

## RESEARCH ARTICLE

# Adjusting for spatial variation when assessing individual-level risk: A case-study in the epidemiology of snake-bite in Sri Lanka

Dileepa Senajith Ediriweera<sup>1\*</sup>, Anuradhani Kasthuriratne<sup>2</sup>,  
 Arunasalam Pathmeswaran<sup>2</sup>, Nipul Kithsiri Gunawardene<sup>3</sup>, Shaluka  
 Francis Jayamanne<sup>4</sup>, Kris Murray<sup>5,6‡</sup>, Takuya Iwamura<sup>7‡</sup>, David Griffith Lalloo<sup>8</sup>,  
 Hithanadura Janaka de Silva<sup>4</sup>, Peter John Diggle<sup>9</sup>

**1** Centre for Health Informatics, Biostatistics and Epidemiology, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka, **2** Department of Public Health, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka, **3** Department of Parasitology, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka, **4** Department of Medicine, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka, **5** Department of Infectious Disease Epidemiology, School of Public Health, Faculty of Medicine, Imperial College London, St Mary's Campus, London, United Kingdom, **6** Grantham Institute—Climate Change and the Environment, Imperial College London, South Kensington, London, United Kingdom, **7** Faculty of Life Sciences, School of Zoology, Tel Aviv University, Tel Aviv, Israel, **8** Liverpool School of Tropical Medicine, Liverpool, United Kingdom, **9** CHICAS, Lancaster University Medical School, Lancaster, United Kingdom

\* These authors contributed equally to this work.

‡ These authors also contributed equally to this work.

\* dileepa@kln.ac.lk



## OPEN ACCESS

**Citation:** Ediriweera DS, Kasthuriratne A, Pathmeswaran A, Gunawardene NK, Jayamanne SF, Murray K, et al. (2019) Adjusting for spatial variation when assessing individual-level risk: A case-study in the epidemiology of snake-bite in Sri Lanka. PLoS ONE 14(10): e0223021. <https://doi.org/10.1371/journal.pone.0223021>

**Editor:** Mohammad Ali, Johns Hopkins Bloomberg School of Public Health, UNITED STATES

**Received:** March 28, 2019

**Accepted:** September 11, 2019

**Published:** October 3, 2019

**Copyright:** © 2019 Ediriweera et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Individual level data used for the present analysis and epidemiological snakebite data used to estimate snakebite incidence are uploaded as Supporting Information. The following third party data can be obtained from Census and Statistics Department of Sri Lanka: GN/Cluster level - population; percentage of males; percentage of agricultural workers; population mean age; percentage of people who had studied up to or above G.C.E. Advanced Level Examination; percentage of the major ethnic group. District level

## Abstract

### Background

Health outcomes and causality are usually assessed with individual level sociodemographic variables. Studies that consider only individual-level variables can suffer from residual confounding. This can result in individual variables that are unrelated to risk behaving as proxies for uncaptured information. There is a scarcity of literature on risk factors for snakebite. In this study, we evaluate the individual-level risk factors of snakebite in Sri Lanka and highlight the impact of spatial confounding on determining the individual-level risk effects.

### Methods

Data was obtained from the National Snakebite Survey of Sri Lanka. This was an Island-wide community-based survey. The survey sampled 165,665 individuals from all 25 districts of the country. We used generalized linear models to identify individual-level factors that contribute to an individual's risk of experiencing a snakebite event. We fitted separate models to assess risk factors with and without considering spatial variation in snakebite incidence in the country.

### Results

Both spatially adjusted and non-adjusted models revealed that middle-aged people, males, field workers and individuals with low level of education have high risk of snakebites. The

- mean income. The following data can be downloaded from NDVI - International Research Institute for Climate and Society, Earth Institute, Columbia University (<http://iridl.ldeo.columbia.edu/>): Elevation and land cover - DIVA-GIS (<http://www.diva-gis.org/gdata>). The authors did not have special access privileges.

**Funding:** This work was supported by the Medical Research Council [MR/P024513/1]. The grant was obtained by AP, KM, TI, DGL, HJdeS and PJD. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

model without spatial adjustment showed an interaction between ethnicity and income levels. When the model included a spatial adjustment for the overall snakebite incidence, this interaction disappeared and income level appeared as an independent risk factor. Both models showed similar effect sizes for gender and age. HEmployment and education showed lower effect sizes in the spatially adjusted model.

## Conclusions

Both individual-level characteristics and local snakebite incidence are important to determine snakebite risk at a given location. Individual level variables could act as proxies for underlying residual spatial variation when environmental information is not considered. This can lead to misinterpretation of risk factors and biased estimates of effect sizes. Both individual-level and environmental variables are important in assessing causality in epidemiological studies.

## Introduction

Both individual characteristics and environmental factors are important to individual health and disease outcomes [1] [2]. Environmental factors can be biological, chemical, physical and social, any of which can lead to disease [3]. Environmental factors typically show geographical or seasonal variation. Studies that consider only the individual-level variables when evaluating disease causation consequently suffer from residual confounding [1]. This can result in individual variables that are unrelated to risk behaving as proxies for uncaptured environmental information, leading to misinterpretation [4]. Environmental factors have also been shown to have interactions with individual-level variables [5], suggesting a need to consider both individual-level and environmental data in epidemiological studies aimed at understanding disease risk [1].

Assessing both individual and environmental factors help to avoid overestimation of individual risk-factor effects [6]. Environmental variables can only explain the average risk at a given location, whereas individual-level variables capture heterogeneity of risk at a location due to differences in behavior and individual characteristics. Estimates obtained without adjusting for spatial variation in risk could be confounded by location, hence masking individual-level causal effects. Estimating the effects of individual-level factors whilst adjusting for geographical variation in epidemiological studies is analogous to the long-established practice of blocking to control for extraneous variation in a randomized field experiment.

Sri Lanka reports 80 000 snakebites annually, with substantial geographical and seasonal variation in incidence [7] [8]. There are more than 100 snake species in the country, amongst which six are considered as medically important (*Naja naja*, *Bungarus ceylonicus*, *Bungarus caeruleus*, *Daboia russelii*, *Echis carinatus* and *Hypnale hypnale*). *Daboia russelii* is the largest venomous snake in Sri Lanka and is widely distributed in the country. *Naja naja* is the largest elapid in Sri Lanka and shows a wide distribution in the country. *Bungarus ceylonicus* and *Bungarus caeruleus* have the most potent venom and can be found in and around human habitats in the dry zone of the country. *Hypnale hypnale* is considered to be a moderately venomous snake and is responsible for 35% to 45% of all human bites. These snakes can commonly be found in human habitats including rubber, tea, coconut and cocoa plantations. *Echis carinatus* is only responsible for 1% to 2% of bites and is confined to the arid dry zones of the country. *Naja naja*, *Bungarus ceylonicus* and *Bungarus caeruleus* cause neurotoxic effects and *Daboia*

*russelii*, *Echis carinatus* and *Hypnale hypnale* mainly show haematological abnormalities [7] [9] [10].

Previous local hospital data have shown that the majority of snakebite victims are middle-aged males [11] [12] [13]. Similar findings have been reported in other countries [14] [15] [16] [17]. There is a scarcity of epidemiological data on risk factors for snakebite, as most snakebite studies are still conducted in hospital settings [18]. In one study, Caiaffa et al investigated the risk factors for snakebite by using hospital patients who do not have a history of snakebites as controls. They reported that agricultural workers and people living in rural areas are high-risk groups for snakebite [19]. Health-seeking behaviour following snakebite shows considerable variation, and hospitals do not provide healthcare service to all snakebite victims [10]. Therefore, hospital data on snakebite are potentially biased, limiting their suitability for assessing risk factors.

There is a paucity of literature on individual-level risk factors for snakebite and this has been highlighted as an area that requires further research [20]. Previous epidemiological studies have presented the collective demographic characteristics of snakebite victims, but have not evaluated individual-level risk factors [7] [14] [15] [16] [17]. Although both geographical and seasonal variation of snakebites have been studied in Sri Lanka, individual-level risk factors have not been evaluated [7] [8]. In this study, we evaluate the individual-level risk factors of snakebite risk in a community-based national representative survey in Sri Lanka. We considered snakebite incidence at the sampled location as an environmental variable, to highlight the effect of spatial confounding on determining risk factors, and the importance of considering both individual and environmental variables in assessing causality in epidemiological studies.

## Materials and methods

### Epidemiological data

Epidemiological snakebite data were collected by conducting an island-wide community-based "National Snakebite Survey". The survey was designed to sample 1% of the Sri Lankan population. This was a cross sectional study and multi-stage cluster sampling technique was used to sample population. A multi-stage cluster sampling technique was used to sample 0.8% of the Sri Lankan population. Sri Lanka has nine provinces, with between two and five districts per province. Each district is further subdivided into a national total of 14 022 Grama Niladari divisions. Grama Niladari divisions are the smallest administrative divisions in Sri Lanka. Each Grama Niladari division was considered as a potential cluster for the survey. In each province, 125 clusters were proportionally allocated among the districts according to their population sizes, and the allocated number of clusters randomly selected from the complete list of Grama Niladari divisions in each district. Within each cluster, 40 consecutive households from the electoral register were sampled, with the first household randomly selected. The survey included all the permanent members in the sampled households. The interviewer used a questionnaire to collect data from a responsible adult household member. All snakebite events that occurred during the preceding year were recorded, along with individual demographic data on the household members and the geo-locations of the household. The survey was conducted from August 2012 to June 2013. It covered all the nine provinces and all the 25 districts in the country. Details of the "National Snake Survey" has been published in our previous publication (i.e. Ediriweera, Dileepa Senajith, et al. "Mapping the risk of snakebite in Sri Lanka—a national survey with geospatial analysis." PLoS neglected tropical diseases 10.7 (2016): e0004813). A location map of Sri Lanka and the administrative divisions are shown in [S1](#) and [S2](#) Figs.

This study was approved by the Ethics Review Committee of the Faculty of Medicine, University of Kelaniya (Ref: P 06/01/2012). All interviews were conducted after obtaining

informed written consent. Approval from District and Divisional level public administrators was obtained for conducting the community-based survey. No animals were used in the study.

### Snakebite incidence data

Sri Lanka does not possess cluster-level information on snakebite. Therefore, we extracted cluster-level snakebite incidence from our previously published snakebite incidence map of Sri Lanka [7]. These incidence estimates were obtained using a geostatistical model that considered population density, height above sea level, occupation distribution and climatic zone. Snakebite incidence at the centroid of each cluster was attached to each sampled individual, which implicitly assumes that the incidence does not vary within clusters. The median area of a cluster is approximately 2.0 (IQR: 0.9–4.2) km<sup>2</sup>. Details of the geostatistical model is given in [S1 Appendix](#).

### Statistical methods

For exploratory analysis, generalized additive models were used to identify non-linear associations in explanatory variables. Generalized linear models were then used to model the probability of snakebite at individual level. All models were fitted using the R programming language version 3.4.2 [21]. We considered age, sex, ethnicity, religion, education, employment and income as exposure variables. We treated age as a continuous variable and the remainder as categorical variables. Categorical variables were collapsed when we found no differences between levels, testing at the conventional 5% level. Multicollinearity between exposure variables were assessed during the model building using variance inflation factor to avoid multicollinearity in the fitted models. The survey collected data from 165 665 individuals (about 0.8% of the Sri Lankan population). After removal of records with missing data, 158 066 records were available for the analysis.

Separate models were fitted to investigate individual-level risk factors with and without adjusting for spatial variation in snakebite incidence, so as to assess the impact of the spatial adjustment on the interpretation of individual-level risk-factors; we refer to these models as spatially adjusted and spatially non-adjusted models, respectively. Log likelihood ratio test and z statistic were considered to select variables in the spatially non-adjusted model and spatially adjusted models respectively.

The spatially adjusted model is a generalized linear mixed model that allows for extra-binomial and spatially correlated variation in risk. Snakebite incidences obtained from geostatistical models are associated with standard errors and conclusions drawn from a single sample of a predictive distribution of snakebite incidence could be error bound. Therefore, we used multiple imputation of these estimates when fitting the spatially adjusted model that included snakebite incidence as an exposure variable. This was done by adopting a two-stage algorithm. Stage 1 generates 10,000 samples from the predictive distribution of the overall risk-map derived from a geostatistical model [7] fitted using the R package *PrevMap* [22]. This gave 10,000 imputed snakebite incidences from the predictive distribution at each location. Stage 2 passes each sample to a generalized linear model to account for individual-level variation in risk within locations. Stage 2 assumes that the responses from different individuals are conditionally independent given the spatial risk surface (i.e. the covariates including the mapped risk surface are sufficient to explain the spatial variation in risk to an individual). Final estimates and standard errors of regression parameters for individual risk-factors were then calculated using a standard result in probability theory (formula 1 and 2):

$$E(\hat{\beta}) = E(E(\hat{\beta}|U)) \quad (1)$$

$$\text{Var}(\hat{\beta}) = \text{Var}(E(\hat{\beta}|U)) + E(\text{Var}(\hat{\beta}|U)) \quad (2)$$

where  $\hat{\beta}$  is a parameter estimate,  $U$  an imputed value of a risk factor and a vertical bar denotes conditioning.

To apply this result, let  $U_i$  denote the  $i^{\text{th}}$  imputed map,  $\hat{\beta}_i$  the corresponding regression parameter estimate and  $v_i$  the variance of  $\hat{\beta}_i$  as reported by the generalized linear model software. Then, the final estimate,  $\hat{\beta}_i$  is the sample mean of the 10,000  $\hat{\beta}_i$  and its variance is the sum of two components: the sample variance of the  $\hat{\beta}_i$  and the sample mean of the  $v_i$ .

Estimated cluster level random effects of the spatially adjusted model were then used to assess for the presence of any residual, unexplained spatial correlation. This was done by calculating the empirical variograms of the predicted random effects and for 1000 random permutations of these. From these random permutations, we obtained pointwise 95% tolerance limits under the assumption of spatially uncorrelated random effects. The variogram of the random effects was not contained within the envelope of the tolerance limits indicating a possible spatial correlation (S3 Fig). We therefore estimated the covariance parameters of spatial correlation using PrevMap package [22]. The estimated variance of the Gaussian process is much smaller than the estimated variance of the nugget effect (0.003 and 0.226 respectively). Therefore, we concluded that the spatial component of the residual, unexplained cluster-level spatial variation is negligible (S1 Table) and there is no evidence of lack of fit on the spatially adjusted model [23].

## Results

### Demography

The survey sampled 165,665 individuals (0.8% of the population of Sri Lanka) living in 44136 households from 1118 clusters. The median and interquartile range of participants' age were 35 and 20–52, respectively. 50.1% of participants were male. The majority of participants' ethnicity was Sinhalese and the majority religion was Buddhism. Nearly 25.9% had been educated up to G.C.E. Advanced Level or above, and 68.4% were field workers including farmers. The survey reported 695 snakebite events during the year preceding interview. Demographic characteristics of survey participants are given in Table 1.

### Individual-level variable analysis from spatially non-adjusted model

In the generalized linear model without adjusting for the local snakebite incidence, age, sex, ethnicity, education, employment and income were associated with snakebite at individual level (Table 2). Age showed a non-linear association with snakebites, males showed higher risk compared to females (Odds ratio = 1.57, 95% CI: 1.34–1.83), and those who have studied below G.C.E. Advanced Level have higher risk than the rest (Odds ratio = 1.55, 95% CI: 1.24–1.90). Employment (Field workers vs non-field workers, Odds ratio = 1.58, 95% CI: 1.30–1.92) appeared as an independent risk factor and there was a strong interaction between ethnicity and income levels.

For Sinhalese, snakebite risk increased along with income, whereas in other ethnic groups snakebite decreased as income increased (Fig 1). According to the fitted model, maximum snakebite risk was observed among the low-educated, field-working, Sinhalese, males aged 55 years, for whom the fitted probability of being bitten by a snake in a year was 0.012 (95% CI: 0.009–0.017).

**Table 1. Demographic characteristics of participants in the national snakebite study, Sri Lanka.**

Characteristics	Number (%) or Median (IQR)
Age (years)	35 (20–52)
< 15	28 663 (17.3%)
15–24	25 502 (15.4%)
25–34	26 449 (16.0%)
35–44	25 090 (15.1%)
45–54	23 508 (14.2%)
55–64	18 884 (11.4%)
> 64	17 559 (10.6%)
Sex	
Males	82 888 (50.1%)
Females	82 705 (49.9%)
Ethnicity	
Sinhalese	123 839 (74.8%)
Tamils	29 852 (18.0%)
Muslims	11 841 (7.1%)
Other	97 (0.1%)
Religion	
Buddhist	120 644 (72.8%)
Catholic/Christian	6 527 (3.9%)
Hindu	26 444 (16.0%)
Islam	11 985 (7.2%)
Other	22 (0.1%)
Education	
No schooling	5658 (0.1%)
Primary	33 535 (20.2%)
Secondary	77 403 (46.8%)
Advanced level	42 888 (25.8%)
Above Advanced level	6060 (0.1%)
Employment	
Field workers	113 310 (68.4%)
Others	52 345 (31.6%)
Monthly income (Sri Lankan rupees)	
<5000	23 570 (14.5%)
5000–10 000	28 533 (17.5%)
10 000–20 000	50 451 (31.0%)
20 000–35 000	50 733 (31.2%)
>35 000	9446 (5.8%)
Estimated snakebite incidence (per 100,000) at sampled locations	336 (216–468)

<https://doi.org/10.1371/journal.pone.0223021.t001>

### Individual-level variable analysis from spatially adjusted model

Individual risk factor analysis after adjusting for spatial variation of snakebite incidence (i.e. spatially adjusted model) showed that snakebite risk is associated with an individual's age, sex, employment, education level, income level and local snakebite incidence (Table 3).

Snakebites showed a non-linear association with age. Snakebite risk was higher in middle-aged people compared both to younger and to older (Fig 2). Males showed higher snakebite risk compared to females (Odds ratio = 1.57, 95% CI: 1.34–1.83) and field workers showed higher risk compared to non-field workers (Odds ratio = 1.43, 95% CI: 1.17–1.75). Individuals

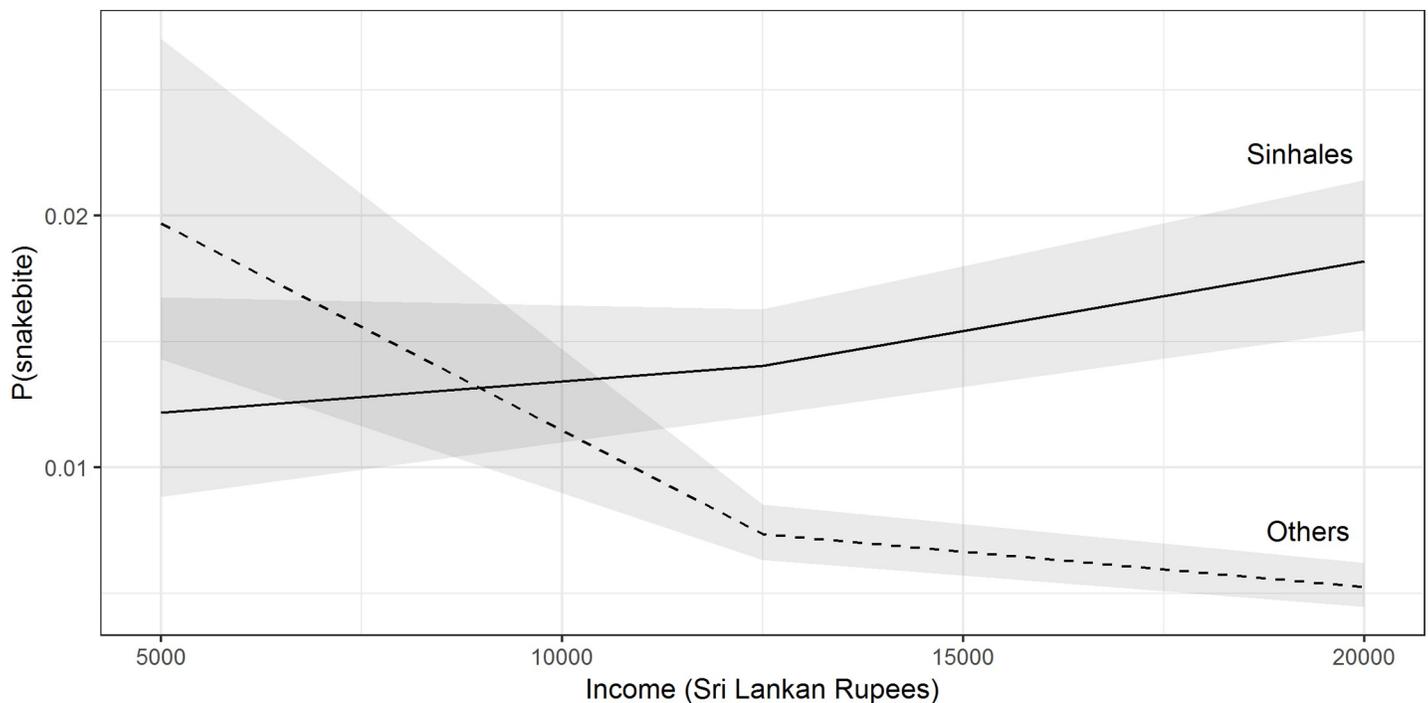
**Table 2. Fitted model for predicting a snakebite event without considering snakebite incidence.**

	Estimate	Std. Error	z value	P(> z )	Odds ratio (95% CI)
(Intercept)	-1.14e+01	4.30e-01	-2.64e+01	< 0.001	-
Age	2.38e-01	1.71e-02	1.39e+01	< 0.001	-
Age^2	-2.17e-03	1.75e-04	-1.23e+01	< 0.001	-
Sex (male)	4.50e-01	7.95e-02	5.66e-00	< 0.001	1.57 (1.34–1.83)
Ethnicity (Sinhalese)	-4.88e-01	1.98e-01	-2.47e-00	0.014	0.61 (0.42–0.90)
Education (advanced level or above)	-4.42e-01	1.01e-01	-4.35e-00	< 0.001	1.55 (1.24–1.90)
Employment (field workers)	4.56e-01	1.01e-01	4.54e-00	< 0.001	1.58 (1.30–1.92)
Income (5-20k)	-9.98e-01	1.99e-01	-5.02e-00	< 0.001	0.37 (0.25–0.54)
Income (>20k)	-1.33e+00	3.40e-01	-3.92e-00	< 0.001	0.26 (0.14–0.51)
Ethnicity (Sinhalese) : Income (5-20k)	1.14e+00	2.62e-01	4.36e-00	< 0.001	3.12 (1.87–5.22)
Ethnicity (Sinhalese) : Income (>20k)	1.74e+00	3.79e-01	4.59e-00	< 0.001	5.70 (2.71–11.96)

Variable units and categories: Age (years); Sex includes males and females; Employment includes field workers and others; Education includes below Advanced level and Advanced level or above; Income includes <5k, 5 – 20k and > 20k; Ethnicity includes Sinhalese and others

<https://doi.org/10.1371/journal.pone.0223021.t002>

who had not studied up to G.C.E. Advanced Level showed a higher risk for snakebite than those who had studied up to or beyond G.C.E. Advanced Level (Odds ratio = 1.33, 95% CI: 1.09–1.63). The lowest income group (i.e. less than Rs. 5000 per month) showed the highest snakebite risk compared to the middle-income group (i.e. between Rs. 5000 to 20 000; Odds ratio = 1.59, 95% CI: 1.30–2.00) and to the high-income group (i.e. more than Rs. 20,000 per



**Fig 1. Interaction plot between ethnicity and income.** Interaction plot between ethnicity and income for 50-year-old male farmers who had a low level of education. Predicted probability of snakebite among Sinhalese increased along with income rise (solid line) and the probability of snakebite decreased in Non-Sinhalese along higher income categories (dashed line). Grey colour bands represent the 95% confidence intervals.

<https://doi.org/10.1371/journal.pone.0223021.g001>

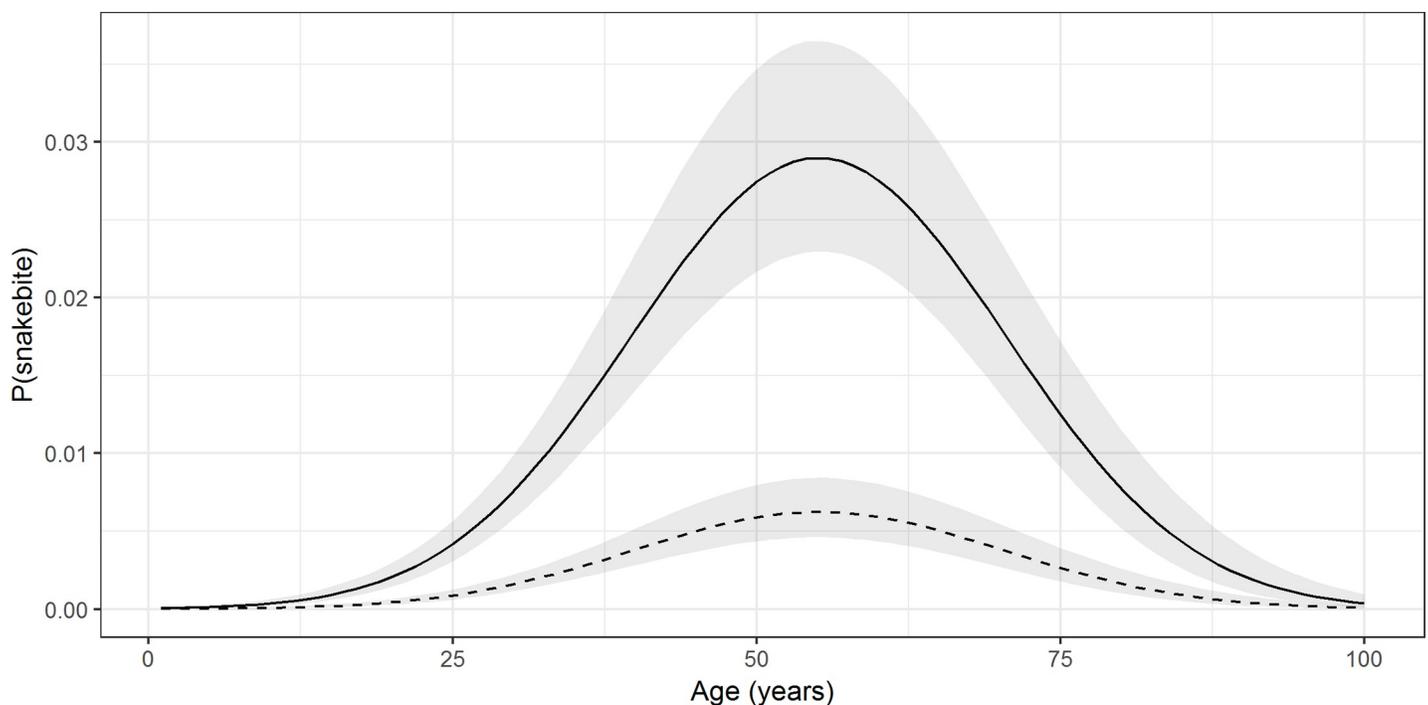
**Table 3. Fitted model for predicting a snakebite event after adjusting for snakebite incidence.**

	Estimate	Std. Error	z value	P(> z )	Odds ratio (95% CI)
(Intercept)	-1.24e+01	4.38e-01	-2.82e+01	< 0.001	-
Age	2.39e-01	1.72e-02	1.38e+01	< 0.001	-
Age^2	-2.16e-03	1.76e-04	-1.23e+01	< 0.001	-
Sex (male)	4.49e-01	8.01e-02	5.60e+00	< 0.001	1.57 (1.34–1.83)
Employment (field workers)	3.56e-01	1.03e-01	3.45e+00	< 0.001	1.43 (1.17–1.75)
Education (advanced level or above)	-2.87e-01	1.04e-01	-2.76e+00	0.003	1.33 (1.09–1.63)
Income (5–20k)	-4.68e-01	1.16e-01	-4.04e+00	< 0.001	1.59 (1.30–2.00)
Income (>20k)	-2.57e-01	1.24e-01	-2.07e+00	0.019	1.29 (1.01–1.65)
Incidence	2.48e-03	2.52e-04	9.86e+00	< 0.001	1.0024 (1.0019–1.00029)

Variable units and categories: Age (years); Sex includes males and females; Employment includes field workers and others; Education includes below Advanced level and Advanced level or above; Income includes <5k, 5 – 20k and > 20k; Incidence (estimated number of bites per 100,000 people in a given location)

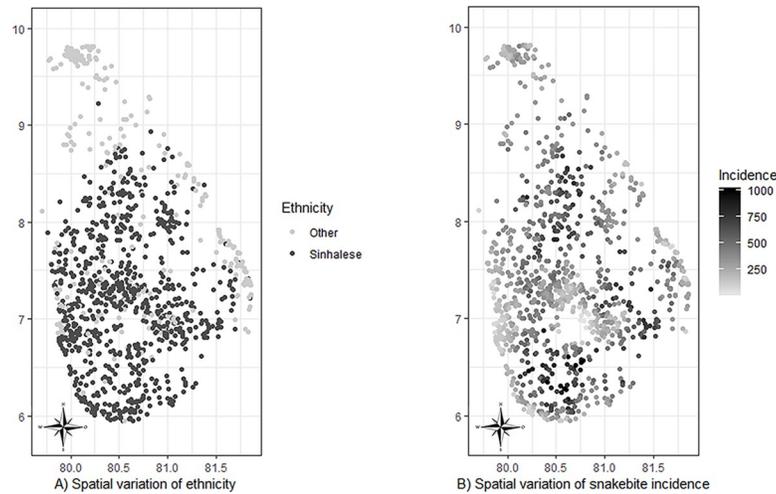
<https://doi.org/10.1371/journal.pone.0223021.t003>

month; Odds ratio = 1.29, 95% CI: 1.01–1.65). According to the spatially adjusted model, in high endemic snakebite areas (e.g. Northcentral province with 600 bites per 100 000), maximum snakebite risk was noted among the low-educated, field working, low-income 55-year-old males, for whom our estimated probability of being bitten was 0.029 (95% CI: 0.023–0.036). In the same areas, educated, non-field working, middle income 55-year-old females showed 0.006 (95% CI: 0.005–0.008) probability of being bitten by snakes.



**Fig 2. Probability of observing a snakebite along with age.** Variation of probability of observing a snakebite along with age for a) low-educated, field-working, low-income males living in a high snakebite endemic area (solid line). b) higher educated, non-field working, middle-income females living in the same area and (dashed line). Grey colour bands represent the 95% confidence intervals.

<https://doi.org/10.1371/journal.pone.0223021.g002>



**Fig 3. Spatial variation of ethnicity and snakebite incidence.** Spatial variation of A) ethnic groups and B) snakebite incidence in the sampled locations. Sinhalese lived in areas with relatively high snakebite incidence compared to other ethnic groups.

<https://doi.org/10.1371/journal.pone.0223021.g003>

### Difference between spatially non-adjusted and adjusted models

Both models showed age, sex, employment and education to be risk factors for snakebite. Ethnicity was a risk factor in the spatially non-adjusted model and showed a significant interaction with income categories. Ethnicity did not appear as a risk factor in the spatially adjusted model whereas income appeared an independent risk factor. Both ethnicity and snakebite incidence demonstrated spatial variation across the country (Fig 3). Sinhalese lived in relatively high snakebite incidence areas compared to other ethnic groups (median (IQR): 372 (244–508) per 100,000 among Sinhalese versus 250 (184–328) per 100,000 among other ethnic groups). Therefore, it is likely that the ethnicity has acted as a proxy variable to underlying spatially varying snakebite incidence when the snakebite incidence is not considered (i.e. spatially non-adjusted model).

### Discussion

Both the individual-level characteristics and local snakebite incidence showed important effects on determining snakebite risk. Individual-level covariates remain significant after adjusting for local snakebite incidence. This highlights that the local snakebite risk (i.e. geographical location) does not full account for variation in individual level risk. Local snakebite incidence provides only an average snakebite risk on a given location and individual level characteristics, such as individual behavior, cause heterogeneity in risk at a location. Location of an individual could be either an inherent risk factor itself or a proxy for unmeasured, spatially structured risk factors, in which case the spatial adjustment helps to estimate the effect size of the individual level covariates that are confounded by their locations [24]. This ambiguity is unavoidable in a non-randomised observational study. Hence, whilst we consider it important to consider the effects of both individual-level and geographical risk factors on health outcomes their interpretation needs to be informed by context. Here, there is strong face validity to the proposition that snakebite risk depends both on environmental factors, not all measured, that vary between locations (here, clusters), and on individual factors that vary within locations.

Both spatially adjusted and non-adjusted models showed that middle-aged, males, field workers and low-educated individuals have high risk for snakebites. Low-income level

appeared as an independent risk factor for snakebite in the spatially adjusted model. In the non-spatially adjusted model, all the exposure variables except religion showed association with snakebites, and the same model showed an interaction between ethnicity and income levels. Religion is closely related to ethnicity in Sri Lanka and the non-significance of religion can be explained by the presence of ethnicity in the non-spatially adjusted model. It is likely that the significance of all other variables, including the interaction term, is at least partially attributable to the attempt to capture the residual variability in the data rather than reflecting direct causation. The interaction between ethnicity with income was non-significant in the spatially adjusted model. It is likely that in the model without spatial adjustment ethnicity acts as proxy variable for geographical variation in snakebite incidence, as ethnicity shows a noticeable geographic variation over the country [25]. A previous national mortality study in India has reported an association between snakebite deaths and religion [26]. It is possible that this result also reflects an underlying geographical variation in snakebite risk rather than a causal effect. Therefore, failure to consider both ecological and individual characteristics can lead to misinterpretation of risk factors that are acting as proxies for unmeasured information [4].

Both spatially adjusted and non-adjusted models give similar parameter estimates and standard errors for gender and for the quadratic effect of age. Neither gender nor age is likely to be spatially confounded in the current setting. In contrast, parameter estimates for employment and education were higher in the spatially non-adjusted model compared to the spatially adjusted model. This could reveal either an overestimation of risk in the spatially non-adjusted model [6] or dilution of risk estimates in the spatially adjusted model [24]. On the other hand, the spatially non-adjusted model estimated a lower overall probability for snakebites for the high-risk individuals compared to the spatially adjusted model (i.e. 0.029 (95%CI: 0.023–0.036) vs 0.012 (95% CI: 0.009–0.017) respectively).

Individual level covariates that are identified by this study (i.e. in the spatially adjusted model) are compatible with previous literature. Both local [11] [12] [13] and regional studies [14] [15] [16] [17] have shown that the males and middle-aged have high snakebite risk. The active workforce of Sri Lanka comprises 63.5% of males and the majority of the workforce belong to middle-aged groups [27], therefore males and middle-aged individuals are likely to encounter more exposures to snakebites while working in rural industries. Previous studies have also reported a high percentage of farmers [12] [28] [29] and low educational levels among snakebite victims who are admitted to hospitals [28]. Snakebite is a disease of poverty; low income and low socio-economic individuals have been identified as risk factors for snakebite [30].

The limitations of our study include the following. Firstly, our analysis relies on recall, rather than direct observation, of snakebite events. Secondly, the snake bitten pattern varies between species, but we do not have data on the biting species. Thirdly, the national snakebite survey captured only the employment status of individuals, not the victim's activity at the time of the bite. Fourthly, we did not consider the seasonal variation of snakebite into the analysis and estimates can be subjected to seasonal bias. Finally, the true snakebite incidence at each sampled location is unknown, hence we used multiple imputation to estimate parameters in the spatially adjusted model in order to allow for the uncertainty in the estimated incidence map.

In conclusion, we highlight the importance of considering environmental information in addition to individual-level factors when designing epidemiological studies so as to avoid the latter acting as proxies for the former, leading in turn to biased estimates of effect sizes. We suggest that it is important to consider both spatial and non-spatial information in future health research [31]. Our results show that males, middle-aged individuals, field workers, individuals with low-education and income are at high risk of snakebite. This highlights the

importance of conducting educational programs and mass media campaigns, to educate the public on vulnerable groups for snakebite and to promote safety measure to avoid snakebites.

## Supporting information

### **S1 Fig. Location map of Sri Lanka.**

(TIFF)

**S2 Fig. Administrative boundaries and sampled Grama Niladari Divisions.** a) Administrative boundaries of Sri Lanka. Dark lines demarcate the provinces and grey lines demarcate the districts of Sri Lanka. b) Locations of the Grama Niladari Divisions sampled by the National Snakebite Survey of Sri Lanka.

(TIFF)

**S3 Fig. Empirical variograms of the estimated cluster level random effects of the spatially adjusted model.** The black line indicates the empirical variogram of the predicted random effects. Dashed lines indicate the 95% pointwise tolerance envelope for the empirical variograms of 1000 random permutations of the random effects.

(TIFF)

### **S1 Table. Estimated covariance parameters of spatial correlation.**

(DOCX)

### **S1 Appendix. Geostatistical modelling of snakebite incidence.**

(DOCX)

### **S1 Data. Individual level data.**

(CSV)

### **S2 Data. Data for snakebite incidence estimation.**

(CSV)

## Acknowledgments

Dileepa Ediriweera wish to acknowledge the Commonwealth Scholarship Commission for awarding the Split-Site PhD Scholarship to study at the Centre for Health Informatics, Computing and Statistics, Lancaster University, United Kingdom.

The authors wish to acknowledge the assistance given by the Director General of Health Services, Provincial Directors of Health Services of the nine provinces, District Secretaries and Regional Directors of Health Services of the 25 administrative districts of Sri Lanka, all the Divisional Secretaries and the respective Grama Niladhari of the sampled Grama Niladhari Divisions for implementation of the study. The technical contributions of Dr. Suwin Hewage, Ms. Dilki Erangika, Ms. Ishara Henayake, Ms. Nayana Harshakanthi, Ms. W.M.A. Amarasinghe, Mr. R.M.G.C.S.B. Jayatissa, Mr. W.P.A. Wickramarathne, Mr. J.H.M.B.T. Jayasundara, Mr. V. Vijayakanth, Mr. E.A.S. Edirisinghe, Mr. R. Tudugala, Ms. T.A.R. Dilrukshi, Mr. T. Thivagar, Mr. B.M.A.I. Balasooriya, Mr. R.M.U.S. Rathnayake, Mr. A. Aravinthan, Mr. S.W. Buddhika, Ms. H.R.I.N. Henayake, Mr. K.A.S.K. Karunathilaka, Mr. S. Ushjenthana, Ms. A.M. B.D.E. Bandarage, Mr. S.M.R. Prasad, Mr. P. Tishanthan, Mr. T.B. Wickramasinghe and Mr. N.W.M. Chathuranga to the study are also acknowledged.

## Author Contributions

**Conceptualization:** Dileepa Senajith Ediriweera, Peter John Diggle.

**Data curation:** Dileepa Senajith Ediriweera.

**Formal analysis:** Dileepa Senajith Ediriweera.

**Funding acquisition:** Kris Murray, Takuya Iwamura, David Griffith Lalloo, Hithanadura Janaka de Silva, Peter John Diggle.

**Investigation:** Dileepa Senajith Ediriweera.

**Methodology:** Dileepa Senajith Ediriweera, Peter John Diggle.

**Project administration:** Anuradhani Kasthuriratne, Arunasalam Pathmeswaran, Nipul Kithsiri Gunawardene, Shaluka Francis Jayamanne, David Griffith Lalloo, Hithanadura Janaka de Silva.

**Software:** Dileepa Senajith Ediriweera.

**Supervision:** Arunasalam Pathmeswaran, David Griffith Lalloo, Hithanadura Janaka de Silva, Peter John Diggle.

**Validation:** Dileepa Senajith Ediriweera.

**Visualization:** Dileepa Senajith Ediriweera.

**Writing – original draft:** Dileepa Senajith Ediriweera, Peter John Diggle.

**Writing – review & editing:** Dileepa Senajith Ediriweera, Anuradhani Kasthuriratne, Arunasalam Pathmeswaran, Nipul Kithsiri Gunawardene, Shaluka Francis Jayamanne, Kris Murray, Takuya Iwamura, David Griffith Lalloo, Hithanadura Janaka de Silva.

## References

1. Smith GD, Hart C, Watt G, Hole D, Hawthorne V. Individual social class, area-based deprivation, cardiovascular disease risk factors, and mortality: the Renfrew and Paisley Study. *J Epidemiol Community Health*. 1998; 52: 399–405. <https://doi.org/10.1136/jech.52.6.399> PMID: 9764262
2. Gordon-Larsen P, Nelson MC, Page P, Popkin BM. Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics*. 2006; 117: 417–24. <https://doi.org/10.1542/peds.2005-0058> PMID: 16452361
3. Wang J, Li X, Christakos G, Liao Y, Zhang T, Gu X, et al. Geographical Detectors-Based Health Risk Assessment and its Application in the Neural Tube Defects Study of the Heshun Region, China. *Int J Geogr Inf Sci*. 2010; 24: 107–127. <https://doi.org/10.1080/13658810802443457>
4. Pickett KE, Pearl M. Multilevel analyses of neighbourhood socioeconomic context and health outcomes: A critical review. *J Epidemiol Community Health*. 2001; 55: 111–122. <https://doi.org/10.1136/jech.55.2.111> PMID: 11154250
5. Ecob R, Jones K. Mortality variations in England and Wales between types of place: An analysis of the ONS longitudinal Study. *Soc Sci Med*. 1998; 47: 2055–2066. [https://doi.org/10.1016/s0277-9536\(98\)00310-4](https://doi.org/10.1016/s0277-9536(98)00310-4) PMID: 10075246
6. O'Campo P, Xue X, Wang MC, Caughy M. Neighborhood risk factors for low birthweight in Baltimore: a multilevel analysis. *Am J Public Health*. 1997; 87: 1113–1118. <https://doi.org/10.2105/ajph.87.7.1113> PMID: 9240099
7. Ediriweera DS, Kasturiratne A, Pathmeswaran A, Gunawardena NK, Wijayawickrama BA, Jayamanne SF, et al. Mapping the Risk of Snakebite in Sri Lanka—A National Survey with Geospatial Analysis. *PLoS Negl Trop Dis*. 2016;10. <https://doi.org/10.1371/journal.pntd.0004813> PMID: 27391023
8. Ediriweera DS, Diggle PJ, Kasturiratne A, Pathmeswaran A, Gunawardena NK, Jayamanne SF, et al. Evaluating temporal patterns of snakebite in Sri Lanka: The potential for higher snakebite burdens with climate change. *Int J Epidemiol*. 2018; 47: 2049–2058. <https://doi.org/10.1093/ije/dyy188> PMID: 30215727
9. Jayawardana S, Gnanathanan A, Arambepola C, Chang T. Chronic Musculoskeletal Disabilities following Snake Envenoming in Sri Lanka: A Population-Based Study. *PLoS Negl Trop Dis*. 2016;10. <https://doi.org/10.1371/journal.pntd.0005103> PMID: 27814368

10. Ediriweera DS, Kasturiratne A, Pathmeswaran A, Gunawardena NK, Jayamanne SF, Laloo DG, et al. Health seeking behavior following snakebites in Sri Lanka: Results of an island wide community based survey. *PLoS Negl Trop Dis*. 2017; 11: e0006073. <https://doi.org/10.1371/journal.pntd.0006073> PMID: 29108023
11. Seneviratne U, Dissanayake S. Neurological manifestations of snake bite in Sri Lanka. *J Postgrad Med*. 2002; 48: 275–8; discussion 278–9. PMID: 12571382
12. Kularatne SAM. Epidemiology and clinical picture of the russell's viper (*daboia russelii russelii*) bite in anuradhapura, sri lanka: A prospective study of 336 patients. *Southeast Asian J Trop Med Public Health*. 2003; 34: 855–862. <https://doi.org/10.1016/j.trstmh.2011.07.010> PMID: 15115100
13. Ariaratnam CA, Sheriff MHR, Theakston RDG, Warrell DA. Distinctive epidemiologic and clinical features of common krait (*Bungarus caeruleus*) bites in Sri Lanka. *Am J Trop Med Hyg*. 2008; 79: 458–462. doi: 79/3/458 [pii] PMID: 18784244
14. Sharma SK, Chappuis F, Jha N, Bovier PA, Loutan L, Koirala S. Impact of snake bites and determinants of fatal outcomes in Southeastern Nepal. *Am J Trop Med Hyg*. 2004; 71: 234–238. <https://doi.org/10.4269/ajtmh.2004.71.234> PMID: 15306717
15. Sinha A, Bhattacharya S, Ram R, Dasgupta U, Ram A, Majumder D. Epidemiological profile of snake bite in South 24 Parganas district of West Bengal with focus on underreporting of snake bite deaths. *Indian J Public Health*. 2014; 58: 17. <https://doi.org/10.4103/0019-557X.128158> PMID: 24748352
16. Hati AK, Mandal M, De MK, Mukherjee H, Hati RN. Epidemiology of snake bite in the district of Burdwan, West Bengal. *J Indian Med Assoc*. 1992; 90: 145–147. PMID: 1522302
17. Rahman R, Faiz MA, Selim S, Rahman B, Basher A, Jones A, et al. Annual incidence of snake bite in rural Bangladesh. *PLoS Negl Trop Dis*. 2010;4. <https://doi.org/10.1371/journal.pntd.0000860> PMID: 21049056
18. Kasturiratne A, Wickremasinghe AR, De Silva N, Gunawardena NK, Pathmeswaran A, Premaratna R, et al. The global burden of snakebite: A literature analysis and modelling based on regional estimates of envenoming and deaths. *PLoS Medicine*. 2008. pp. 1591–1604. <https://doi.org/10.1371/journal.pmed.0050218> PMID: 18986210
19. Caiaffa WT, Antunes CM, de Oliveira HR, Diniz CR. Epidemiological and clinical aspects of snakebite in Belo Horizonte, southeast Brazil. *Rev Inst Med Trop Sao Paulo*. 1997; 39: 113–118. <https://doi.org/10.1590/s0036-46651997000200009> PMID: 9394525
20. Morandi N, Williams J. Snakebite injuries: Contributing factors and intentionality of exposure. *Wilderness Environ Med*. 1997; 8: 152–155. [https://doi.org/10.1580/1080-6032\(1997\)008\[0152:SICFAI\]2.3.CO;2](https://doi.org/10.1580/1080-6032(1997)008[0152:SICFAI]2.3.CO;2) PMID: 11990155
21. R Core Team. R: A language and environment for statistical computing. R Found Stat Comput Vienna, Austria. 2017; <https://doi.org/10.1017/CBO9781107415324.004>
22. Giorgi E, Diggle P. PreVMap: an R package for prevalence mapping. *J Stat Softwar*. 2017;10.
23. Diggle PJ, Ribeiro PJ. *Model-based Geostatistics*. Springer; 2007. <https://doi.org/10.1111/1467-9876.00113>
24. Clayton DG, Bernardinelli L, Montomoli C. Spatial Correlation in Ecological Analysis. *Int J Epidemiol*. 1993; 22: 1193–1202. <https://doi.org/10.1093/ije/22.6.1193> PMID: 8144305
25. Ross R, Savada A. Sri Lanka: A country study. Washington, DC: GPO for the Library of Congress; 1988.
26. Mohapatra B, Warrell DA, Suraweera W, Bhatia P, Dhingra N, Jotkar RM, et al. Snakebite mortality in India: A nationally representative mortality survey. *PLoS Negl Trop Dis*. 2011;5. <https://doi.org/10.1371/journal.pntd.0001018> PMID: 21532748
27. Department of Census and Statistics. Sri Lanka Labour Force Statistics [Internet]. 2016. Available: <http://www.statistics.gov.lk/>
28. Kularatne AM, Silva A, Maduwage K, Ratnayake I, Walathara C, Ratnayake C, et al. Victims' response to snakebite and socio-epidemiological factors of 1018 snakebites in a tertiary care hospital in Sri Lanka. *Wilderness Environ Med*. 2014; 25: 35–40. <https://doi.org/10.1016/j.wem.2013.10.009> PMID: 24412659
29. Kularatne SAM, Silva A, Weerakoon K, Maduwage K, Walathara C, Paranagama R, et al. Revisiting Russell's Viper (*Daboia russelii*) bite in Sri Lanka: Is abdominal pain an early feature of systemic envenoming? *PLoS One*. 2014;9. <https://doi.org/10.1371/journal.pone.0090198> PMID: 24587278
30. Harrison RA, Hargreaves A, Wagstaff SC, Faragher B, Laloo DG. Snake envenoming: A disease of poverty. *PLoS Negl Trop Dis*. 2009;3. <https://doi.org/10.1371/journal.pntd.0000569> PMID: 20027216
31. Wang F, Luo W. Assessing spatial and nonspatial factors for healthcare access: Towards an integrated approach to defining health professional shortage areas. *Health and Place*. 2005. pp. 131–146. <https://doi.org/10.1016/j.healthplace.2004.02.003> PMID: 15629681