International Journal for
Equity in Health

RESEARCH

Open Access

Bayesian estimation of the effect of health inequality in disease detection

Dinah Jane Lope, Haydar Demirhan[*] and Anil Dolgun

Abstract

Background: Measuring health inequality is essential to ensure that everyone has equal accessibility to health care. Studies in the past have continuously presented and showed areas or groups of people affected by various inequality in accessing the health resources and services to help improve this matter. Alongside, disease prevention is as important to minimise the disease burden and improve health and quality of life. These aspects are interlinked and greatly contributes to one's health.

Method: In this study, the Gini coefficient and Lorenz curve are used to give an indication of the overall health inequality. The impact of this inequality in granular level is demonstrated using Bayesian estimation for disease detection. The Bayesian estimation used a two-component modelling approach that separates the case detection process and incidence rate using a mixed Poisson distribution while capturing underlying spatio-temporal characteristics. Bayesian model averaging is used in conjunction with the two-component modelling approach to improve the accuracy of estimates by incorporating many candidate models into the analysis instead of using fixed component models. This method is applied to an infectious disease, influenza, in Victoria, Australia between 2013 and 2016 and the corresponding primary health care of the state.

Result: There is a relatively equal distribution of health resources and services pertaining to general practitioners (GP) and GP clinics in Victoria, Australia. Roughly 80 percent of the population shares 70 percent of the number of GPs and GP clinics. The Bayesian estimation with model averaging revealed that access difficulty to health services impacts both case detection probability and incidence rate. Minimal differences are recorded in the observed and estimated incidence of influenza cases considering social deprivation factors. In most years, areas in Victoria's southwest and eastern parts have potential under-reported cases consistent with their relatively lower number of GP or GP clinics.

Conclusion: The Bayesian model estimated a slight discrepancy between the estimated incidence and the observed cases of influenza in Victoria, Australia in 2013-2016 period. This is consistent with the relatively equal health resources and services in the state. This finding is beneficial in determining areas with potential under-reported cases and under-served health care. The proposed approach in this study provides insight into the impact of health inequality in disease detection without requiring costly and time-extensive surveys and relying mainly on the data at hand. Furthermore, the application of Bayesian model averaging provided a flexible modelling framework that allows covariates to move between case detection and incidence models.

Keywords: Gini coefficient, Lorenz curve, Bayesian statistics, Bayesian model averaging, Influenza, Case detection, Incidence

*Correspondence: haydar.demirhan@rmit.edu.au

School of Science, Mathematical Sciences Discipline, RMIT University, Melbourne 3000, Australia

Background

Measuring health inequality is crucial as it reflects the performance of the country's health system, which is crucial towards the healthy development of individuals,

families and society as a whole [1]. Australia has an extensive health prevention and promotion plan in keeping all healthy as long as possible. According to the Australian Institute of Health and Welfare, it is essential to have access to timely, appropriate and quality health care at the right cost during ill health. In addition, all is not equal. Location, income, the existence of a disability, access to health services; all contribute to one's health. The health system can be assessed by measuring how health care varies across geographic areas, the number of potentially preventable hospitalisations and adverse events in hospitals [2]. Given this, it overlays the importance of health care and disease studies to keep the nation healthy.

Recognising its importance, the World Health Organisation has been actively promoting to reduce the health inequalities within and across countries [1]. In the year 2000, Wilkinson and Symon [3] first measured the inequality of general practitioners in Australia. The study showed that non-metropolitan areas were relatively under-served of general practitioners. On the other hand, Hann and Gravelle [4] explored the same study on general practitioners in England and Wales from 1974 to 2003, and the findings showed increased maldistribution of general practitioners from mid-1980s to 2003. In Albania, Theodorakis et al. [5] found that there is a trend in the inequality of general practitioners while Toyabe [6] in Japan observed that the number of physicians in the hospital has increased significantly in urban areas but not in low population areas. Matsumoto et al. [7] assessed the distribution of primary care physicians in Japan and Britain to compare the primary care systems between the two countries. Results suggested that the distribution in Japan is less balanced than Britain's by about twice the observed inequality [7].

Ohba et al. [8] compared the geographical inequalities of radiotherapy resources in two districts of Japan, Hokkaido-Tohoku and Tokyo. Results demonstrated that the inequality is higher in the Tokyo district. In China, Liu et al. [9] analysed the inequality of emergency medical services (EMS) in Chongqing city. This study aimed to determine the population and geographic distribution of EMS-related facilities and human resources. Findings suggested that there is a low allocation of resources and the provision of services did not meet the patients' needs [9]. While Zhang et al. [10] studied the overall distribution of health resources and services in China as the country's major goal is to achieve equity in the health system reform. It is observed that inequality is more apparent in the geographic distribution than per capita and better resources and services to the rich as compared to the poor, a risk for two-tiered health care system [10].

Disease studies and surveillance are important to minimise the burden of the disease that affects the health and quality of life. According to the Centers for Disease Control and Prevention (CDC), the end goal of disease surveillance is to keep people healthy. This can be done by detecting the disease when and where it happens, stopping before it spreads, studying the disease and improving prevention and control. According to CDC, spatial analysis of disease has crucial value in order to stop the spread and keep people healthy. Infectious diseases have a tricky nature of transmitting the virus in the community at a particular time and speed [11].

Influenza, being an epidemic infection, affects millions of people around the world on yearly basis [12]. The spread of a virus is very complex in local and global contexts, with many factors affecting its onset, time span, and severity. In 2018, Geoghegan et al. [13] conducted an in-depth spatio-temporal analysis of influenza in Australia driven by the study in the United States showing a strong geographical correlation. Given the complexity of the climate and a smaller population in Australia, albeit similar in land-size of the United States, this trend is yet unknown to apply in Australia. Results showed that despite the wide spatial and temporal variations, there is remarkable epidemiological synchronicity in the continent. This emphasises the importance of well-coordinated responses in times of outbreaks and human-to-human transmissions [13]. Moa et al. [12] explored Australia's inter-seasonality of influenza wherein results indicated a relatively constant summer-to-winter influenza notifications ratio in the country, which is consistently supported by the constant yearly hospitalisation rates.

In terms of case detection, Shaweno et al. [14] used a Bayesian binomial mixture geospatial model to estimate tuberculosis (TB) incidence and case detection rate in Ethiopia. Stoner et al. [15] proposed a hierarchical Bayesian approach that uses prior information and corrects for under-reporting of the disease. They implemented their approach to analyse TB incidence and case detection in Brazil. These two studies provide us with a practical modelling framework to estimate the number of under-reported cases concerning other health services and socio-economic factors. With a very comprehensive study, Zipfel et al. [16] recently proposed a Bayesian framework to determine important factors on health disparities in influenza cases in the US. They consider socioeconomic covariates and transmission of the disease in relation to the socio-economic status, and aim to capture spatial variations in health inequalities. They develop an epidemiological model and a spatial inferential model which is very similar to those of Shaweno et al. [14] and Stoner

Lope *et al. International Journal for Equity in Health* (2022) 21:118

Page 3 of 17

et al. [15]. Although Zipfel et al. [16] consider the spatial variability, they do not take the spatial correlation between adjacent areas into the modelling as Shaweno et al. [14] and Stoner et al. [15] did.

These studies show the importance of investigating two main health factors: the health resources and services and the disease concerning the geographical variation. In this study, we aim to extend the overall measure of inequality by exploring its impact on a granular level through Bayesian estimation for influenza cases in Victoria, Australia. As Zipfel et al. [16], we also consider socioeconomic factors through their spatial variation in our modelling. We utilise the estimates of under-reported cases due to the lack of health resources and services to infer spatial health inequality. To achieve this aim, we improve the case detection models developed for TB disease and implement them for influenza cases. In the Bayesian model frameworks of Shaweno et al. [14], Stoner et al. [15], and surveillance model of Zipfel et al. [16], the covariates belonging to the incidence rate and detection probability components of the model are determined by the practitioner before running the model. However, some of the socioeconomic covariates can influence either component of the model. Instead of assigning all the covariates to specific components, letting them float between the model components contributes to the performance of the model and captures the impact of floating covariates on both the incidence rate and case detection probability. We improve the modelling framework by introducing Bayesian model averaging (BMA) that allows flexibility in the covariates to move between detection probability and incidence rate components. In this way, we improve the modelling and fill in the gap by estimating the effect of health inequality on the disease detection rate, which does not only identify areas that have under-served health resources and services but also areas with possible under-reported disease cases as a result of this health inequality. The contributions of this study are that i) we use the concept of under-reporting to detect spatial health inequality considering the correlations between adjacent areas. ii) We utilise socioeconomic factors to model both incidence rate and detection probability without restricting them only to affect either the model's case detection probability component or its incidence rate component. iii) We provide a picture of health inequality across the local government areas of Victoria, Australia in relation to influenza disease. Our approach in this study is readily generalisable to other countries and diseases to analyse spatial health inequality.

Methods and data

This study will review the impact of access to health care on disease detection. Firstly, an overall inequality of health resources and services is measured using inequality indices. Secondly, we estimate the effect of this degree of inequality in detecting the disease using Bayesian estimation.

Inequality measure

The Gini coefficient is a widely used inequality measure mainly applied in income and wealth. Several studies in the past have explored the distribution of health services from general practitioners to specialists and health resources such as emergency department and radiotherapy by using the Gini coefficient. Through this coefficient, these studies determined areas or groups of people experiencing inequality in the distribution. In Albania, Theodorakis et al. [5] used the Gini coefficient to compare the inequality of distribution of GPs in 2000-2004 and recorded a declining trend across the years. In Japan, Toyabe [6] compared the Gini coefficient with other inequality indices such Atkinson index and Theil index to assess the distribution of physicians where all the indices generated relatively the same findings. Also in Japan, Ohba et al. [8] determined that the inequality of radiotherapy health resources in Tokyo district is higher than in the Hokkaido-Tohoku by using the Gini coefficient. Furthermore, Matsumoto et al. [7] used Gini coefficient to validate that primary care physicians' distribution in Britain is relatively equal than in Japan. On the other hand, in China, Liu et al. [9] used the Gini coefficient to determine that distribution of physicians and nurses in Chongqing city is relatively equal. In China as a whole, Zhang et al. [10] found that the geographic distribution of the health resources and health services is less equal than per capita using the Gini coefficient.

The value of Gini coefficient ranges from 0 to 1, wherein a lower Gini coefficient means a more equal distribution while a higher Gini coefficient means a more unequal distribution. In conjunction, the Lorenz curve is used to visualise the Gini coefficient. In this study, the application of these indices will determine the overall value and graphical representation of the inequality of health resources and services using Eq. (1) [17]:

$$G = 1 - \sum_{k=1}^{n}(Y_{k-1} + Y_k)(X_k - X_{k-1}), \qquad (1)$$

where X_k is the cumulative proportion of the population for $k=0,...,n$, with $X_0=0$, $X_n=1$ and Y_k is the

Lope *et al. International Journal for Equity in Health* (2022) 21:118

Page 4 of 17

cumulative proportion of the health resources and services for $k=0,...,n$ with $Y_0 = 0$, $Y_n=1$. The relationship between Gini coefficient and Lorenz curve is given in Eq. (2):

$$G = A/(A + B), \qquad (2)$$

where A is the area above the curve and B is the area below the curve.

Bayesian estimation

Shaweno et al. [14] introduced the use of the Bayesian technique in estimating tuberculosis incidence and case detection rate in Ethiopia. They used a two-component Bayesian model composed of a model for incidence rate and another model for the case detection rate. Their study was able to determine TB hotspots that were not previously identified. Further, the model used in their study clearly separated true incidence and case detection that enables to identify whether a high notification rate is due to a failing health system, failing to detect patients rapidly or an efficient health system diagnosing all incidences quickly. Results revealed a vast difference between notification rate and estimated incidence rate in areas with no health facilities, highlighting the importance of reducing the number of missed cases. This approach provided an alternative and practical approach to estimate incidence by only using routinely collected surveillance data [14]. Stoner et al. [15] developed a similar approach by correcting the under-reporting in count data using enhanced models and additional prior information to supplement partial information in the data. In their study, the model for incidence includes various social deprivation covariates such as homelessness and incarceration as some spatial variability can be attributed to these covariates [15]. These approaches from Shaweno et al. and Stoner et al. are used in this study to estimate the effect of unequal health resources and services in disease detection, taking into consideration the spatio-temporal diversity and the correlation between neighbouring areas.

Under the assumption that individual cases are detected independently at a fixed rate and are conditional on individual incidence [14, 15], two main component models to capture (true) incidence and case detection probability are explored in this study using the following modelling framework:

$$\text{Incidence}_{t,s} \sim \text{Poisson}(\lambda_{t,s}), \qquad (3)$$

such that,

$$\log(\lambda_{t,s}) = \alpha_0 + \sum_{k=1}^{K} \alpha_k \, u_{t,s}^{(k)}, \qquad (4)$$

where $\{u^{(k)}\}$ includes covariates that give rise to the number of cases including balancing effect such as population counts. The expected number of detected cases, $Z_{t,s}$, follows a binomial distribution conditional on the hidden true incidence rate and the probability of the case being detected:

$$Z_{t,s} \sim \text{Binomial}(\pi_s, \text{Incidence}_{t,s}), \qquad (5)$$

where $\text{Incidence}_{t,s}$ is defined in Eq. (3) and

$$\log\left(\frac{\pi_s}{1 - \pi_s}\right) = \beta_0 + \sum_{j=1}^{J} \beta_j \, v_{t,s}^{(j)}. \qquad (6)$$

In Eq. (6), $\{v^{(j)}\}$ is the vector of covariates related to the disease detection process. This allows to record the severity of under-reporting and what it pertains to. Using integration and Bayes' rule [15],

$$\text{Incidence}_{t,s} - Z_{t,s} \sim \text{Poisson}((1 - \pi_s)\lambda_{t,s}), \qquad (7)$$

where s represents the space index and t the time index to capture both the spatial and temporal effects.

One of the risks highlighted by Stoner et al. [15] is the uncertainty in identifying or classifying whether a covariate belongs to the case detection process or the true incidence; hence, the distinction between if a covariate impacts case detection rate or incidence rate is unclear. We call this type of covariate *floating covariate* in this study. To overcome this ambiguity, we propose to use Bayesian model averaging (BMA) within the implementation of the modelling framework of Stoner et al. [15]. We establish a model space $\wp(S)$ such that all elements are considered to estimate the parameter of interest Δ given the data D [18]. Then, the posterior distribution of the parameter of interest is calculated as in Eq. (8):

$$P(\Delta|D) = \sum_{m \in \wp(S)} P(\Delta|m, D)P(m|D). \qquad (8)$$

BMA allows all the alternative models in a model space $\wp(S)$ to have a chance to impact the overall parameter estimates proportional to their posterior model probabilities that are derived by

$$P(m|D) = \frac{P(D|m)P(m)}{\sum_{m' \in \wp(S)} P(D|m')P(m')}. \qquad (9)$$

In this study, a model space is created by considering floating covariate(s) in both of case detection probability and the incidence rate components of the overall model, and we let the BMA capture the part where the floating covariate will most likely be. This is our main contribution over the existing work of Shaweno et al. [14] and Stoner et al. [15].

Data and descriptive analysis

The method will be applied using the notified cases of influenza in Victoria, Australia. The notified cases of influenza are requested from National Notifiable Diseases Surveillance System from 2013-2016 spatially aggregated in 79 Local Government Areas (LGA) [19]. The data for Victoria, which contains the social deprivation covariates including the availability of health resources and services and the access difficulty covariate, are all accessed from the Australian Urban Research Infrastructure Network (AURIN) Portal [20, 21]. The estimated resident population is sourced from the Australian Bureau of Statistics [22]. The influenza cases and all other covariates were aggregated by LGA and year to evaluate the spatial and temporal effects.

The distribution of influenza notifications and their trends by LGA over the observation period of 2013-2016 are given in Fig. 1.

The right panel of Fig. 1 shows the yearly distribution of the notified cases of influenza regardless of type in Victoria from 2013-2016. The number of cases across the years has been substantially increasing with an increasing variation. Consequently, this variation may cause more variable estimates to arise during Bayesian estimation. The number of influenza cases is plotted across time for each LGA of Victoria on the left panel of Fig. 1. There is a quadratic increase in the number of influenza cases in the LGAs of Victoria. Based on this observation, we introduce a quadratic function of time into the model.

Figure 2 shows density plots of the covariates considered in the models and Fig. 3 shows the correlation of the incidence covariates against the influenza cases. For the incidence model in Eq. (10), the included covariates are population, socio-economic indexes for areas (SEIFA), household size, homelessness and year as the temporal lag. With various social factors incorporated in deriving SEIFA, it is transformed to an orthogonal polynomial with a degree of 2 to minimise multi-collinearity since it may interact with other social factors in our model, as also suggested by Stoner et al. ([15], see page 1484). From Fig. 3, there is a significant and negative correlation between the influenza cases and SEIFA, and significant positive correlations between influenza cases and household size and homelessness. For the case detection probability model in Eq. (11), GP and GP clinic per 1,000 population are considered. Further, the floating covariate, access difficulty, is to be explored in both models. Since Markov Chain Monte Carlo (MCMC) methods iteratively optimise the model parameters [23], it is time-consuming to get representative and accurate chains that do not have significant autocorrelation when the covariates' distributions have a large range and are skewed with outliers. When there are outliers in the covariates, they have the potential to make the parameter structure highly correlated and reduce the speed of convergence ([24], see page 177). In order to shorten the tails of the covariate's distributions and move outliers towards the centre of their distributions, different transformations, including trigonometric transformations and the family of Box-Cox transformations that includes log, reciprocal, square root, and polynomial transformations, are
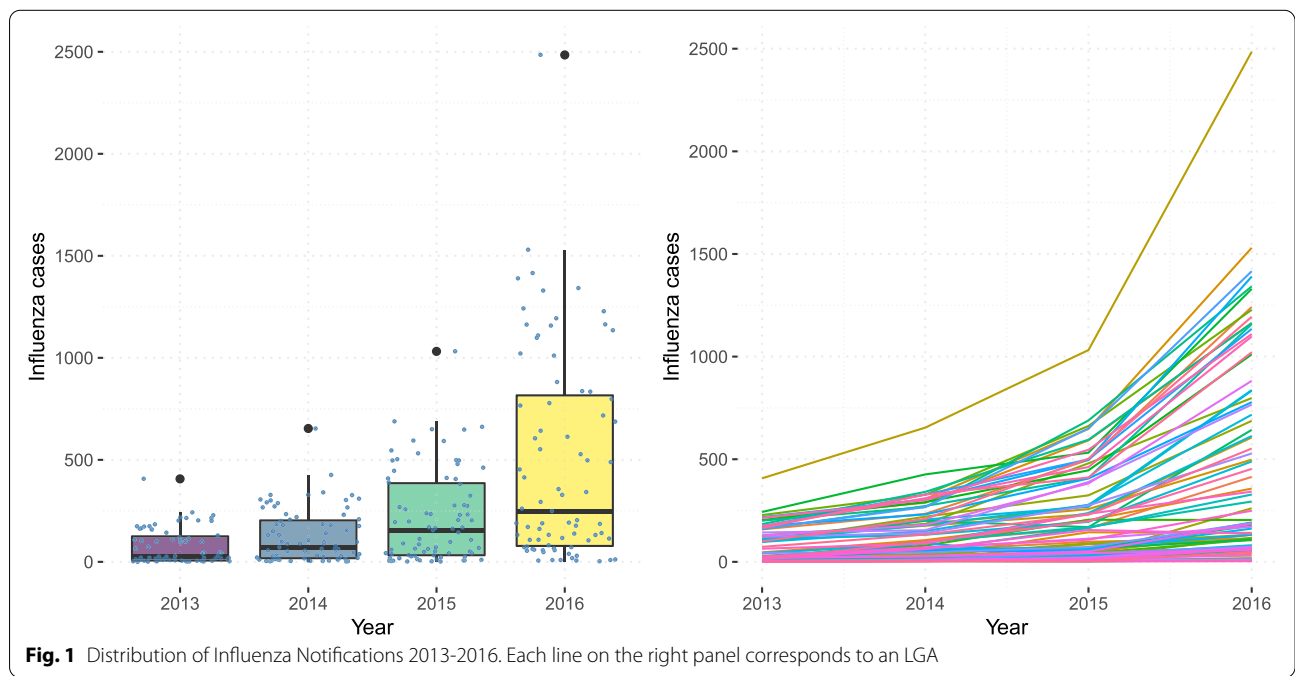
Fig. 1 Distribution of Influenza Notifications 2013-2016. Each line on the right panel corresponds to an LGA

Fig. 2 Exploratory analysis of the raw and transformed covariates
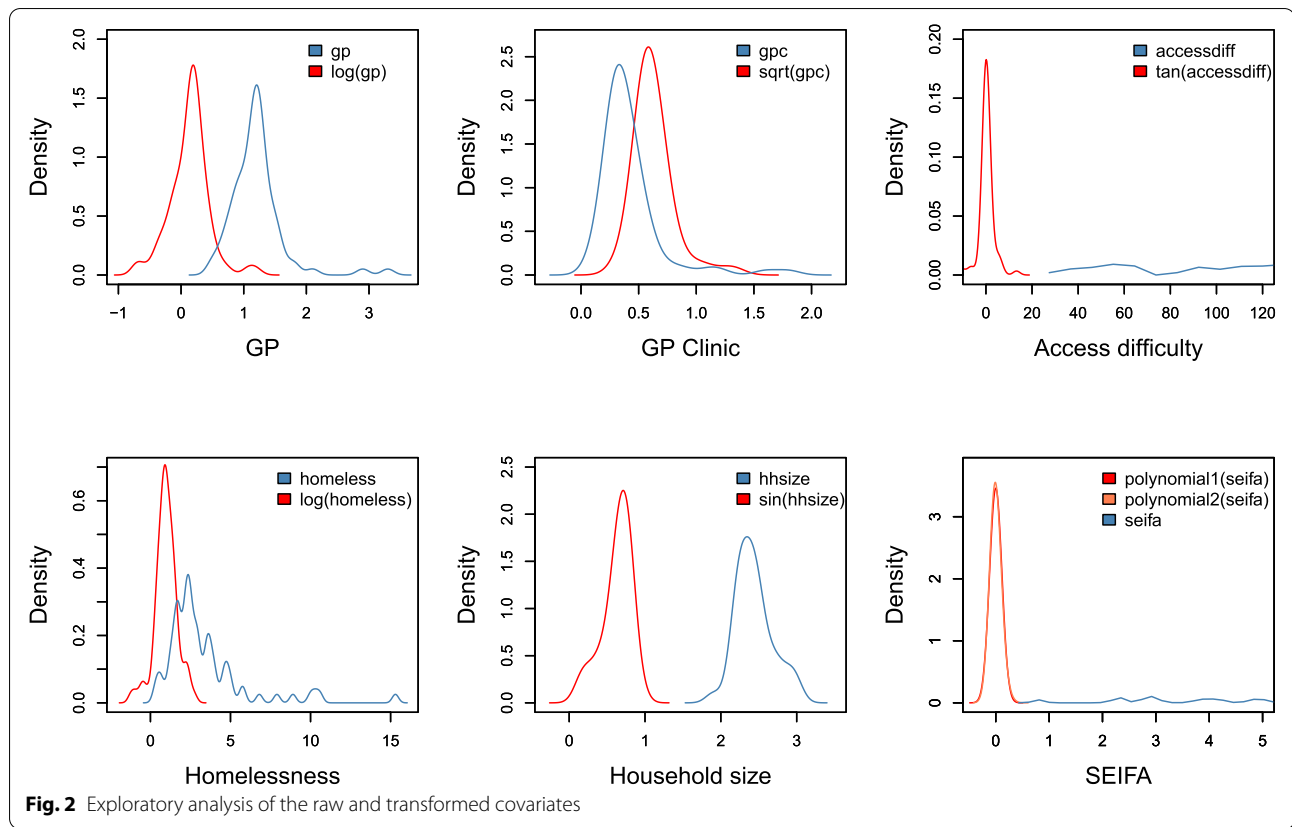
considered. The transformation that reduces the range of each covariate's distribution and makes it most symmetric is implemented. For this aim, adopting suitable transformations is considered. Common transformations include log, reciprocal, square root, in general, the family of Box-Cox transformations, and trigonometric transformations ([25], see page 169). Standardisation by centring and scaling, Box-Cox transformations by optimising the power parameter, and basic trigonometric transformations such as sine, cosine, and tangent are applied for GP, GP clinic, Access difficulty, Homelessness and household size covariates. The transformation that reduces the range of each covariate's distribution and makes it most symmetric is implemented. Red density plots in Fig. 2 show the transformed versions of the covariates.

Bayesian modelling and model space

We apply our Bayesian framework using the dataset to capture (true) incidence and case detection probability in Eqs. (3)-(7). For the incidence rate component in Eq. (3), we set the model in Eq. (10):

$$
\begin{aligned}
\log(\lambda_{t,s}) = {} & \log(\text{Pop}_{t,s}) + \alpha_0 + \boldsymbol{\alpha_1}\ \text{seifa}_s + \alpha_2\ \text{hhsize}_s + \alpha_3\ \text{homeless}_s + \phi_s + \theta_s \\
& + f(t) + I(\cdot)\alpha_4\ \text{AccessDifficulty}_s,
\end{aligned}
\tag{10}
$$

where $\boldsymbol{\alpha_1}=(\alpha_{1.1},\alpha_{1.2})$ show the parameters resulting from the polynomial transformation of SEIFA, $f(t)$ is a quadratic function of time defined as $f(t)=\alpha_5 t + \alpha_6 t^2$ and the presence of the floating uncategorised covariate AccessDifficulty$_s$ is controlled by the indicator function $I(\cdot)$ (in Eq. (11) as well). The expected number of detected cases, Z$_{t,s}$, given the distribution of the health resources and services follows a binomial distribution conditional on the hidden true incidence rate and the probability of the case being detected as given in Eq. (5) where we define Incidence$_{t,s}$ as in Eqs. (3) and (10), and set the model in Eq. (11) for the case detection probability:

$$
\begin{aligned}
\log\left(\frac{\pi_s}{1-\pi_s}\right) = {} & \beta_0 + \beta_1\ \text{GP}_s + \beta_2\ \text{GPclinic}_s \\
& + I(\cdot)\beta_3\ \text{AccessDifficulty}_s,
\end{aligned}
\tag{11}
$$

where space index s represents the LGAs in Victoria $s=0,1,...,79$ and t and the time index $t=1,2,3,4$ represents years 2013-2016, respectively.
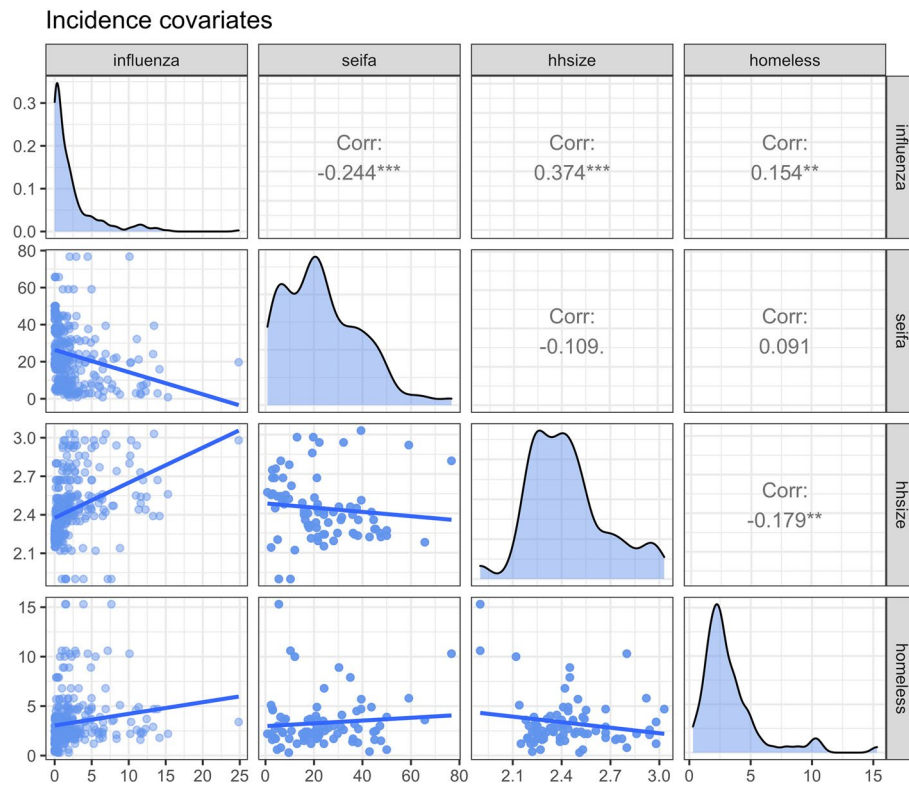
Fig. 3 Correlation matrix of the incidence covariates. Influenza counts are divided by 100 for plotting purposes. *: *P*-value <0.05, **: *P*-value <0.01, ***: *P*-value <0.001

The incidence model in Eq. (10), represented by Poisson distributed $\lambda_{t,s}$, is modelled using the population, SEIFA, homelessness and household size covariates. These covariates were considered in relation to our descriptive analysis and the study by Stoner et al. [15] on tuberculosis for social deprivation factors given the similar infectious nature of the disease. In Eq. (10), the $\log(\text{Pop}_{t,s})$ term incorporates the varying population, seifa_s is the index of relative socio-economic disadvantage (IRSD), hhsize_s is the average household size and homeless_s is the estimated number of homeless people per 1,000 population varying in space and time. ϕ_s represents the spatially structured effect modelled as an intrinsic Gaussian autoregressive model [26] to capture the effects of contiguous areas wherein areas are contiguous when $s' \neq s$ shares a common geographical boundary. θ_s is the spatially unstructured random effect. The function of time $f(t)$ is implemented in a quadratic form to emphasize the substantial increase in influenza cases through the years as shown in Fig. 1.

The case detection probability in Eq. (11), represented by the logit transformed π_s, is modelled using GP and GP clinic covariates. These covariates were considered noting that the data compiled in the NNDSS are greatly influenced by patients who seek healthcare [27]. In Eq. (11), the term GP_s is the distribution of general practitioner per 1,000 population and GPclinic_s is the distribution of general practitioner clinic per 1,000 population. The floating covariate $\text{AccessDifficulty}_s$ is the estimated number of people 18 years old and over who experienced difficulty in accessing healthcare in the last 12 months when needed, with a major reason being the cost of service. This is the floating covariate of our study for the reason that it has potential to affect both the case detection and incidence models. Therefore, it is left unclassified whether to belong in the case detection probability or the incidence rate.

There is also the possibility that the sub-models of the models in Eqs. (10) and (11) perform better than the full models. Therefore, the sub-models are also considered alongside the full models at the model averaging stage. To reflect this, 24 candidate models are formulated in the model space of our study. Figure 10 of Appendix A shows the 24 models considered in the model space $\mathscr{P}(S)$ the study. These models are combinations of the individual case detection probability and incidence models. The floating covariate AccessDifficulty is considered under the incidence rate component in Models 1-12, while it

appears under the case detection probability component in Models 13-24. By this setting for AccessDifficulty, we do not need to decide on the place of AccessDifficulty covariate before running the Bayesian model. Instead, the information in the data designates the most suitable part of the model where AccessDifficulty belongs.

The model diagrams for the implementation are presented in Figs. 11 and 12 of Appendix A, respectively, for single model implementation and Bayesian model averaging. In Fig. 11 of Appendix A, non-informative prior and hyper-prior distributions are induced on all coefficients of the models in Eqs. (10) and (11), given the absence of prior information about model parameters. The vector of residual spatial variation parameters $\boldsymbol{\phi}$ has a priori conditional autoregressive Gaussian process (CAR) [28] with zero men vector and covariance matrix of $[\tau(D-W)]^{-1}$, where τ controls precision, W is the adjacency matrix and D is the diagonal matrix with the sum of the adjacent neighbours for each LGA. The CAR process is to capture the spatial dependence between irregularly spaced sites. For the spatially unstructured random effect parameter θ_s, a non-informative normal prior with normal and half-normal hyper-priors on its parameters are set to induce prior variability to improve the accuracy of the generated chains during MCMC implementation. Figure 12 of Appendix A represents the Bayesian model averaging through a non-informative categorical prior distribution that assigns an equal prior probability to each model in the model space.

Model evaluation and implementation

The diagnostics of the models are evaluated as a preliminary step to determine the representativeness and accuracy of the MCMC samples. Moreso, a posterior predictive model checking is conducted to assess that the model fits the observed data.

The data are mainly processed and modelled in R software version 3.6.0 using various required packages. The final processed data are used to calculate the Gini coefficient and visualise the Lorenz curve. The Bayesian models are fitted using MCMC in JAGS; leveraging its capability to perform model averaging. In addition, the module GeoJAGS is used to generate structured spatial effect of the model through the CAR process [29].

The implementation of the proposed modelling framework in JAGS provides the opportunity to do model averaging which was not implemented in the previous studies that used WinBUGS and Nimble [14, 15]. However, it is noted that the usage of the module GEOJAGS in JAGS to run the complex spatial model such as ICAR is still at its early stage as compared to WinBUGS and Nimble [29].

The 24 models listed in Fig. 10 of Appendix A ran in JAGS with 140,000 iterations, 2 chains, 19 thinning while discarding the first 200,000 iterations as burn-in to configure the most probable models to represent the data.

Results

Figure 4 shows the inequality of the number of GP and GP clinics across LGAs in Victoria and the corresponding Gini coefficient. The closer the curve is to the line of perfect equality, the more equal the health resources and services are across areas. Overall, the distributions of both numbers of GP and GP clinics in Victoria are relatively equal, wherein 80 percent of the population shares roughly 70 percent of GP (Gini = 0.16) and GP clinics (Gini = 0.20). Further, the number of GPs is slightly more equal as compared to GP clinics in the population.

Figure 5 shows the posterior model probabilities of all the models in the model space, and Table 1 gives the top 5 models in terms of posterior model probability along with their model equations. In Fig. 5, the 24 models show a competitive contribution to the predictions with the domination of Model 22, dominating almost half of the model visits during MCMC.

Four out of the top five models include access difficulty in the detection probability component. However, in Model 4, the access difficulty covariate contributes to the incidence rate component. This result implies that the access difficulty covariate should neither be classified under the case detection probability nor the incidence model. Both components seem to benefit from the effect of the access difficulty covariate. Looking further, in the top 5 models in Table1, Model 4 and Model 16 are model counterparts with access difficulty being swapped between the incidence rate and case detection probability components of the model. Apparently, the use of BMA provides a more inclusive prediction by taking the contribution of all the models that have a non-zero weight and eliminating the need for a premodelling decision for fixed covariate categorization.

Figure 6 shows the posterior distributions of model parameters. The parameter estimates of β_1, β_2 and β_3 in π_s component which has health resources and resources predictors, have bi-modality. For β_3, when access difficulty is not in the π_s component, the realisations of this coefficient come from the prior distribution centred around zero. When it is in the π_s component, the realisations are created by the posterior distribution. As a result, we have a bi-modal posterior for this coefficient. For β_1 and β_2, the reason for having bi-modal posteriors is the variability caused by the inclusion of the access difficulty in half of the 24 models. This also depicts varying efficiency in case of detection rates among GP and GP clinics across LGAs regardless of the number of available health resources and services or telehealth services, which can create cross-border detection.

Lope *et al. International Journal for Equity in Health*   (2022) 21:118

Page 9 of 17
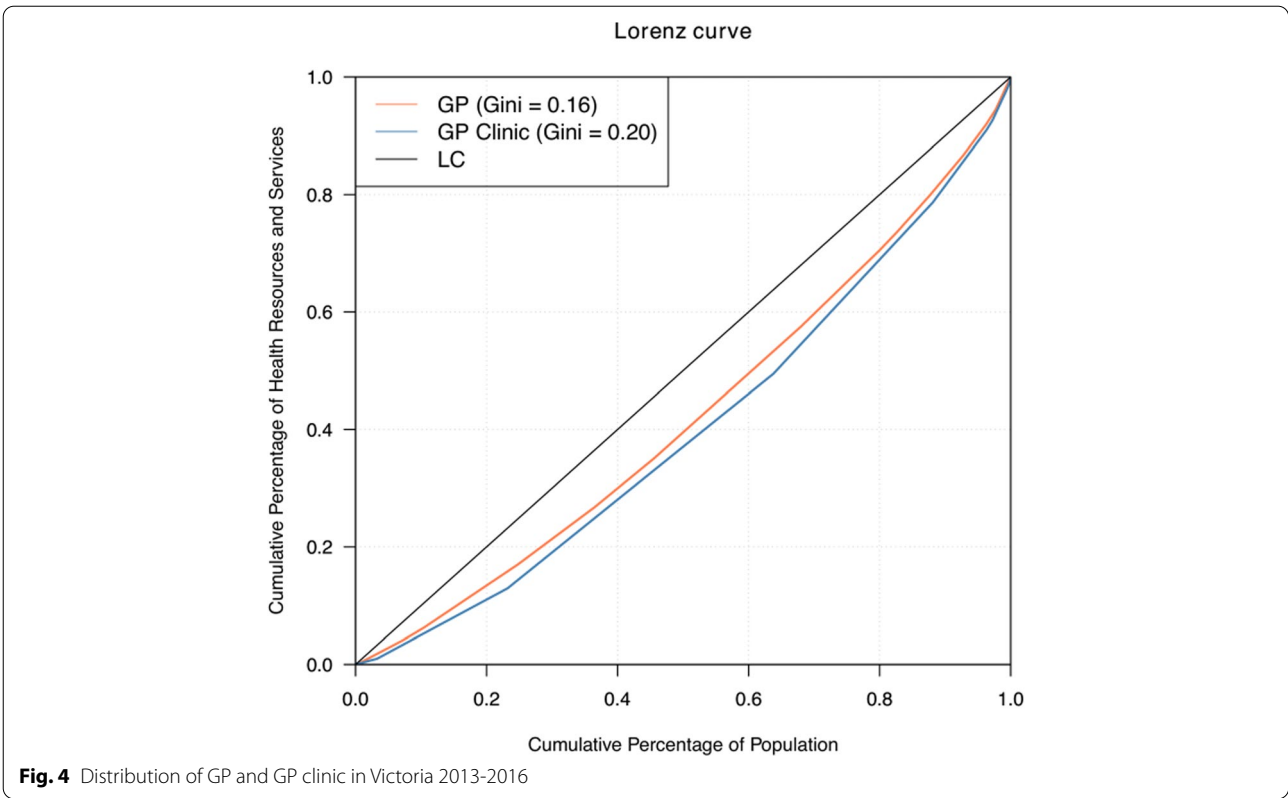
Fig. 4 Distribution of GP and GP clinic in Victoria 2013-2016

Fig. 5 Posterior model probabilities

Lope *et al. International Journal for Equity in Health*   (2022) 21:118

Page 10 of 17
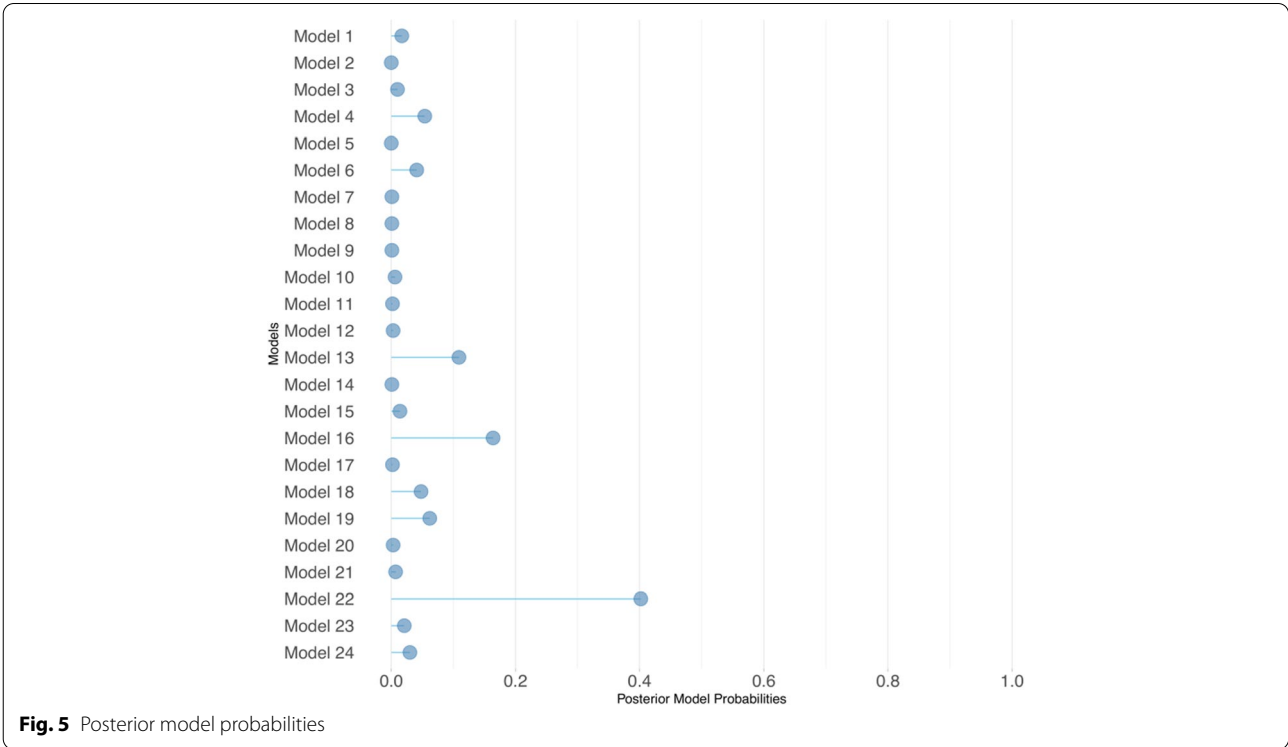
Table 1 5 Models streamlined during model selection

Model No.	Combination of $\lambda_{t,s}$ and π_s	$P(m\|z)$
Model 4	$\lambda_{t,s} = \log(\text{Pop}_s) + \alpha_0 + \alpha_{1.1}\,\text{seifa}_s + \alpha_{1.2}\,\text{seifa}_s + \alpha_2\,\text{hhsize}_s + \alpha_4\,\text{AccessDifficulty}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1\,\text{GP}_s + \beta_2\,\text{GPclinic}_s$	0.054
Model 13	$\lambda_{t,s} = \log(\text{Pop}_s) + \alpha_0 + \alpha_{1.1}\,\text{seifa}_s + \alpha_{1.2}\,\text{seifa}_s + \alpha_2\,\text{hhsize}_s + \alpha_3\,\text{homeless}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1\,\text{GP}_s + \beta_2\,\text{GPclinic}_s + \beta_3\,\text{AccessDifficulty}_s$	0.109
Model 16	$\lambda_{t,s} = \log(\text{Pop}_s) + \alpha_0 + \alpha_{1.1}\,\text{seifa}_s + \alpha_{1.2}\,\text{seifa}_s + \alpha_2\,\text{hhsize}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1\,\text{GP}_s + \beta_2\,\text{GPclinic}_s + \beta_3\,\text{AccessDifficulty}_s$	0.164
Model 19	$\lambda_{t,s} = \log(\text{Pop}_s) + \alpha_0 + \alpha_{1.1}\,\text{seifa}_s + \alpha_{1.2}\,\text{seifa}_s + \alpha_3\,\text{homeless}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1\,\text{GP}_s + \beta_2\,\text{GPclinic}_s + \beta_3\,\text{AccessDifficulty}_s$	0.062
Model 22	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_{1.1}\,\text{seifa}_s + \alpha_{1.2}\,\text{seifa}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1\,\text{GP}_s + \beta_2\,\text{GPclinic}_s + \beta_3\,\text{AccessDifficulty}_s$	0.402

Incidence and case detection rates

Figure 7 shows the comparison maps of the estimated incidence and observed notification of influenza in Victoria in 2013-2016, wherein the darker the colour depicts a higher number of influenza cases. Figures 8 and 9 show the comparison of predicted and observed influenza cases by year and LGA. For Fig. 9, the dark red colour means the predicted number of influenza cases in the area is higher than the observed number of cases. This indicates the possibility of having under-reported cases for that specific area.

From Fig. 7, we observe that the estimated incidence and actual notifications of influenza are correlated with highly recorded cases in the southern and eastern LGAs of the state. Further, the differences are very minimal. This depicts a good detection process among LGAs in Victoria considering the relatively equal distribution of health resources and services. Figure 7 shows the validity of the models used given the close fit of the predictions and actual notifications.

In Fig. 8, looking closely at the predicted cases in comparison to the observed cases of influenza by year, it is apparent that increasing case detection rate is observed through the years. Reviewing the potential LGAs with under-reported cases by year in Fig. 9, a changing pattern is recorded. In 2013, most of the LGAs were deemed to have under-reported cases, and 2014 seems to concentrate at the southwest and eastern parts, moving towards a more spread out under-reported cases in 2015-2016. Overall, the number of GP and GP clinics in the dataset from 2013-2016 is constant. The changing pattern in the number of cases detected across areas and years depicts either a generally increasing number of influenza cases or improvement in the case detection process in the state. Performing a more detailed look into the potential under-reported areas where consistently observed in most years in Fig. 9, potential areas for under reported cases in the southwest are Glenelg, Southern Grampians, Ararat, Moyne, Corangamite, Colac-Otway, Surf Coast and eastern such as Alpine, Towong and East Gippsland. These areas are recorded accordingly to have relatively lower the number of GP or GP clinic.

Discussion

This study demonstrated the impact of the degree of inequality in health resources and services into disease detection using two-component Bayesian modelling. Using an inequality measure in conjunction with Bayesian estimation provides a picture of the severity of underserved areas in terms of access to health resources and services in addition to its efficiency to detect disease cases as it happens by accounting for existing social factors. The Bayesian framework proposed in this study is a practical approach that does not require time-intensive and costly surveying of detailed geography levels such as LGA and expert opinion on disease rates, though acknowledges their importance. In addition, it provides an estimation approach that relies mainly on the observed data at hand.

In line with the relatively equal number of GP and GP clinics (with Gini coefficients of 0.16 and 0.20, respectively) across the LGAs in Victoria, Australia in 2013-2016, the estimated incidence of the influenza cases showed minimal discrepancy with the observed influenza cases. This result takes into account the social deprivation factors that can have an impact on the actual count of influenza incidence. Through the years, there has been an increasing trend in the number of influenza cases in the state. Looking closely at the areas, areas in eastern and southwestern parts of Victoria recorded potential under-reporting in most years. These areas accordingly are recorded to have

Fig. 6 Predictors of incidence and case detection rate
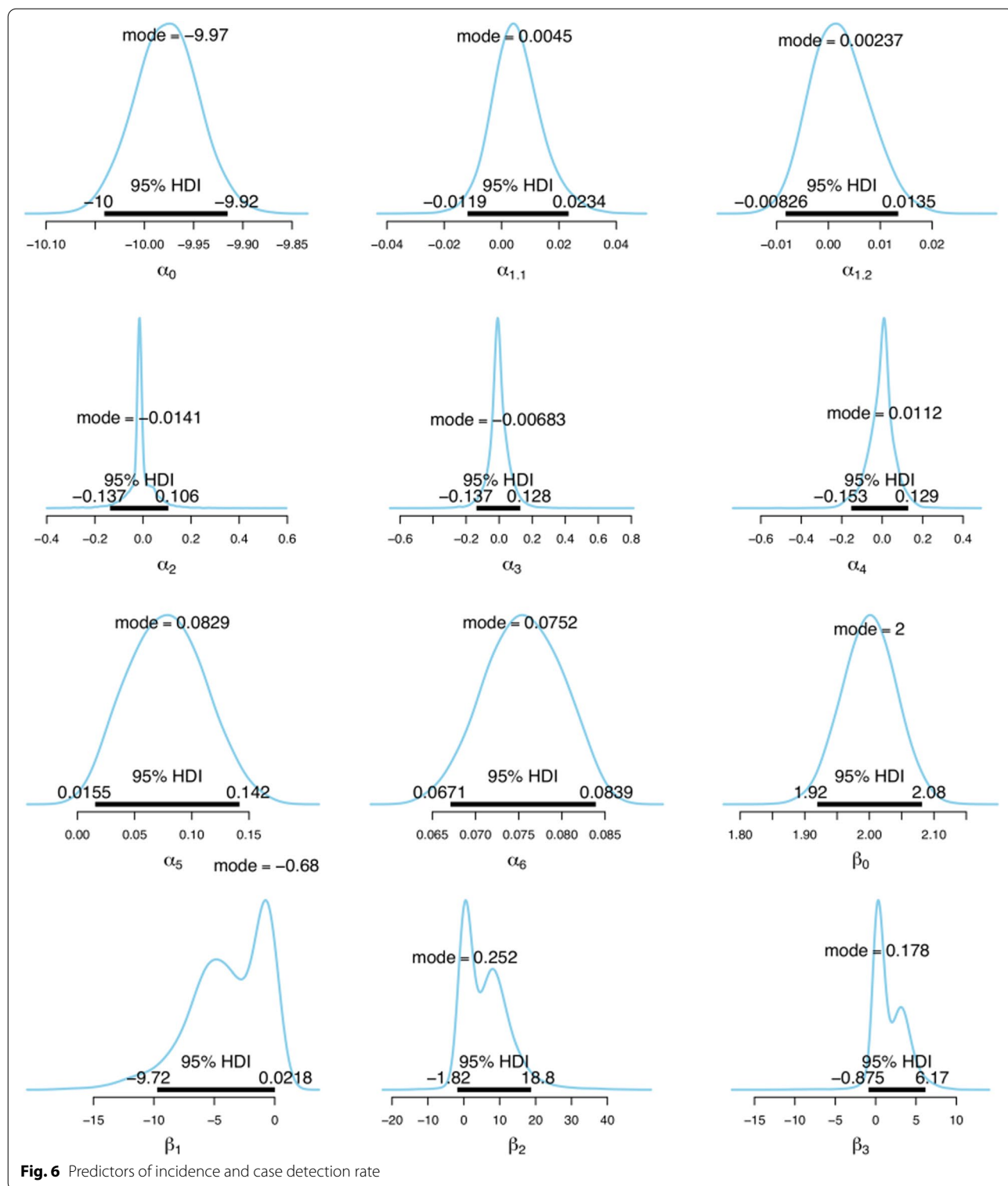
either lower number of GP or GP clinic. This result depicts where the potential increase is happening vis-à-vis availability of health resources, services, and other social factors which can provide insight into areas with the possibility of

an outbreak. Moreover, it identifies not only possible areas with a relatively lower number of GP and GP clinics that are unable to detect cases immediately, but also it explores the increase in estimated incidence rate that can happen

Fig. 7 Estimated influenza incidence and notification in Victoria 2013-2016 (per 100,000 population)
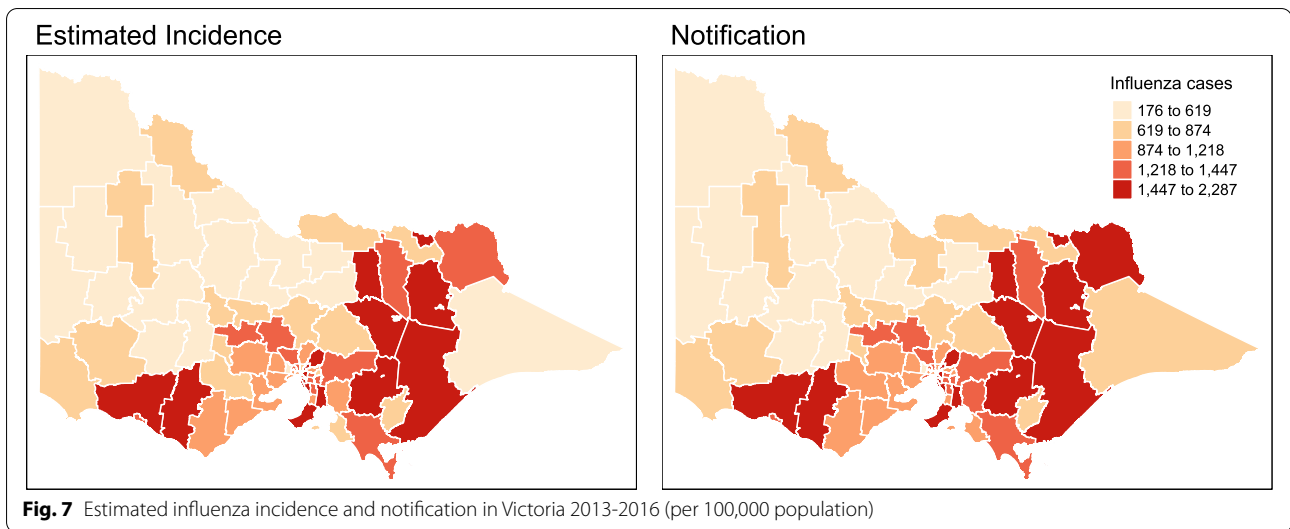
due to social deprivation factors such as homelessness and higher index of relative socio-economic disadvantage in an area and its neighbouring areas.

Use of model averaging in this study, which is an improvement over the studies of Shaweno et al. [14] and Stoner et al. [15], helped to determine the classification of the floating or uncategorised covariate access difficulty. This covariate can be seen to impact both the case detection probability and the incidence rate of influenza. Bayesian model averaging provided a model framework that lets a covariate influence both model components without

implementing fixed covariate categorisation. After running our Bayesian framework with 24 candidate models, the access difficulty was neither classified as a fixed covariate into the case detection probability component nor the incidence rate component. The posterior model probabilities suggest that all models have a contribution, and there is no need for the selection of one particular model or a fixed covariate categorisation. This finding provides an alternative and flexible modelling framework when any of the covariates seem to have overlapping effects between the case detection and incidence models.

Fig. 8 Observed and predicted influenza cases (per 100,000 population) 2013-2016
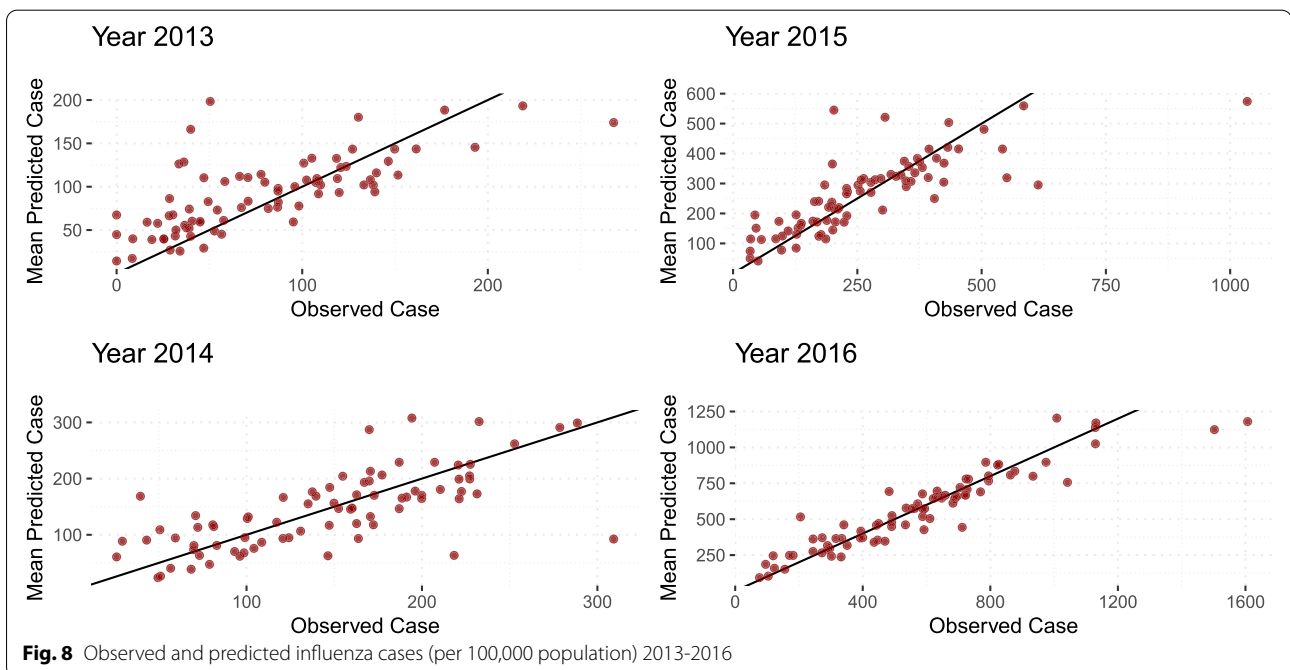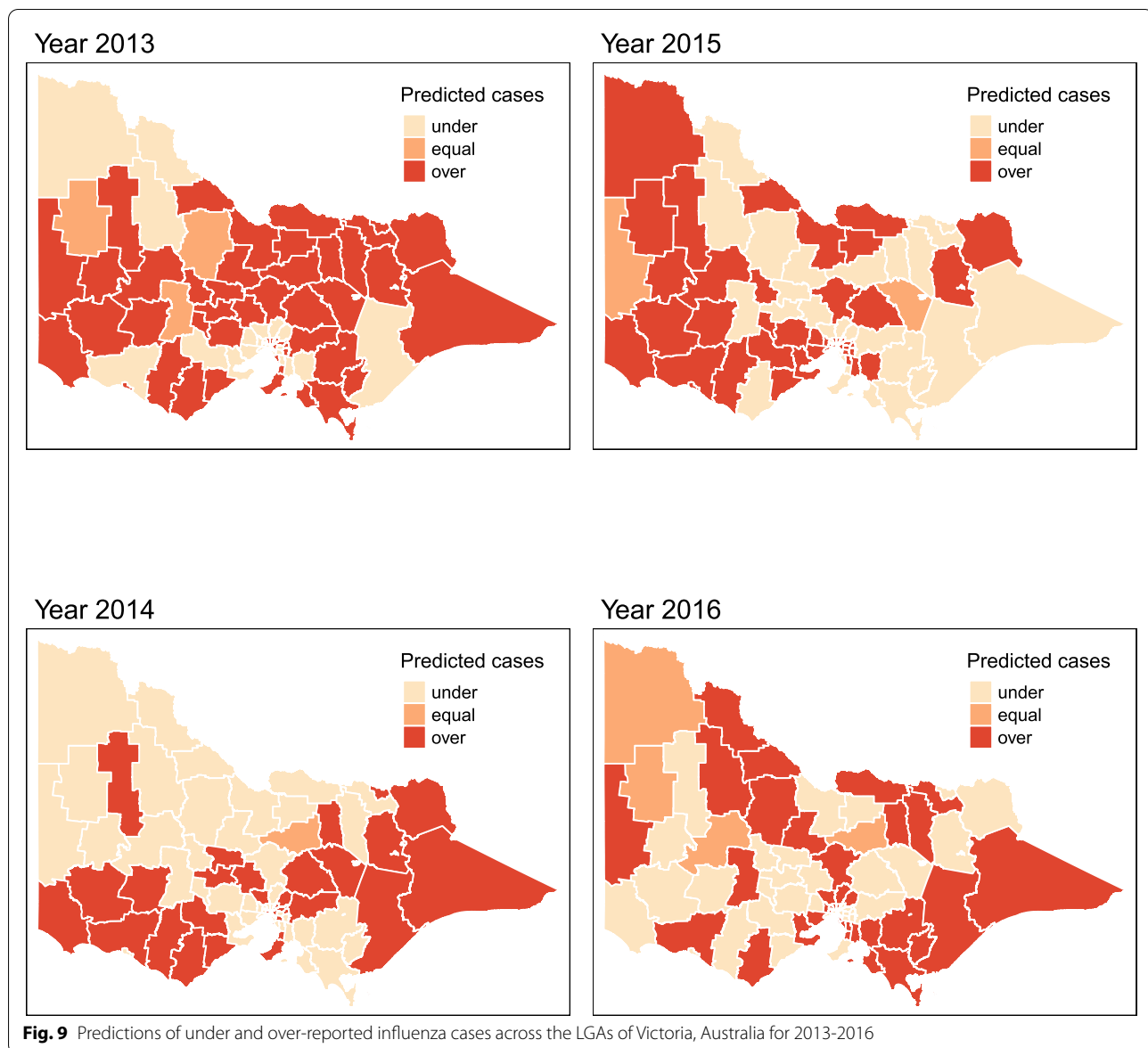
Fig. 9 Predictions of under and over-reported influenza cases across the LGAs of Victoria, Australia for 2013-2016

Conclusion

This study showed the connection of health inequality into a crucial matter of disease detection. The estimation of the impact of health inequality in terms of problems in disease detection can encourage policymakers to actively pursue equal access, especially in underserved areas and prevention in the spread of disease in the areas with potentially under-reported cases. Further, this practical and timely modelling using Bayesian estimation provides an alternative contemporary approach, especially during time-restricted situations such as an outbreak or unexpected surge of disease cases since it does not require costly and time-consuming surveying and preparations.

This hand-in-hand approach of using an inequality measure and Bayesian estimation mainly helped to identify finer level of areas with under-served health resources and services in addition to areas with possible under-reported disease cases. Moreover, by using two-component modelling Bayesian estimation for case detection probability and incidence rate, it distinguishes effects coming from the inequality of health resources and services or other social deprivation factors.

Lastly, this study demonstrated the use of Bayesian model averaging to enable researchers to eliminate the need for covariate categorisation, especially with overlapping covariate effects to both incidence and case detection models. The use of Bayesian model averaging

Model No.	$\lambda_{t,s}$ and π_s
Model 1	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_2 \text{ hhsize}_s + \alpha_3 \text{ homeless}_s + \alpha_4 \text{ AccessDifficulty}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s + \beta_2 \text{ GPclinic}_s$
Model 2	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_2 \text{ hhsize}_s + \alpha_3 \text{ homeless}_s + \alpha_4 \text{ AccessDifficulty}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_2 \text{ GPclinic}_s$
Model 3	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_2 \text{ hhsize}_s + \alpha_3 \text{ homeless}_s + \alpha_4 \text{ AccessDifficulty}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s$
Model 4	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_2 \text{ hhsize}_s + \alpha_4 \text{ AccessDifficulty}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s + \beta_2 \text{ GPclinic}_s$
Model 5	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_2 \text{ hhsize}_s + \alpha_4 \text{ AccessDifficulty}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + +\beta_2 \text{ GPclinic}_s$
Model 6	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_2 \text{ hhsize}_s + \alpha_4 \text{ AccessDifficulty}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s$
Model 7	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_3 \text{ homeless}_s + \alpha_4 \text{ AccessDifficulty}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s + \beta_2 \text{ GPclinic}_s$
Model 8	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_3 \text{ homeless}_s + \alpha_4 \text{ AccessDifficulty}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + +\beta_2 \text{ GPclinic}_s$
Model 9	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_3 \text{ homeless}_s + \alpha_4 \text{ AccessDifficulty}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s$
Model 10	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_4 \text{ AccessDifficulty}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s + \beta_2 \text{ GPclinic}_s$
Model 11	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_4 \text{ AccessDifficulty}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + +\beta_2 \text{ GPclinic}_s$
Model 12	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_4 \text{ AccessDifficulty}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s$
Model 13	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_2 \text{ hhsize}_s + \alpha_3 \text{ homeless}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s + \beta_2 \text{ GPclinic}_s + \beta_3 \text{ AccessDifficulty}_s$
Model 14	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_2 \text{ hhsize}_s + \alpha_3 \text{ homeless}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_2 \text{ GPclinic}_s + \beta_3 \text{ AccessDifficulty}_s$
Model 15	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_2 \text{ hhsize}_s + \alpha_3 \text{ homeless}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s + \beta_3 \text{ AccessDifficulty}_s$
Model 16	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_2 \text{ hhsize}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s + \beta_2 \text{ GPclinic}_s + \beta_3 \text{ AccessDifficulty}_s$
Model 17	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_2 \text{ hhsize}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_2 \text{ GPclinic}_s + \beta_3 \text{ AccessDifficulty}_s$
Model 18	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_2 \text{ hhsize}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s + \beta_3 \text{ AccessDifficulty}_s$
Model 19	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_3 \text{ homeless}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s + \beta_2 \text{ GPclinic}_s + \beta_3 \text{ AccessDifficulty}_s$
Model 20	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_3 \text{ homeless}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_2 \text{ GPclinic}_s + \beta_3 \text{ AccessDifficulty}_s$
Model 21	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \alpha_3 \text{ homeless}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s + \beta_3 \text{ AccessDifficulty}_s$
Model 22	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s + \beta_2 \text{ GPclinic}_s + \beta_3 \text{ AccessDifficulty}_s$
Model 23	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_2 \text{ GPclinic}_s + \beta_3 \text{ AccessDifficulty}_s$
Model 24	$\lambda_{t,s} = \log(\text{Pop}_{t,s}) + \alpha_0 + \alpha_1 \text{ seifa}_s + \phi_s + \theta_s + f(t)$ $\pi_s = \beta_0 + \beta_1 \text{ GP}_s + \beta_3 \text{ AccessDifficulty}_s$

Fig. 10 24 Candidate Models. Here $f(t) = a_5 t + a_6 t^2$

Fig. 11 Model Diagram

Fig. 12 Bayesian Model Averaging (BMA)

provides the inclusion of all candidate Bayesian models' effects to the predictions and allows the models to find their chance during implementation.

To note some limitations of this study, the time frame of this cross-sectional data analysis is only 2013-2016. Due to the five-year interval of the census data collection, some covariates are not observed for each year; hence, the same information is repeated across the years of study. Nonetheless, as a future study, exploring the application of our Bayesian modelling approach to other types of diseases and the inclusion of other disease-specific covariates can be considered. It would be helpful to explore the impact of various degrees of health inequality on the disease detection estimates. Doing so can provide ideal and achievable resource allocation without compromising the delivery of an efficient disease detection process.

Appendix A: additional figures

Figure A1 lists the candidate models that are implemented in the study. Figure A2 shows the diagram of the Bayesian model, and Figure A3 outlines the proposed Bayesian model averaging approach.

Authors' contributions
D.J.L. did the conceptualization, developed the methodology and software, carried on the formal analysis, did the data curation, and wrote the original manuscript. H.D. did the conceptualization, developed the methodology and software, and wrote the original manuscript. A.D. did the conceptualization and developed the methodology. All authors read and approved the final manuscript.

Lope *et al. International Journal for Equity in Health*   (2022) 21:118

Page 17 of 17

Declarations

Ethics approval and consent to participate
Not Applicable

Consent for publication
Not Applicable

Competing interests
The authors declare that they have no competing interests.

References

1. World Health Organization. The world health report 2000. 2000. https://apps.who.int/gb/archive/pdf_files/WHA53/ea4.pdf. Accessed 07 June 2021.
2. Australian Institute of Health and Welfare. Australia's Health 2018. 2018. https://www.aihw.gov.au/reports/australias-health/australias-health-2018-in-brief. Accessed 07 June 2021.
3. Wilkinson D, Symon B. Inequitable distribution of general practitioners in australia: estimating need through the robin hood index. Aust N Z J Public Health. 2000; 24(1):71–75.
4. Hann M, Gravelle H. The maldistribution of general practitioners in england and wales: 1974–2003. Br J Gen Pract. 2004; 54(509):894–8.
5. Theodorakis PN, Mantzavinis GD, Rrumbullaku L, Lionis C, Trell E. Measuring health inequalities in albania: a focus on the distribution of general practitioners. Hum Resour Health. 2006; 4(1):5.
6. Toyabe S. -i.Trend in geographic distribution of physicians in japan. Int J Equity Health. 2009; 8(1):5.
7. Matsumoto M, Inoue K, Farmer J, Inada H, Kajii E. Geographic distribution of primary care physicians in japan and britain. Health Place. 2010; 16(1):164–6.
8. Ohba H, Narumi M, Hosokawa Y, Aoki M. Measuring the inequalities in radiotherapy health resources in japan: comparison of the hokkaido-tohoku and tokyo districts. Jpn J Radiol. 2010; 28(1):20–6.
9. Liu Y, Jiang Y, Tang S, Qiu J, Zhong X, Wang Y. Analysis of the equity of emergency medical services: a cross-sectional survey in chongqing city. Int J Equity Health. 2015; 14(1):150.
10. Zhang T, Xu Y, Ren J, Sun L, Liu C. Inequality in the distribution of health resources and health services in china: hospitals versus primary care institutions. Int J Equity Health. 2017; 16(1):42.
11. Center for Disease Control and Prevention. Public Health Surveillance, Preparing for the Future. 2018. https://www.cdc.gov/surveillance/pdfs/Surveillance-Series-Bookleth.pdf. Accessed 07 June 2021.
12. Moa AM, Adam DC, MacIntyre CR. Inter-seasonality of influenza in australia. Influenza Other Respir Viruses. 2019; 13(5):459–64.
13. Geoghegan JL, Saavedra AF, Duchene S, Sullivan S, Barr I, Holmes EC. Continental synchronicity of human influenza virus epidemics despite climactic variation. PLoS Pathog. 2018;14(1):e1006903.
14. Shaweno D, Trauer JM, Denholm JT, McBryde ES. A novel bayesian geospatial method for estimating tuberculosis incidence reveals many missed tb cases in ethiopia. BMC Infect Dis. 2017; 17(1):1–8.
15. Stoner O, Economou T, da Silva GDM. A hierarchical framework for correcting under-reporting in count data. J Am Stat Assoc. 2019;114(528):1481–92.
16. Zipfel CM, Colizza V, Bansal S. Health inequities in influenza transmission and surveillance. PLoS Comput Biol. 2021; 17(3):1008642.
17. Brown MC. Using gini-style indices to evaluate the spatial patterns of health practitioners: theoretical considerations and an application based on alberta data. Social Sci Med. 1994; 38(9):1243–56.
18. Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: a tutorial (with comments by m. clyde, david draper and ei george, and a rejoinder by the authors. Stat Sci. 1999; 14(4):382–417.
19. National Notifiable Diseases Surveillance System. Reports. http://www9.health.gov.au/cda/source/cda-index.cfm. Accessed 07 June 2021.
20. Government of Victoria - Department of Health and Human Services. Local Government Area (LGA) profiles data 2015 for VIC. 2016. https://portal.aurin.org.au. Accessed 07 June 2021.
21. TUA_PHIDU. LGA15 Access to Services - 2014. 2017. https://portal.aurin.org.au. Accessed 07 June 2021.
22. Australian Bureau of Statistics. Regional population. 2022. https://www.abs.gov.au/statistics/people/population/regional-population/2021. Last accessed 16 August 2022.
23. Kruschke J. Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan, 2nd ed.New York: Academic Press; 2014.
24. Justel A, Peña D. Gibbs sampling will fail in outlier problems with strong masking. J Comput Graph Stat. 1996; 5(2):176–89.
25. Osborne JW. Eight: improving the normality of variables through Box-Cox transformation: debunking the myth of distributional irrelevance. In: Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data. New York: SAGE Publications: 2013. p. 169–190.
26. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Stat Math. 1991; 43(1):1–20.
27. Sullivan SG, Pennington K, Raupach J, Franklin LJ, Bareja C, de Kluyver R, Do H. A summary of influenza surveillance systems in australia, 2015. Commun Dis Intell Canberra. 2016;1–52.
28. Pettitt AN, Weir IS, Hart AG. A conditional autoregressive gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data. Stat Comput. 2002; 12(4):353–67.
29. de Freitas Severino MT. Extending jags for spatial data. Depatamento de Estatistica; 2018.

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