic to retrospectively identify suicide clusters that largely match areas CCV and the Suicide

Prevention and Response Office were concerned about. We have also been able to apply these settings to near real-time data to simulate how suicide cluster monitoring would be done in practice; that is, using a rolling 2-year window of data for monitoring. We were able to identify a similar set of clusters with this data and produce outputs showing the timing and location of these clusters. These encouraging findings provide an ideal platform from which to consider what a suicide cluster monitoring system might look like, if implemented, in Victoria, and potentially across Australia.

References

1. Australian Bureau of Statistics. Causes of death in Australia, 2020. Canberra: ABS, 2021.

2. Hawton K, Hill NTM, Gould M, John A, Lascelles K, Robinson J. Clustering of suicides in children and adolescents. *The Lancet Child & Adolescent Health* 2020; **4**(1): 58-67.

3. Hill NT, Spittal MJ, Pirkis J, Torok M, Robinson J. Risk factors associated with suicide clusters in Australian youth: Identifying who is at risk and the mechanisms associated with cluster membership. *EClinicalMedicine* 2020; **29**: 100631.

4. Zalsman G, Hawton K, Wasserman D, et al. Suicide prevention strategies revisited: 10-year systematic review. *The Lancet Psychiatry* 2016; **3**(7): 646-59.

5. Department of Health and Ageing. Developing a community plan for preventing and responding to suicide clusters, 2012.

6. Lai CCS, Law YW, Shum AKY, Ip FWL, Yip PSF. A Community-Based Response to a Suicide Cluster. *Crisis* 2020; **41**(3): 163-71.

7. Public Health England. Identifying and Responding to Suicide Clusters: A Practice Resource. London, 2019.

8. Baran A, Gerstner R, Ueda M, Gmitrowicz A. Implementing Real-Time Data Suicide Surveillance Systems. *Crisis* 2021; **42**(5): 321-7.

9. Sutherland G, Milner A, Dwyer J, et al. Implementation and evaluation of the Victorian Suicide Register. *Australian and New Zealand Journal of Public Health* 2018; **42**(3): 296-302.

10. Kulldorff M. SaTScan user guide2022. http://www.satscan.org (accessed 17/01/2023).

11. Kleinman K. rsatscan: Tools, Classes, and Methods for Interfacing with SaTScan Stand-Alone Software. R package version 0.3.9200 ed; 2015.

12. Pebesma E. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* 2018; **10**(1): 439-46.

13. Jones P, Gunnell D, Platt S, et al. Identifying probable suicide clusters in Wales using national mortality data. *PLoS One* 2013; **8**(8): e71713.