

# Geographically regulated designs of incidence surveys can match the precision of classical survey designs whilst requiring smaller sample sizes: the case of snakebite envenoming in Sri Lanka

Dileepa Senajith Ediriweera,<sup>1</sup> Tiloka de Silva,<sup>2</sup> Anuradhani Kasturiratne,<sup>3</sup> Hithanadura Janaka de Silva,<sup>4</sup> Peter Diggle<sup>5</sup>

**To cite:** Ediriweera DS, de Silva T, Kasturiratne A, *et al.* Geographically regulated designs of incidence surveys can match the precision of classical survey designs whilst requiring smaller sample sizes: the case of snakebite envenoming in Sri Lanka. *BMJ Global Health* 2022;**7**:e009500. doi:10.1136/bmjgh-2022-009500

**Handling editor** Seye Abimbola

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjgh-2022-009500>).

Received 2 May 2022  
Accepted 24 September 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Dileepa Senajith Ediriweera; [dileepa@kln.ac.lk](mailto:dileepa@kln.ac.lk)

## ABSTRACT

**Background** Snakebite envenoming is a neglected tropical disease. Data from the worst affected countries are limited because conducting epidemiological surveys is challenging. We assessed the utility of inhibitory geostatistical design with close pairs (ICP) to estimate snakebite envenoming incidence.

**Methods** The National Snakebite Survey (NSS) in Sri Lanka adopted a multistage cluster sampling design, based on population distribution, targeting 1% of the country's population. Using a simulation-based study, we assessed predictive efficiency of ICP against a classical survey design at different fractions of the original sample size of the NSS. We also assessed travel distance, time taken to complete the survey, and sensitivity and specificity for detecting high-risk areas for snake envenoming, when using these methods.

**Results** A classical survey design with 33% of the original NSS sample size was able to yield a similar predictive efficiency. ICP yielded the same at 25% of the NSS sample size, a 25% reduction in sample size compared with a classical survey design. ICP showed >80% sensitivity and specificity for detecting high-risk areas of envenoming when the sampling fraction was >20%. When ICP was adopted with 25% of the original NSS sample size, travel distance was reduced by >40% and time to conduct the survey was reduced by >75%.

**Conclusions** This study showed that snakebite envenoming incidence can be estimated by adopting an ICP design with similar precision at a lower sample size than a classical design. This would substantially save resources and time taken to conduct epidemiological surveys and may be suited for low-resource settings.

## INTRODUCTION

Snakebite envenoming is a neglected tropical disease, mainly affecting poor rural communities.<sup>1</sup> Robust survey data on snakebite

## WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Snakebite envenoming is a neglected tropical disease, and robust survey data on snakebite incidence are scarce in most of the affected countries. Only a few community-based surveys have been conducted in the affected countries due to the lack of resources and logistical difficulties

## WHAT THIS STUDY ADDS

⇒ Inhibitory geostatistical designs with close pairs (ICP) for incidence surveys can match the precision of classical survey designs at lower sample sizes.  
⇒ The ICP designs showed a lower predictive variance than classical designs, indicating ICP designs were able to produce more reliable predictions.  
⇒ The ICP method showed a lower time to complete the survey than the classical sampling method.

## HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Resource requirements and time to complete surveys can be reduced without increasing the distance to travel by adopting ICP design for epidemiological surveys.

incidence are scarce in most of these countries, and most estimates of snakebite incidence rely on hospital-based data.<sup>2</sup> However, hospital records are known to underestimate the burden.<sup>3</sup> Among countries where snakebite envenoming is a public health problem, a small number, including Thailand and Brazil,<sup>4,5</sup> have attempted to estimate incidence by making snakebite notifiable, but here again, the data relies on reports from health-care institutions. Very few community-based surveys have been conducted to estimate snakebite incidence, especially nationwide

surveys, due to the lack of resources and logistical difficulties.<sup>3,6</sup>

Sri Lanka is a tropical island nation with a high incidence of snakebite envenoming.<sup>7</sup> The island is located in the Indian Ocean, with a land area of 65 000 km<sup>2</sup> and a population exceeding 20 million. There are 100 terrestrial snake species in the country, 6 of which are considered medically important (*Naja naja*, *Bungarus ceylonicus*, *Bungarus caeruleus*, *Daboia russelii*, *Echis carinatus* and *Hypnale hypnale*).<sup>8</sup> Sri Lanka was the first country to estimate the country-wide community incidence of snakebite and envenoming by conducting a National Snakebite Survey (NSS) in 2012–2013.<sup>3</sup> The NSS adopted a multi-stage cluster sampling design, targeting 1% of the total population with cluster selection based on the population distribution; more clusters were chosen in population-dense areas and vice versa.

From a geographical prediction point of view, the selection of clusters in NSS could be inefficient due to the risk of oversampling and undersampling in areas with high and low population densities, respectively.<sup>9</sup> The resulting geographical predictions are likely to be associated with varying degrees of precision over the region of interest, with especially low precision in areas with low population densities.<sup>10</sup> This can be addressed by adopting geostatistical sampling methods. Geostatistical sampling designs have been shown to be effective in prevalence mapping, including in malaria<sup>11</sup> and in several neglected tropical disease settings such as lymphatic filariasis<sup>10</sup> and soil transmitted helminth infections.<sup>12,13</sup> An inhibitory geostatistical design with close pairs (ICP) is one example of geostatistical sampling design. In an ICP design, where primary sample units are separated by a pre-specified minimum distance, which safeguards against the geographically unbalanced selection of sampling units as occurred in the NSS.<sup>9</sup>

The NSS in Sri Lanka was feasible given the country's relatively small geographical size. Even so, the survey took 11 months to complete and was logistically challenging. As such, undertaking national surveys on the same scale in other countries with high snakebite burdens and resource constraints (or even repeating the same survey within Sri Lanka) would be very challenging. Therefore, this study aimed to apply a geostatistical sampling design, specifically ICP, to estimate the incidence of snakebite envenoming and highlight the utility of geostatistical designs to achieve a similar precision to more resource intensive sampling designs such as those based on population distribution.

## METHOD

### Sampling

The NSS in Sri Lanka is one of the largest surveys conducted in neglected tropical disease settings. Therefore, we considered NSS as the basis for this simulation study. The NSS was a country-wide community-based cross-sectional survey conducted between August 2012

and June 2013. It was the first survey to be conducted at a national level to estimate snakebite incidence.<sup>3</sup> The true snakebite incidence in Sri Lanka was unknown at the time, and approximately 1% of the population of the country was sampled.

### Original sampling design adopted in NSS: classical survey design

Sri Lanka has nine provinces and 25 districts. The last census of population and housing carried out before the NSS was conducted in 2001 and covered only 18 districts, where the population was found to be 18.8 million living in 3.9 million households. Based on these numbers, the investigators estimated the population of 2012 to be 20 million and the number of households to be 4.5 million. Accordingly, to cover 1% of the population, 45 000 households were sampled, divided equally among the nine provinces.

The NSS adopted a classical survey design (ie, multi-stage cluster sampling method) to recruit participants. The Grama Niladhari (GN) divisions are the smallest administrative unit in the country; there are 14 022 such divisions in the country. The GN divisions were considered as the first-stage sampling units (eg, clusters), with 125 GN divisions selected from each province. A total of 125 GN divisions were then proportionately allocated to the districts within a province based on the population distribution between districts (ie, proportional allocation to strata). The GN divisions within a district were selected by simple random sampling. The sampled GN divisions are shown in online supplemental figure 1.

A household was considered as the second-stage sampling unit, with 40 households sampled from each GN division. The first household was randomly selected from the electoral register, after which proximity selection was adopted to select the next 39 households (ie, 'next nearest'). All members of the household were recruited for the survey, which collected data on all snakebites as well as snakebites with clinical features of systemic and/or local envenoming (presence of local tissue necrosis at the site of the bite, presence of neurotoxicity or bleeding manifestations) that took place within the previous 12 months. All bite events in a household were recorded with their time and location. These included bites that occurred to a person more than once and those that resulted in death.

### Inhibitory geostatistical design with close pairs

Snakebite incidence estimated using data from the NSS showed substantial geographical variation. The previous geostatistical models indicated the existence of a geographical correlation between events that persisted after adjustment for covariate effects, and a further a non-negligible component of unexplained, geographically uncorrelated component of variation (ie, a nugget effect in the residual covariance structure of the data). Since an inhibitory geostatistical design alone cannot reliably estimate all aspects of this complex geographical distribution (ie, covariance structure associated with a

combination of geographical and non-geographical variation), we chose an ICP to select the first-stage sampling units (ie, GN divisions) in this study.<sup>9</sup> An ICP has two types of sample points; primary sample points which maintain a minimum distance between any two sampled locations, and supplementary sample points, each of which is chosen to lie within a given radius of a primary sample point, so generating close pairs.

The centroids of all the 14022 GN divisions in the country formed our sampling frame. The steps followed in the ICP sampling method are outlined below, where  $n$  is the total number of required sample points,  $\delta$  is the minimum distance between any given two primary sample points,  $k$  is the number of supplementary sample points (ie, close pairs), and  $\zeta$  is the radius of the disc within which a supplementary or paired point is added.

1. Draw a sample of locations  $x_i: i=1, \dots, n-k$  randomly from the sampling frame (ie, 14022 GN divisions).
2. Set  $i=1$ .
3. Calculate the minimum distance,  $d_{\min}$ , from  $x_i$  to all other  $x_j$  in the sample.
4. If  $d_{\min} \geq \delta$ , increase  $i$  by 1 and return to step 3 if  $i \leq n$ , otherwise stop.
5. If  $d_{\min} < \delta$ , replace  $x_i$  with a new location drawn at random from the sampling frame and return to step 3.
6. Sample  $k$  locations from  $x_1, \dots, x_{n-k}$  without replacement and denote them  $x_j^*, j=1, \dots, k$ .
7. For,  $j=1, \dots, k$ ;  $x_{n-k+j}$  is chosen at random from all as-yet unsampled locations within the disc with centre  $x_j^*$  and radius  $\zeta$ .

In our study,  $k$  was selected as 20% of the total number of required sample points,  $n$ ,  $\delta$  was set at the maximum value for which it was possible to sample  $n-k$  locations, and  $\zeta$  was selected as one third of  $\delta$ .

### Simulation study

We undertook a simulation study to assess the predictive efficiency of classical survey design against ICP with different fractions (ie, thinning) of the original NSS sample size. We followed the same modelling approach to estimate the snakebite envenoming incidence using the same explanatory variables as in the original study.<sup>3</sup>

The true envenoming bite incidence in Sri Lanka is unknown, therefore, predictive incidence maps which were already published based on the original NSS data were assumed to be the gold standard. The Medical Officer of Health (MOH) divisions are the smallest administrative divisions in the public healthcare system in Sri Lanka. Public health initiatives are implemented through these divisions. We defined the MOH areas having an average envenoming bite incidence of more than the national envenoming bite incidence (ie, 151 per 100 000) as high-risk MOH areas.

Incidence maps were then developed, and predictive variances were calculated as follows:

1. Simulate the number of envenoming bites in the chosen GN divisions.

2. Fit geostatistical models for the simulated envenoming bites in step 1 and develop predictive maps using a prediction grid surface.
3. Calculate the average predictive variance (APV) over the prediction grid surface, also the sensitivity and specificity for detecting high-risk MOH areas.

### Geostatistical model

We used the same geostatistical model that was developed from the original NSS data<sup>3</sup> to model the simulated envenoming bites. At location  $x_p$ , the number of envenoming bites,  $y_i$ , out of  $n_i$  individuals sampled are independent binomial events conditional on an unobserved geographical stochastic process  $S(x)$ , with probabilities  $p(x_i)$ , where

$$\log \{p(x)/[1-p(x)]\} = \beta_0 +$$

$$\beta_1 * \text{elevation} + \beta_2 * I(\text{elevation} > 195) * (\text{elevation} - 195) +$$

$$\beta_3 * \text{climatezone}_2 + \beta_4 * \text{climatezone}_3 +$$

$$\beta_5 * \text{populationdensity} + S(x) \text{equation (1)}$$

In equation (1),  $I(\text{elevation} > 195)$  is equal to one of elevation is greater than 195m, and zero otherwise. The term  $S(x)$  captures any unexplained (ie, residual) geographical variation after adjusting for the explanatory variables (ie, covariates) and is modelled as a Gaussian process with mean zero, variance  $\sigma^2$  and correlation structure  $\text{Corr}[S(x), S(x')] = \exp(-u/\phi)$ , where  $u$  is the distance between  $x$  and  $x'$ , and  $\phi$  represents the scale of geographical correlation. Hence, the envenoming bite incidence at location  $x_i$  depends on the explanatory variables observed at location  $x_i$  and on  $S(x_i)$ , while  $p(x)$  is the probability that a person at location  $x$  will experience a bite.<sup>14</sup>

### Simulating the number of envenoming bites

Sri Lanka has 140 22 GN divisions. We selected different fractions of the original NSS sample size (ie, 50%, 33%, 25%, 20% and 10%) to conduct the simulations for both classical survey and ICP designs.

The simulation of envenoming bites was conducted as follows:

1. Draw a sample of GN divisions  $c_i: i=563$  (50%), 372 (33%), 282 (25%), 225 (20%), 169 (15%), 113 (10%) using each of the two sampling methods.
2. Extract envenoming bite incidence in the selected GN divisions from the gold standard incidence map.
3. Draw a random number of envenoming bites (ie, observations) at each selected location using the binomial distribution for 150 trials and envenoming bite probability extracted from step 2. Note that the NSS surveyed 40 households (ie, second-stage sampling units) in each GN division, which included approximately 150 people from a GN division (ie, 3–4 people per household).
4. Extract the values of explanatory variables in equation 1 for the selected GN divisions from the respective raster maps.



### Geostatistical modelling and predictive mapping

The simulated envenoming bites were modelled using binomial geostatistical models (equation 1). Parameter estimates of the models were obtained by the Monte Carlo maximum likelihood method.<sup>14</sup> A predictive grid surface of 2 km by 2 km was developed for the entire country and plug-in geographical predictions were obtained for mapping the results. The PrevMap package implemented in V.1.5.4 of the R programming language was used to fit the geostatistical models.<sup>15</sup>

### Calculating APV, sensitivity and specificity

Both the original predictive map (ie, the gold standard) based on the NSS and simulated maps used the same 2 km by 2 km predictive grid surface. The original predictive map and simulated maps were compared at each grid point and the APV (an estimate of the mean squared prediction error) was calculated as follows;

$$APV = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$$

At location  $x_i$ ,  $i=1,2,\dots,n$ ,  $y_i$  is the actual envenoming bite incidence and  $\hat{y}_i$  is the envenoming bite incidence predicted by the simulations.

Subsequently, we calculated for each simulated map the sensitivity and specificity of detecting high-risk MOH divisions.

We replicated the above simulation 100 times for the classical survey and ICP designs for each fraction of the original NSS sample size (ie, 100%, 50%, 33%, 25%, 20% and 10%) to obtain the mean and SD of APV, and the sensitivity and specificity of detecting MOH areas with high and low envenoming bite incidences.

### Distance calculation

The data collection of the NSS was carried out at the provincial level, with a single group completing data collection in one province before moving on to the other provinces. Because the ICP design imposes a minimum distance between sampling units, the travel distances involved for data collection within a province are unlikely to reduce in the same proportion as the reduction in sample size. Therefore, we estimated the travel distance for data collection within each province with ICP designs and compared it with the sampling strategy used in the NSS.

The distance travelled within a province was computed by randomly selecting a sampled GN division (through ICP or the original sampling strategy used in the NSS) from a given province and then using the travelling salesman algorithm in the TSP package within the R programming language.<sup>16</sup> This algorithm estimates the shortest route that covers all the selected GN divisions without any repetitions within a province. The sample selection and travel distance calculation procedures were then repeated 100 times to obtain mean travel distances for each ICP design from a random starting point.

The above computation requires a matrix of distances between each pair of GN divisions within a province. Given the cost constraints involved in estimating actual road distances for high-dimensional matrices, the distances were first approximated as the Euclidean (straight-line) distances between the centroids of a selected pair of clusters and then adjusted to better reflect road distances based on a district-level correction factor. The adjustment was based on the actual road distances from a randomly selected GN division to each of the remaining GN divisions within a district. This was more parsimonious than computing road distances between every pair of GN divisions within a district. For instance, in a district with 500 GN divisions, this approach required only 500 distances to be computed rather than  $500^2$ ). The correction factor for a district was calculated as the average of the ratios between the road distance and Euclidean distance from the randomly selected GN division to the remaining GN divisions within the district. The actual road distances used for the calculation were obtained using the Distance Matrix application programming interface (API).

### Survey time calculation

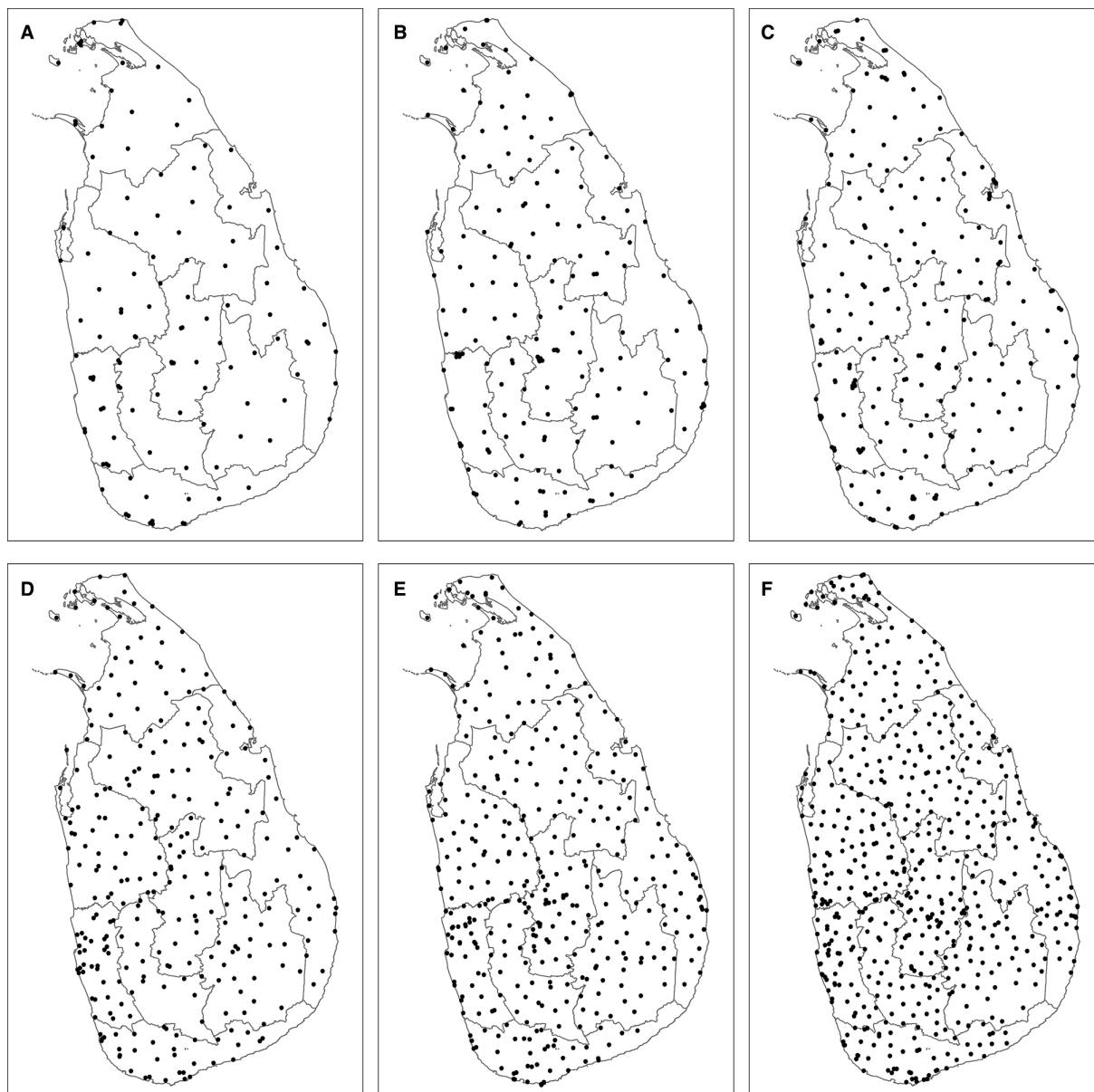
In the NSS, the time taken to complete the survey in a household depended on whether a member of that household had been a victim of a snakebite or not. In households that did not have victims of snakebite, the average time taken to complete the survey was 10 min, compared with an average of 40 min in households that did. Using these assumptions, we calculated the time taken to complete the survey for the simulated envenoming bites as described above. We replicated the above simulation 100 times for the classical survey design and ICP for each fraction of the original NSS sample size (ie, 100%, 50%, 33%, 25%, 20% and 10%) to obtain the mean and SD for time taken to complete the survey.

## RESULTS

Online supplemental figure 2 and figure 1 respectively, demonstrate an instance of a sampled location from classical survey and ICP designs at different sampling fractions.

### Prediction variance

The APV of the 100 replicates of simulations are shown in figure 2. The APV of classical survey designs showed relatively higher deviation from the original sample size compared with ICP. The 95% confidence limits of the original sample size overlapped with the APV of 33% and 50% sampling fractions in classical survey designs, and with the APV of 25%, 33% and 50% sampling fractions in ICP designs. This indicated that a similar predictive efficiency can be achieved with a sample fraction of 33% in a classical survey design and 25% in an ICP, which in turn represents a 25% reduction in the required sample size for an ICP design by comparison with a classical survey design. For all the sampling fractions considered, the 95% confidence limits of the variance ratio between



**Figure 1** Grama Niladhari divisions selected by inhibitory geostatistical design with close pairs (ICP). Panel a) 10%, b) 15%, c) 20%, d) 25%, e) 33% and f) 50% of the original sample size in the National Snakebite Survey.

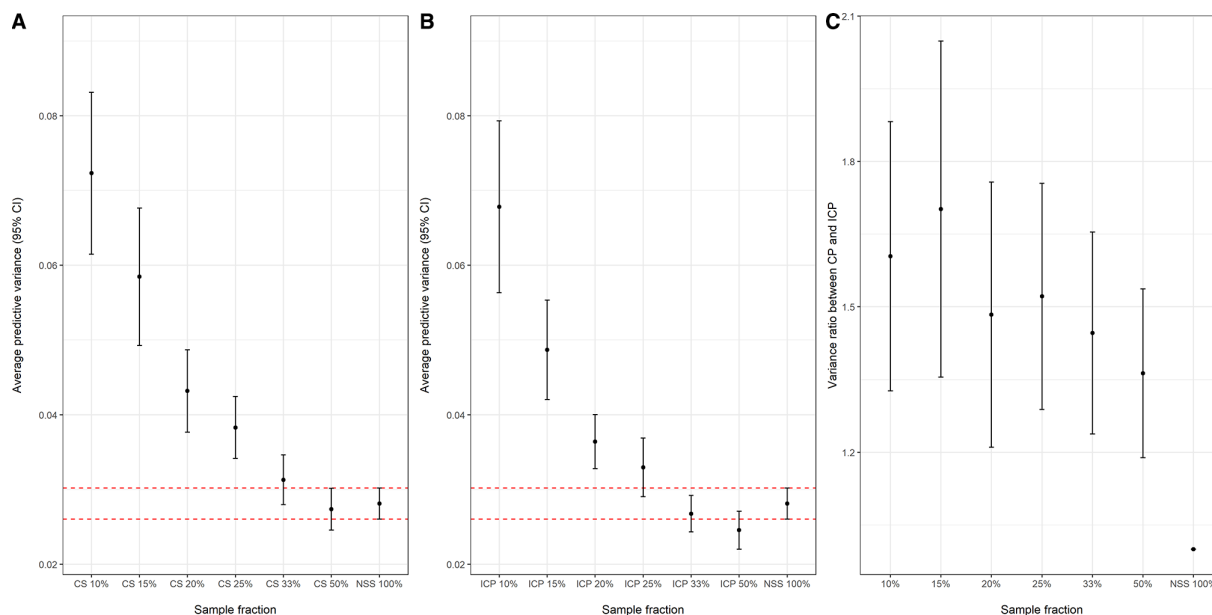
classical and ICP designs were substantially bigger than one, indicating the uniformly greater precision achieved by ICP designs relative to their classical design counterparts.

### Sensitivity and specificity

Table 1 shows the sensitivity and specificity of the 100 replicates of simulations. The original NSS sampling showed 81.1% (95% CI 79.5% to 82.6%) sensitivity and 91.5% (95% CI 89.9% to 93.0%) specificity in detecting high-risk MOH areas for snakebite envenoming. Classical survey designs showed a lower sensitivity and a higher specificity than ICP designs with the same sample size. All the ICP designs showed more than 80% sensitivity and 75% specificity. Both statistics were more than 80% when the sampling fractions were more than 25% in ICP designs.

### Distance

Figure 3 plots the average percentage change in travel distance from reducing the sample size to different fractions of the original for both classical survey and ICP designs using the travel distance from the NSS as the base. Reducing the number of clusters leads to a substantial reduction in average travelling distance. For instance, reducing the number of clusters to 25% of the original sample resulted in a 44% reduction in average travel distance. The reductions in travelling time with the ICP design were slightly smaller than the reductions with classical survey design, as ICP imposes a minimum distance between clusters to ensure geographical representativeness. However, in all cases the 95% confidence limits contained one, indicating that the ICP designs did not increase travel distance by comparison with classical designs.



**Figure 2** Results of the simulation. The average predictive variance (APV) of 100 replications of simulations a) classical design, b) inhibitory close pairs (ICP) design and c) variance ratio between classical design and ICP design with at 10%, 15%, 20%, 25%, 33%, 50% and 100% of the original sample size in the National Snakebite Survey (NSS). Red dashed lines indicate 95% CI of the APV of the NSS sampling in panel a) and b), and variance ratio of one in panel c). CP, close pair; ICP, inhibitory close pair.

### Time taken to complete the survey

Both classical design and ICP design showed a reduction in the time taken to complete the survey at each fraction of the NSS sample size (figure 4). The 95% confidence limits were very narrow due to the high number of participants without envenoming bites. The lower 95% confidence limits of the ratio between time taken to complete the survey in classical and ICP designs were greater than

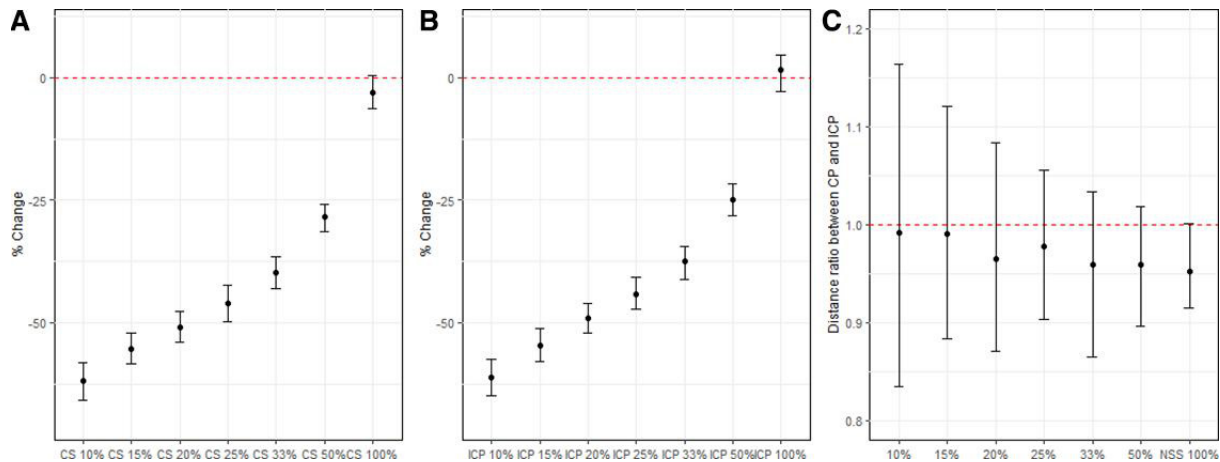
one at sampling fractions less than 50%, indicating that the ICP design can save time by comparison with the classical design.

### DISCUSSION

In this study, we compared the classical survey design with ICP in order to assess the utility of the ICP design in estimating snakebite envenoming incidence. Our results showed, first, that a similar predictive efficiency can be obtained at lower sample size with both the classical survey and ICP designs. This is a consequence of the geographical correlation in the incidence surface, whereby, *provided* that the data are analysed using geostatistical methods, the observed incidence of snakebite at any one location is predictive of the underlying risk in neighbouring locations. The classical survey design could yield almost the same predictive efficiency at one-third of the original sample size, indicating the NSS has oversampled the country's population. By comparison, the ICP design was able to produce the same results at one-fourth of the original sample size of the NSS, highlighting an advantage of adopting ICP over classical survey design in estimating envenoming bite incidence in the country. Geostatistical methods, including ICP designs, are designed for efficient disease mapping over a region of interest and so comparatively smaller sample sizes are required to be compared with non-geographical, conventional survey sampling methods.<sup>9</sup> Further, ICP designs showed lower predictive variance than classical designs at all the sampling fractions, indicating ICP designs were able to produce more reliable predictions than conventional survey sampling designs. Therefore,

**Table 1** Sensitivity and specificity of 100 replicates of simulations in detecting high-risk medical officer of Health areas with classical and inhibitory geostatistical design with close pairs (ICP) design

Sample size	Sensitivity (95% CI)	Specificity (95% CI)
Classical design		
563 (50%)	85.5 (83.4 to 87.6)	86.5 (84.2 to 88.7)
372 (33%)	83.2 (80.6 to 85.8)	87.8 (85.3 to 90.3)
282 (25%)	80.7 (77.9 to 83.5)	85.2 (82.5 to 88.0)
225 (20%)	80.7 (77.4 to 84.1)	85.4 (83.0 to 87.7)
169 (15%)	78.9 (74.8 to 83.1)	80.7 (77.8 to 83.6)
113 (10%)	73.5 (68.6 to 78.5)	79.5 (75.8 to 83.1)
ICP design		
563 (50%)	86.5 (84.8 to 88.3)	81.3 (78.8 to 83.7)
372 (33%)	86.8 (84.9 to 88.6)	80.9 (77.9 to 84.0)
282 (25%)	87.0 (85.1 to 88.9)	80.9 (81.3 to 84.4)
225 (20%)	84.7 (82.1 to 87.2)	81.7 (78.3 to 85.0)
169 (15%)	83.2 (80.3 to 86.1)	77.6 (74.1 to 81.1)
113 (10%)	83.6 (80.9 to 86.3)	75.0 (71.2 to 78.8)



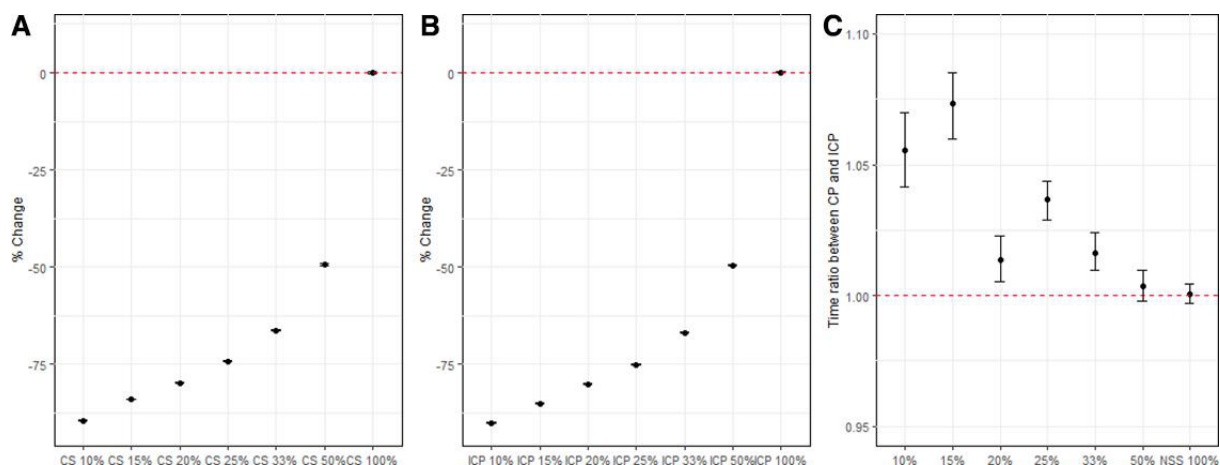
**Figure 3** Change in mean travel distance. The average change in mean travel distance of 100 replications of simulations a) classical design, b) inhibitory close pairs (ICP) design and c) ratio between travel distance in classical and ICP designs at 10%, 15%, 20%, 25%, 33%, 50% and 100% of the original sample size in the National Snakebite Survey (NSS). CP, close pair; ICP, inhibitory close pair.

resource requirements can be reduced by adopting ICP designs and geostatistical analysis in epidemiological surveys.

The smallest healthcare administrative divisions in Sri Lanka are the MOH areas. The NSS estimated the national envenoming bite incidence in Sri Lanka as 151 per 100 000 people. We classified the MOH areas as high risk and low risk based on the national envenoming bite incidence (ie, gold standard). This was done to assess the sensitivity and specificity of ICP design in detecting high-risk MOH divisions. The ICP designs showed more than 80% sensitivity and specificity even at 25% of the original NSS sample size, showing their practical utility in snakebite research. Therefore, geostatistical methods, including ICP, can be considered a viable alternative for epidemiological surveys, particularly in low-resource settings.

According to the NSS, snakebites show a substantial geographical correlation in Sri Lanka.<sup>3</sup> Therefore,

samples obtained from close locations can be expected to provide essentially the same information. Sampling methods based on random sampling can easily select sampling units close to each other which could be disadvantageous in these circumstances.<sup>17</sup> Further, conventional survey sampling methods may be inefficient for predicting local risk, as the estimates can be associated with high standard errors, especially in areas with lower sampling efforts. The adoption of geographically regulated designs such as ICP designs, in conjunction with geostatistical methods of analysis, has been proposed to lower the mean square prediction errors and reduce uncertainty in predictions.<sup>10</sup> The inclusion of close pairs in the ICP design enables estimation of geographical and non-geographical variation of the underlying geographically varying risk surface (ie, both spatial and non-spatial random effects in the covariance structure), should these be needed, in addition to binomial sampling variation.



**Figure 4** Change in mean survey time. The average change in mean time taken to complete the survey of 100 replications of simulations a) classical design, b) inhibitory close pairs (ICP) and c) ratio between time taken to complete the survey in classical and ICP designs fraction of the NSS sample size design corresponding to 10%, 15%, 20%, 25%, 33%, 50% and 100% of the original sample size in the National Snakebite Survey (NSS). CP, close pair; ICP, inhibitory close pair.



In an ICP design, primary sampling units maintain a minimum distance between any two units. Consequently, a reduction in travel time to complete the survey does not necessarily follow a reduction in sample size. Therefore, we compared the total travel distance in ICP and classical survey designs at different sampling fractions of original sample size of the NSS survey. Our analysis showed that the ICP designs did not increase the travel distance compared with classical survey designs of the same size. However, since the required sample size is smaller in an ICP design, this will reduce the travel-related costs incurred to complete the survey. When we evaluated the time taken to complete the survey, an ICP design took less time than a conventional survey sampling design at the same sampling fraction.

However, there are challenges to adopting geostatistical methods. First, an ICP design strictly requires prior geolocations of all units in the sampling frame. In the absence of these, a pragmatic solution is the following. A provisional ICP design is constructed in which every point in the design region is considered as a potential sampling location. The field team is then asked to identify the available sampling location closest to each point in the provisional design. Depending on the geometry of the potential sampling locations, this may or may not result in a good approximation to an ICP design. A second limitation is that expertise in geostatistical analysis methods is needed to get the maximum benefit from georeferenced data. This highlights the importance of knowledge-sharing and capacity building across the scientific community, especially in low-resource settings.<sup>17</sup>

## CONCLUSIONS

In conclusion, our study showed that snakebite envenoming incidence can be estimated with similar precision to classical survey designs, such as, multistage cluster sampling by adopting an ICP design with smaller sample size and travel time. This would substantially save human and financial resources and time taken to conduct surveys. These designs do require georeferenced data and expertise in geostatistical methods of analysis, but this can and should be addressed by capacity-building activities in the relevant countries assisted by the development of more user-friendly software interfaces. ICP designs should be considered for epidemiological studies in low-resource settings.

### Author affiliations

<sup>1</sup>Health Data Science Unit, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka

<sup>2</sup>Department of Decision Sciences, Faculty of Business, University of Moratuwa, Moratuwa, Sri Lanka

<sup>3</sup>Department of Public Health, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka

<sup>4</sup>Department of Medicine, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka

<sup>5</sup>Centre for Health Informatics, Computing and Statistics, Lancaster University, Lancaster, UK

**Twitter** Dileepa Senajith Ediriweera @edileepa

**Acknowledgements** The authors wish to acknowledge the research team who designed and undertook the National Snakebite Survey.

**Contributors** DSE, TdS, AK, HJdS and PD designed the Study. DSE and TdS designed the analytical strategy, conducted the analysis and interpret the results. DSE and TdS drafted the manuscript and AK, HJdS and PD critically revised the article. DSE is responsible for the overall content as guarantor.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not applicable.

**Ethics approval** Ethical approval for the National Snakebite Survey was obtained from the Ethics Review Committee of the Faculty of Medicine, University of Kelaniya. All interviews were conducted after obtaining informed written consent. Permission for conducting the survey was obtained from district-level and divisional-level public administrators. This study used only the secondary data.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** The data underlying this article will be shared on reasonable request to the corresponding author.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

- Chippaux J-P. Snakebite envenomation turns again into a neglected tropical disease! *J Venom Anim Toxins Incl Trop Dis* 2017;23:38.
- Kasturiratne A, Wickremasinghe AR, de Silva N, *et al*. The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths. *PLoS Med* 2008;5:e218–1604.
- Ediriweera DS, Kasturiratne A, Pathmeswaran A, *et al*. Mapping the Risk of Snakebite in Sri Lanka - A National Survey with Geospatial Analysis. *PLoS Negl Trop Dis* 2016;10:e0004813.
- Patikorn C, Ismail AK, Abidin SAZ, *et al*. Situation of snakebite, antivenom market and access to antivenoms in ASEAN countries. *BMJ Glob Health* 2022;7:e007639 <https://gh.bmj.com/content/7/3/e007639>
- Schneider MC, Min K-duk, Hamrick PN, *et al*. Overview of snakebite in Brazil: possible drivers and a tool for risk mapping. *PLoS Negl Trop Dis* 2021;15:e0009044–18.
- Hossain J, Biswas A, Rahman F. Snakebite epidemiology in Bangladesh—A national community based health and injury survey. *Health* 2016;8:479–86 <http://www.scirp.org/journal/PaperInformation.aspx?PaperID=64732>
- Jayawardana S, Arambepola C, Chang T, *et al*. Prevalence, vulnerability and epidemiological characteristics of snakebite in agricultural settings in rural Sri Lanka: a population-based study from South Asia. *PLoS One* 2020;15:e0243991 <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0243991>
- Ediriweera DS, Kasturiratne A, Pathmeswaran A, *et al*. Adjusting for spatial variation when assessing individual-level risk: a case-study in the epidemiology of snake-bite in Sri Lanka. *PLoS One* 2019;14:e0223021.
- Chipeta M, Terlouw D, Phiri K, *et al*. Inhibitory geostatistical designs for spatial prediction taking account of uncertain covariance structure. *Environmetrics* 2017;28:e2425–11.



- 10 Fronterre C, Amoah B, Giorgi E, *et al.* Design and analysis of elimination surveys for neglected tropical diseases. *J Infect Dis* 2020;221:S554–60.
- 11 Kabaghe AN, Chipeta MG, McCann RS, *et al.* Adaptive geostatistical sampling enables efficient identification of malaria hotspots in repeated cross-sectional surveys in rural Malawi. *PLoS One* 2017;12:e0172266–14.
- 12 Midzi N, Montresor A, Mutsaka-Makuvaza MJ, *et al.* Elimination of STh morbidity in Zimbabwe: results of 6 years of deworming intervention for school-age children. *PLoS Negl Trop Dis* 2020;14:e0008739 [https://journals.plos.org/plosntds/article?id=](https://journals.plos.org/plosntds/article?id=10.1371/journal.pntds.1008739)
- 13 Johnson O, Fronterre C, Amoah B, *et al.* Model-Based Geostatistical methods enable efficient design and analysis of prevalence surveys for Soil-Transmitted helminth infection and other neglected tropical diseases. *Clin Infect Dis* 2021;72:S172–9.
- 14 Diggle PJ, Giorgi E. Model-based Geostatistics for Global Public Health. In: *Model. Geostatistics Glob. public heal.* 1st ed.. New York: Chapman and Hall/CRC, 2019..
- 15 Giorgi E, Diggle PJ. PrevMap: an R package for prevalence mapping.. *J Stat Softw* 2015;78:1–29.
- 16 Hahsler M, Hornik K. TSP—Infrastructure for the traveling Salesperson problem. *J Stat Softw* 2008;23:1–21 <https://www.jstatsoft.org/index.php/jss/article/view/v023i02>
- 17 Diggle PJ, Amoah B, Fronterre C, *et al.* Rethinking neglected tropical disease prevalence survey design and analysis: a geospatial paradigm. *Trans R Soc Trop Med Hyg* 2021;115:208–10.