Supplementary Information for

Role of Multi-resolution Vulnerability Indices in COVID-19 Spread in India: A Bayesian Model-based Analyses

Rupam Bhattacharyya[1], Anik Burman[2], Kalpana Singh[3], Sayantan Banerjee[4], Subha Maity[5], Arnab Auddy[6], Sarit Kumar Rout[7], Supriya Lahoti[8], Rajmohan Panda[8, *, #], Veerabhadran Baladandayuthapani[1, *, $]

[1]Department of Biostatistics, University of Michigan, Ann Arbor, USA

[2]Indian Statistical Institute, Kolkata, India

[3]Hamad Medical Corporation, Doha, Qatar

[4]Operations Management & Quantitative Techniques Area, Indian Institute of Management, Indore, India

[5]Department of Statistics, University of Michigan, Ann Arbor, USA

[6]Department of Statistics, Columbia University, New York, USA

[7]Indian Institute of Public Health, Bhubaneswar, Public Health Foundation of India, India

[8]Public Health Foundation of India, New Delhi, India

[*]Co-corresponding authors

[#]Plot No. 47, Sector 44, Institutional Area Gurugram 122002, India; raj.panda@phfi.org

[$]Department of Biostatistics, 1415 Washington Heights, Ann Arbor, Michigan 48109-2029, USA; veerab@umich.edu

Bhattacharyya R, *et al. BMJ Open* 2022; 12:e056292. doi: 10.1136/bmjopen-2021-056292

1. SUPPLEMENTARY TEXT

1.1 Data collation and cleanliness

Survey-based data are available at a nationwide scale and quality checks are performed before they are shared publicly. After getting the data downloaded from the sources mentioned in Supplementary Table 1, we have further performed quality checks of our own. We have accounted for the complex survey design in all analyses by adjusting standard errors for clustering and incorporating sampling weights. The details of maintaining data quality and processing at the survey level are provided in individual reports from the agencies, such as the National Family Health Survey 2015-16 (NFHS-4) report.[1] The data processing involved office editing, data entry using CSPro software, verification of data entry, secondary editing, and final cleaning of data at IIPS.

1.2 Computation of relative COVID-19 vulnerability indices

We adapt the algorithm by Acharya and Porwal, 2020 for the calculation of the relative vulnerability indices for each district.[2] The algorithm consists of three steps, as summarized in the left panel of Figure 2 in the main paper. In the Step 1, ranks are assigned to the values of a particular indicator for different districts of the state. The ranking is done in a manner so that the ‘riskier’ the value is (relative to the other values corresponding to the same indicator), the higher is the rank assigned, i.e., if the value of the indicator for a certain district suggest minimum risk, then it is given the least rank 1, and if the value suggests maximum risk, then it is ranked the highest. Using these ranks, the *indicator-specific relative cVI* for a district is computed (Equation 1). In Step 2 we calculate *theme-specific cVI* by aggregating the *indicator-specific vulnerability index* of the indices under that particular theme (Equation 2). In the Step 3, the *overall vulnerability index (oVI)* is calculated by aggregating the *theme-specific cVI’s* (Equation 3).

Bhattacharyya R, *et al. BMJ Open* 2022; 12:e056292. doi: 10.1136/bmjopen-2021-056292

Equation 1

For each of the n districts of interest, and for a particular indicator denoted by I_j, let its values for the n districts be $\{x_{1j}, x_{2j}, \ldots, x_{nj}\}$. Now, each of these x_{ij} values are assigned ranks R_{iI_j} in such a way so that the 'riskier' the value is (relative to the other values corresponding to the same indicator), the higher is the value of R_{iI_j}. Thus, for the set of n values of the indicator across the n districts, let the ranks assigned be $\{R_{1I_j}, R_{2I_j}, \ldots, R_{nI_j}\}$. Now, using these indicator-specific ranks for I_j, we define the *indicator-specific relative vulnerability index* for a district D_i as:

$$cVI_{iI_j} = \frac{R_{iI_j} - 1}{n-1}.$$

The definition of cVI_{iI_j} implies that it takes a value between 0 and 1. As mentioned previously, the ranks have been assigned in such a way so that higher the value of cVI_{iI_j}, cVI_{ij}, the more vulnerable that district is w.r.t to I_j relative to other districts; i.e., a value of $cVI_{iI_j} = 1$ makes that district the most vulnerable one and a value of 0 suggests that the district is least vulnerable relative to other districts. In case of ties, the minimum value of R_{iI_j} among the ties is assigned to the districts having the tied values.

Equation 2

For a particular theme T_j, let the indicators associated with it be $\left\{I_{n_{j_1}}, I_{n_{j_2}} \ldots, I_{n_{j_{k_j}}}\right\}$. After calculating the indicator-wise relative vulnerability index for the indicators grouped under T_j (using Equation 1), we now add up those relative vulnerability index values $\left\{cVI_{iI_{n_{j_1}}}, cVI_{iI_{n_{j_2}}}, \ldots, cVI_{iI_{n_{j_{k_j}}}}\right\}$ for a particular district D_i and define new variables t_{ij} as:

$$t_{ij} = \sum_{m=1}^{k_j} cVI_{iI_{n_{j_m}}}.$$

Thus, for theme T_j, we have values $\{t_{1j}, t_{2j}, \dots, t_{nj}\}$. Now we assign ranks R_{iT_j} to them such that the higher the value of t_{ij}, higher is the rank assigned and vice-versa. Based on these, we then define the *theme-wise relative vulnerability index* for theme T_j and district D_i as:

$$cVI_{iT_j} = \frac{R_{iT_j} - 1}{n-1}.$$

Here, a similar interpretation is possible as in using the same definition used for cVI_{iI_j} with I_j replaced by T_j.

Equation 3

For M different themes, we add up the individual indices. That is, once all the *theme-wise relative vulnerability indices* $\{cVI_{iT_1}, cVI_{iT_2}, \dots, cVI_{iT_M}\}$ are available, we aggregate the theme-wise indices for a particular district D_i and define new variables o_{ij} as:

$$o_{ij} = \sum_{m=1}^{M} cVI_{iT_m}.$$

Here, we assign and then use these values for obtaining increasing ranks R_{ij}. The same process is then applied to o_{ij} to compute the *overall relative vulnerability index* oVI_{ij} as the final aggregate across the theme-specific indices:

$$oVI_{ij} = \frac{R_{ij} - 1}{n-1}.$$

Here, we can also interpret oVI_{ij} like cVI_{iI_j}or cVI_{iT_j}.

1.3 Example of computing cVIs for a given region

We exemplify our steps for the computations of the vulnerability indices at the covariate and theme levels for the Mayurbhanj district of Odisha. Similar steps have been adopted for the other districts. First, we focus on the covariate named "General no. of beds per 10k." As per our data (available at the repository https://github.com/bayesrx/COVID_vulnerability_India), the two districts having the least number of general beds per 10k are Debagarh (0.0104 per 10k) and Boudh (0.0169 per

10k), and the two districts having the highest number of general beds per 10k are Ganjam (0.0943 per 10k) and Mayurbhanj (0.0826 per 10k). We rank the districts such that a higher rank puts a particular district in a more vulnerable position than a district with a lower rank. Since less number of general beds per unit of population indicates more risk or vulnerability, the ranking for this particular variable is in the decreasing order – i.e. the lower the number of beds per 10k for a district, the lower it features in the rank list and gets a higher numeric rank. For example, as Mayurbhanj has the second-highest number of beds per 10k, it will get a rank of 2. As per the covariate-specific vulnerability index formula, the VI for Mayurbhanj corresponding to the above covariate will be (2-1)/(30-1) = 0.0345. In the same manner, we compute the VIs for other covariates for Mayurbhanj. If, on the other hand, we were focusing on a variable for which a higher value indicates higher risk or vulnerability, the rank assigned to the district of Mayurbhanj with everything else unchanged would have been (30 - 2 + 1) = 29.

After computing the covariate-specific VIs in this way for all the covariates and all the districts, we compute one theme-specific VI. Let us consider the theme "Preparedness of COVID" for the purpose of illustration here. This theme comprises of the following covariates: 1) Total Beds Capacity per 10k, 2) Total ICU Beds per 10k, 3) Temporary medical camps Rural, 4) Temporary medical camps Urban, and 5) COVID hospital testing center. The covariate-specific VIs of these five covariates for Mayurbhanj are 0.7931, 0.6207, 0.1724, 0.6897, and 0.1034, which sum up to 2.3793. It is clear that the higher the above sum for a district, the more vulnerable it will potentially be for COVID preparedness. Mayurbhanj has the eighth highest value when ordering this total across the districts, with the two highest values belonging to the districts Boudh (3.2069) and Gajapati (3.1724), and the two lowest values belonging to the districts Puri (1.0689) and Koraput (0.9310). Thus, Mayurbhanj is assigned the rank (30-8+1) = 23, since a higher sum here indicates higher vulnerability). The target theme-specific VI for Mayurbhanj thus turns out to be (23-1)/(30-1) = 0.7586.

Once we repeat the above procedure for all the themes and all the districts, we can compute the overall VI for Mayurbhanj. The values of the five theme-specific VIs for Mayurbhanj are 0.7586, 0.9655, 0.0345, 0.75862069, and 0.7241, which add up to 3.2413. This sum is the fourth-highest value among the values of the sum of theme-specific VIs across the different districts. Thus, similar to the ranking in the case of theme-specific VI computation, the rank assigned to Mayurbhanj will be (30-4+1) = 27. Hence, the overall VI for Mayurbhanj will be (27-1)/(30-1) = 0.8966.

1.4 Estimation and summarization of time-varying reproductive number (R)

For the i^{th} location of interest (a district, in our case) the input to the EpiEstim package is represented by $\left\{I^i(t), t \in \{1, \dots, T\}\right\}$ where $I^i(t)$ indicates the incidence (count of new COVID-19 cases) in that location on day t, and T denotes the total number of days for which this data is available. Using this input, a Bayesian estimation procedure utilizing user-supplied mean and standard deviation for the serial interval distribution yields a time series of effective reproduction number estimates $\left\{R^i(t), t \in \{1, \dots, T\}\right\}$. This procedure is performed for each location separately across $i \in \{1, \dots, n\}$, say.

The details of the parameter and function choices for the estimation procedure are summarized in Supplementary Table 2. Instead of using the full time series of $\left\{R^i(t), t \in \{1, \dots, T\}, i \in \{1, \dots, n\}\right\}$ as a response variable, we decided to compute scalar summaries of the time series at the district levels. We compute two scalar summaries, namely, a 14-day average of estimated time-varying R, called the instantaneous R (iR) (Equation 4), and the first principal component of the time-varying R, called the variability in R (vR) (Equation 5).

Equation 4

We summarize the location information the time-varying R profiles by taking a mean of the profile at a location level over a specified period of estimation and call this summary the instantaneous R (iR).

$iR_K^i = \frac{\sum_{t=T-K+1}^{T} R(t)}{K}, i \in \{1, \ldots, n\}.$

The normality for this metric was checked visually via density and quantile-quantile plots, as exhibited in Supplementary Figure 4.

Equation 5

Two estimated time-varying R profiles with the same computed iR may represent potentially different trajectories of the pandemic depending on how dispersed the actual profile is around that representative value. Therefore, in order to obtain a summary of the variability in the R profiles, we performed a principal component analysis (PCA) on $\left\{R^i(t), t \in \{T-K+1, \ldots, T\}, i \in \{1, \ldots, n\}\right\}$.[3] Denoting the matrix of these estimated R values as $\mathrm{P}_{n\times K}, \mathrm{P}_{it} = R^i(t), t \in \{T-K+1, \ldots, T\}, i \in \{1, \ldots, n\}$, the PCA is performed on P and the weight matrix ($W_{K\times K}$) is found as the matrix of eigenvectors of the covariance matrix of P. Then, we define the variability in R as the first principal component.

$vR_K^i = (\mathrm{P}W)_{i1}, i \in \{1, \ldots, n\}.$

The details regarding the values of n, K, M and other choices are summarized in Supplementary Table 2.

1.5 Regression analyses using summaries of time-varying R profiles

Our response variable is denoted by $iR_K^i, i \in \{1, \ldots, n\}$. We simplify the notations for the cVIs a bit first. For a particular theme T_j, the indicator-level covariates (cVIs) are denoted by $cVI_{\mathrm{iq}}^j, i \in \{1, \ldots, n\}, q \in \{1, \ldots, k_j\}$ where k_j is the number of indicators available for that particular theme.

Across the M available themes, the theme-level cVIs are denoted by $cVI_{ij}, i \in \{1, \dots, n\}, j \in \{1, \dots, M\}$. Using these, we respectively fit one overall model (Equation 6) and M theme-specific models (Equation 7) with the iR values as the response.

Equation 6

The overall model looks like the following.

$$iR_K^i = \beta_0 + \sum_{j=1}^{M} \beta_j \, cVI_{ij} + \epsilon_i, i \in \{1, \dots, n\},$$

$$where\ \epsilon_i \sim N(0, \sigma^2)\ iid\ across\ i, \beta_0\ and\ \beta_j s\ jointly\ follow\ N_{M+1}(\mathbf{0}, g(\sigma^2 V^T V)^{-1}),$$

$$i.e. the\ Zellner's\ g - prior, V\ being\ the\ full\ design\ matrix,$$

$$\log(\sigma) \sim U(\mathbb{R}), g = n.$$

Equation 7

The theme-specific indicator-level model for the j^{th} theme looks like the following.

$$iR_K^i = \beta_0^j + \sum_{q=1}^{k_j} \beta_q^j \, cVI_{\text{iq}}^j + \epsilon_i^j, i \in \{1, \dots, n\},$$

$$where\ \epsilon_i^j \sim N\left(0, \sigma_j^2\right)\ iid\ across\ i, \beta_0^j\ and\ \beta_q^j s\ jointly\ follow\ N_{k_j+1}\left(\mathbf{0}, g\left(\sigma_j^2 V_j^T V_j\right)^{-1}\right),$$

$$i.e. the\ Zellner's\ g - prior, V_j\ being\ the\ full\ design\ matrix,$$

$$\log\left(\sigma_j\right) \sim U(\mathbb{R}), g = n.$$

For the vR regressions, our response variable is denoted by $vR_K^i, i \in \{1, \dots, n\}$. Using the same notation for the cVIs as before, we respectively fit one overall model (Equation 8) and M theme-specific models (Equation 9) with the vR values as the response.

Equation 8

The overall model looks like the following.

$$vR_K^i = \gamma_0 + \sum_{j=1}^{M} \gamma_j \, cVI_{ij} + \delta_i, i \in \{1, \dots, n\},$$

$$where\ \delta_i \sim N(0, \tau^2)\ iid\ across\ i, \gamma_0\ and\ \gamma_j s\ jointly\ follow\ N_{M+1}(\mathbf{0}, g(\tau^2 V^T V)^{-1}),$$

$i.e.\, the\ Zellner's\ g-prior, V\ being\ the\ full\ design\ matrix,$

$\log(\tau) \sim U(\mathbb{R}), g = n.$

Equation 9

The theme-specific indicator-level model for the j^{th} theme looks like the following.

$vR_K^i = \gamma_0^j + \sum_{q=1}^{k_j} \gamma_q^j\, cVI_{\mathrm{iq}}^j + \delta_i^j, i \in \{1, \dots, n\},$

$where\ \delta_i^j \sim N\left(0, \tau_j^2\right)\ iid\ across\ i, \gamma_0^j\ and\ \gamma_q^j s\ jointly\ follow\ N_{k_j+1}\left(\mathbf{0}, g\left(\tau_j^2 V_j^T V_j\right)^{-1}\right),$

$i.e.\, the\ Zellner's\ g-prior, V_j\ being\ the\ full\ design\ matrix,$

$\log\left(\tau_j\right) \sim U(\mathbb{R}), g = n.$

We use the R package BMS with default choices for the bms function for fitting our models which performs a Bayesian model averaging procedure (BMA) to obtain posterior estimates, credible intervals and posterior inclusion probabilities (PIPs) for the parameters of interest.[4] Briefly, BMA performs a search across the weighted posterior probabilities of the plausible models and selects the model with the highest posterior probability.

1.6 The Bayesian model averaging approach and computation of posterior inclusion probabilities

Following Zeugner and Feldkircher, 2015, we describe the Bayesian model averaging (BMA) procedure implemented in the BMS R package utilized in context of our regression modeling framework.[3] Recall the general form of our models as described previously in Equation 6. Our response variable is denoted now by $iR^i, i \in \{1, \dots, n\}$. Assuming that there are M available themes, the theme-level cVIs are denoted by $cVI_{ij}, i \in \{1, \dots, n\}, j \in \{1, \dots, M\}$. The across-theme overall vulnerability model can then be written as the following.

$$iR^i = \beta_0 + \sum_{j=1}^{M} \beta_j\, cVI_{ij} + \epsilon_i, i \in \{1, \dots, n\},$$

$$\epsilon_i \sim N(0, \sigma^2)\ iid\ across\ i, (\beta_0, \beta_1, \dots, \beta_M)^T \sim N_{M+1}(\mathbf{0}, g(\sigma^2 V^T V)^{-1}),$$

$$V_{(ij)} = cVI_{ij}, \log(\sigma) \sim U(\mathbb{R}), g = n.$$

Here βs are the model coefficients of interest, σ is the standard deviation of the error distribution, g is the scaling parameter of the Zellner's g-prior taken to be equal to the sample size in all our analyses, and $U(\mathbb{R})$ denotes an improper uniform distribution over the real number line. In context of this model, then, BMA estimates models for all possible combinations of the M vulnerabilities of interest by constructing a weighted average on them. This means that the procedure estimates $\gamma = 1, \dots, 2^M$ many models (since each covariate may or may not be included in a model), where the model weights can be computed as the following using Bayes' theorem.

$$p\left(Model_\gamma \middle| (iR^1, \dots, iR^n)^T, V\right) = \frac{p\left((iR^1, \dots, iR^n)^T \middle| Model_\gamma, V\right) p(Model_\gamma)}{p((iR^1, \dots, iR^n)^T | V)} =$$

$$\frac{p\left((iR^1, \dots, iR^n)^T \middle| Model_\gamma, V\right) p(Model_\gamma)}{\sum_{s=1}^{2^M} p((iR^1, \dots, iR^n)^T | Model_s, V) p(Model_s)}.$$

Once we have computed these posterior model probabilities (PMPs), it is then possible to compute the model-weighted posterior distributions for any parameters or functions thereof. For example, if we are interested in the posterior of the coefficient β_1 corresponding to the first themed cVI, we can compute the following.

$$p(\beta_1 | (iR^1, \dots, iR^n)^T, V) = \sum_{s=1}^{2^M} p(\beta_1 | Model_s, (iR^1, \dots, iR^n)^T, V) p(Model_s | (iR^1, \dots, iR^n)^T, V).$$

Similar to the posterior distributions of the parameters, we can compute the posterior inclusion probabilities (PIPs) for a variable by summing up the PMPs for all models out of the 2^M where that variable was included. For example, the PIP for the first themed cVI can be computed in the following way.

$$PIP_1 = \sum_{s=1}^{2^M} p(Model_s | (iR^1, \dots, iR^n)^T, V)\, I(\beta_1 \neq 0 | Model_s).$$

Bhattacharyya R, *et al. BMJ Open* 2022; 12:e056292. doi: 10.1136/bmjopen-2021-056292

Similar calculations can then be extended to all our theme-specific models and models using variability in R (vR) as outcome instead of iR. For the model priors $p(Model_\gamma)$, we use the default choice of setting $p\left(Model_\gamma\right) \propto 1$ i.e., uniform priors due to the lack of additional knowledge.

1.7 Simulation study to assess the selection performance of posterior inclusion probabilities computed using the BMS package

We perform a set of ground-truth simulations to mimic scenarios similar to the data structure used in our analyses to assess the performance of the Bayesian model averaging procedure described above. Fixing a sample size n and number of covariates p, we first generate the design matrix $\boldsymbol{X}_{n\times p}$ where each element follows a $U(0,1)$ distribution independently. We choose this distribution since the range of the cVIs which serve as covariates in our COVID-19 data fall between [0, 1]. Then, we set 100a% of these p covariates (pa many) to have a true (non-zero) effect on the outcome and the rest ($p(1-a)$ many) to have no effect on the outcome. In essence, the tuning parameter a controls the sparsity of the true signals. The non-zero coefficients (βs) are generated independently from $U(b, b+1)$ distribution (to include different effect sizes in low, medium, and high ranges) – to cover a range of corresponding associations. These βs are then each multiplied by independent random variables taking values ± 1 with probability $\frac{1}{2}$, to include effects of mixed signs in a single scenario – to include both positive and negative associations. The outcome vector $\boldsymbol{Y}_{n\times 1}$ is then generated as $\boldsymbol{X}_{n\times p}\boldsymbol{\beta}_{p\times 1} + \boldsymbol{\epsilon}_{n\times 1}$, where the ϵ_is are iid $N(0, s^2)$ to control the spread of the noise in the generating model. The noise standard deviation s controls the signal-to-noise ratio for a fixed absolute value of the coefficients.

We investigate all possible combinations of the following values for the model generation inputs: n = 30 (akin to our real dataset), 60, 90 (low, medium and high sample sizes); a = 0.2, 0.4, 0.6, 0.8 (sparse to dense true signals); b = 0.5, 1, 1.5 (low to high absolute values of the true effects); s =

Bhattacharyya R, *et al. BMJ Open* 2022; 12:e056292. doi: 10.1136/bmjopen-2021-056292

1, 2, 3 (high to low signal-to-noise ratio for a fixed absolute coefficient). The number of covariates p is fixed at 10 for all scenarios, since the number of covariates in our vulnerability models do not exceed this, and this is the scale at which we are interested in assessing the performance of the BMA procedure. Each simulation scenario is iterated 100 times and model evaluation metrics are computed and summarized across the 100 simulations for each scenario. We compute one threshold-free metric of model selection performance based on the PIPs, namely, the AUC (area under the receiver operating characteristics curve). This metric is intended to provide summary-level performances of the PIPs as continuous quantifications of variable importance as we use them in our analyses. Additionally, we perform threshold-specific variable selection by computing a selection cut-off based on the PIPs by performing false discovery rate (FDR) control at a 10% level treating 1-PIPs as p-value type quantities.[4] We compute the observed FDR, the true positive rate (TPR), and the false positive rate (FPR) at the threshold thus computed. The results from the simulation studies are summarized in Supplementary Table 5 and Supplementary Figures 5-8.

Some evident patterns emerge from these summaries. First, the AUCs show steady increase with increase in sample size in all scenarios – in particular, for less noisy data, an effect size of one with medium sample size ensures near-perfect AUC for any proportion of true signals. The observed FDR remains low across all scenarios, with only a few small-sample scenarios going beyond the 10% threshold. Keeping in line with this, the FPR also remains low, with typically low to zero falsely identified signals even in small-sample cases comparable to our real data analyses. The true positive rates are for low sample size and low effect size scenarios but undergoes drastic improvement with increase in either of those features. Overall, these results indicate several positive aspects for our real-data applications: (1) the smoothing applied on the reproduction number profiles by computing a 14-day average denoises the response variable in our regression models and ensures improved performance, (2) although we deal with the small-sample scenario of 30 samples in our applications, we are confident about the findings being true given the low false discoveries and false positives in the simulation scenarios – a higher sample size could probably allow better identification of more signals.

1.8 Uniqueness and utility of our pipeline

One of the goals of our paper is to compute COVID-19 Vulnerability Indices (cVIs) across multiple thematic resolutions for different geographical and spatial administrative regions based on publicly reported socio-economic, demographic, health-based and epidemiological data from national surveys in India. While our algorithm has similarities with the algorithm proposed by Acharya and Porwal (2020) for computing the relative vulnerability indices for each district, there are some differences.[2] One of the key differences in our approach is that we have used additional variables and indicators which are relevant for infectious disease outbreaks and COVID preparedness (such as COVID hospital testing centers) across the themes to come up with these indices.

More importantly, while computation of vulnerability indices (indicator specific relative cVI, theme-specific cVI and overall vulnerability indices) constitutes the first goal, we have taken a step further and used these multi-resolution cVIs in regression models to assess their impact on indicators of the spread of COVID-19 such as the average time-varying instantaneous reproduction number. This multi-layer aggregation approach examines the potential heterogeneities in the themed vulnerabilities across districts via exploring their association with pandemic growth metrics. Relative ranking for overall and theme specific vulnerabilities are also performed using regression models under a Bayesian paradigm using standard metrics of variable importance like the posterior inclusion probabilities. Furthermore, our paper demonstrates novelty via identification of specific target areas (for example, setting up more rural temporary healthcare facilities) for policy implementation in order to mitigate the crisis.

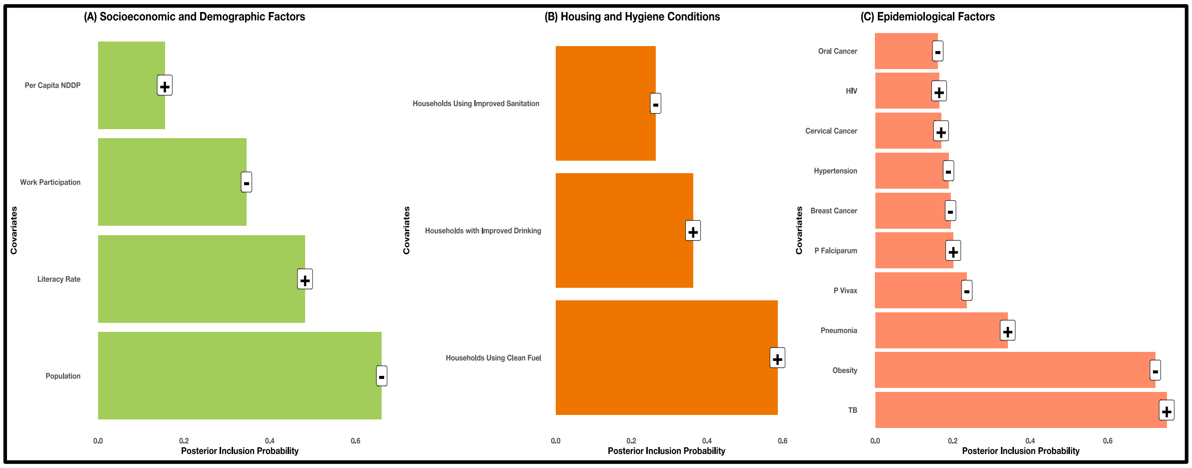
Bhattacharyya R, *et al. BMJ Open* 2022; 12:e056292. doi: 10.1136/bmjopen-2021-056292

2. SUPPLEMENTARY FIGURES

Algorithm 1: Algorithm for computing various VIs

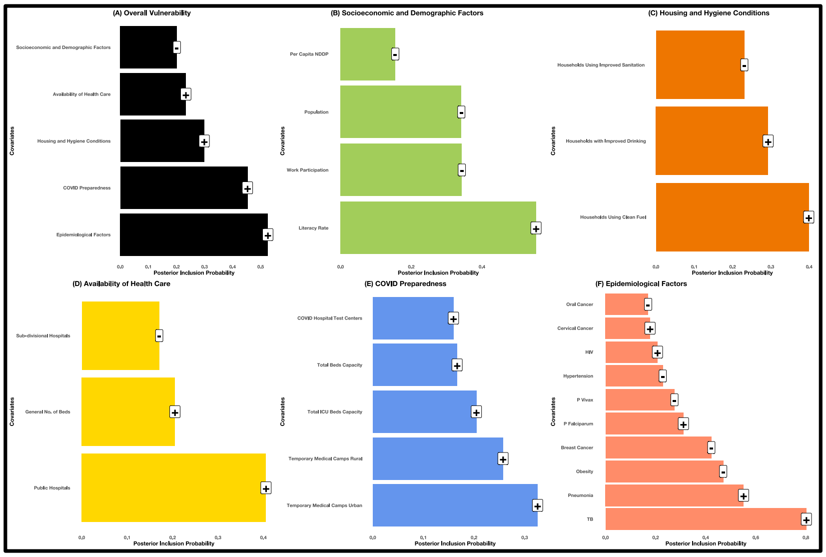
```
Input: d = no. of districts ;
Input: n = no. of covariates ;
Input: t = no. of themes (along with covariate composition of each
  theme) ;
while 1 ≤ i ≤ d do
    while 1 ≤ j ≤ n do
        if j^th covariate I_j is negative sensed then
            R_ij^(c) ← Σ_{1≤k≤d} I(x_ij ≥ x_kj) ;  /* Sum of indicators that
             value of covariate I_j for i^th district (x_ij) is more
             than that of k^th district (x_kj) */
        else
            if j^th covariate I_j is positive sensed then
                R_ij^(c) ← Σ_{1≤k≤d} I(x_ij ≤ x_kj) ;    /* Sum of indicators
                 that value of covariate I_j for i^th district (x_ij)
                 is less than that of k^th district (x_kj) */
            end
        end
        V_ij^(c) ← (R_ij^(c) − 1)/(d − 1) ;             /* Covariate Specific VI */
    end
    while 1 ≤ j ≤ t do
        define S_j = {k : I_k ∈ T_j} ;       /* Collection of covariates
         under theme T_j */
        define t_ij = Σ_{k∈S_j} V_ik^(c) ;  /* Sum of Covariate specific VIs
         for covariates under theme T_j for i^th district */
        R_ij^(t) ← Σ_{1≤k≤d} I(t_ij ≥ t_kj) ;  /* Same notation as above */
        V_ij^(t) ← (R_ij^(t) − 1)/(d − 1) ;             /* Theme Specific VI */
    end
    define o_i = Σ_{1≤k≤t} V_ik^(t) ; /* Sum of all theme specific VI for
     i^th district */
    R_i^(o) ← Σ_{1≤k≤d} I(o_i ≥ o_k) ;        /* Same notation as above */
    V_i^(o) =← (R_i^(o) − 1)/(d − 1) ;                    /* Overall VI */
end
```

Supplementary Figure 1. Algorithm for computing variable-specific, theme-level, and overall vulnerability indices for a given set of districts and covariates.



Supplementary Figure 2. Summary of regression models for instantaneous R with the vulnerability indices as covariates.

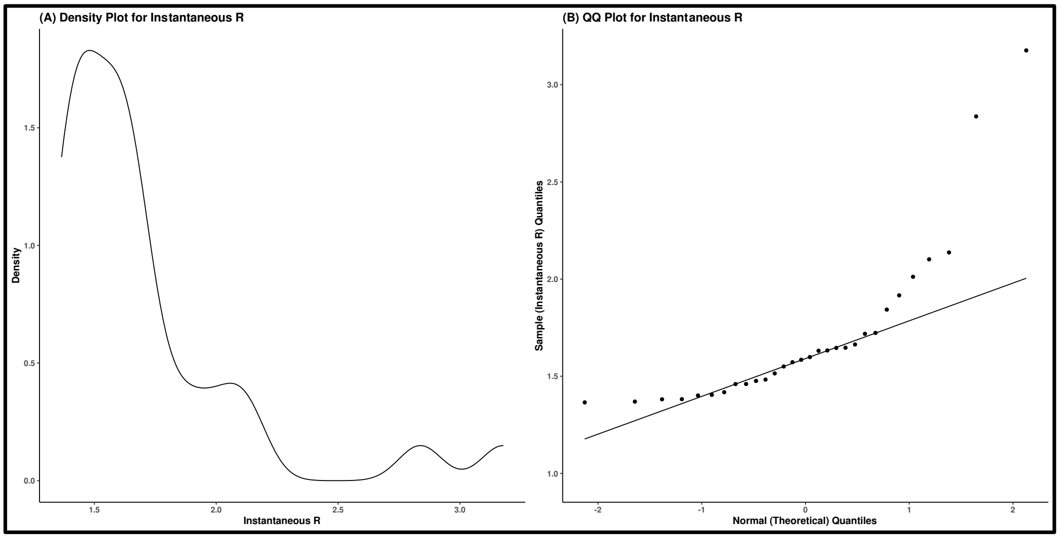
Panels A-C depict the indicators within the themes ‘Socioeconomic and Demographic Factors’, ‘Housing and Hygiene Conditions’ and ‘Epidemiological Factors’ respectively, as mentioned in the individual panel titles. In each case, a Bayesian model averaging-based linear regression model is fit using instantaneous R as response and the indicators/themes as covariates. The heights of the bars indicate the posterior inclusion probabilities for the covariates in the fitted models, and the labels on top of the bars indicate the signs of the estimated coefficient as obtained in those models.



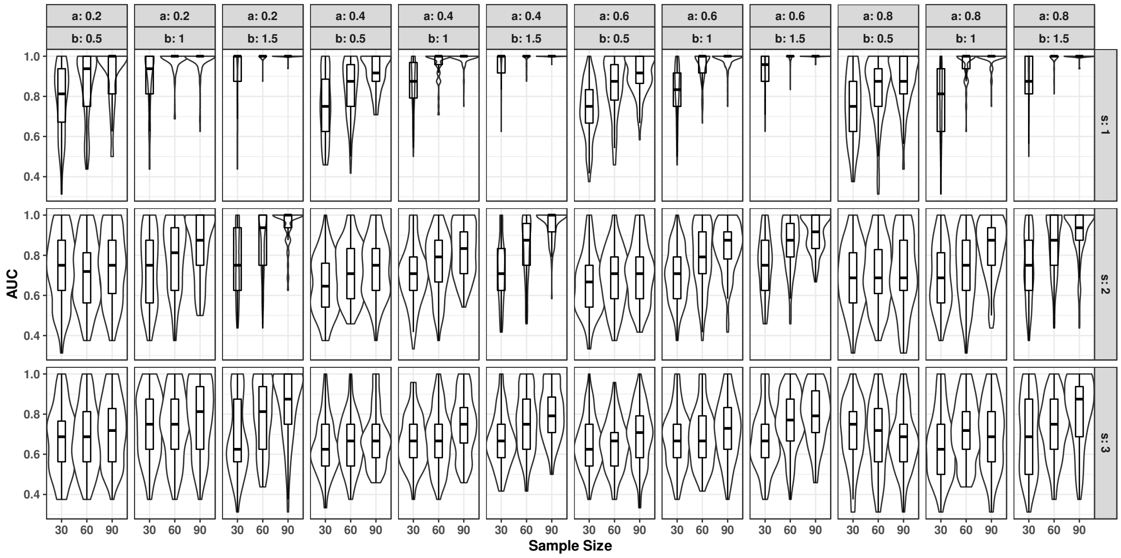
Supplementary Figure 3. Summary of regression models for variability of R (vR) with the vulnerability indices as covariates.

Panel A corresponds to the themed vulnerability indices constituting the overall vulnerability index and Panels B-F depict the indicators within the five themed vulnerability indices, as mentioned in the individual panel titles. In each case, a Bayesian model averaging-based linear regression model is fit using vR as response and the indicators/themes as covariates. The heights of the bars indicate the posterior inclusion probabilities for the covariates in the fitted models, and the labels on top of the bars indicate the signs of the estimated coefficient as obtained in those models.

Bhattacharyya R, *et al. BMJ Open* 2022; 12:e056292. doi: 10.1136/bmjopen-2021-056292



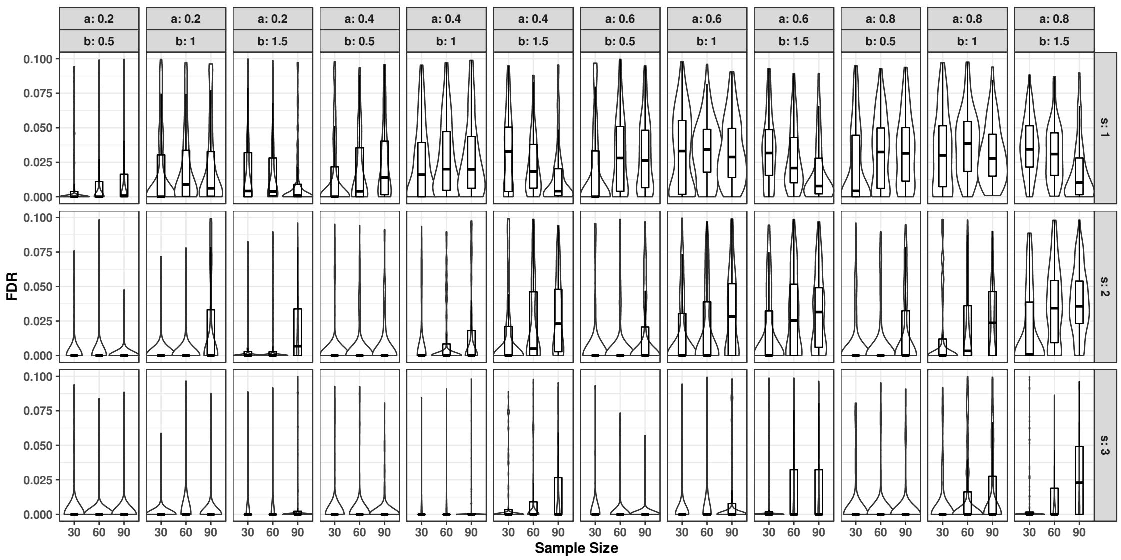
Supplementary Figure 4. Density plot and quantile-quantile plot for the last fortnight mean of estimated R (instantaneous R or iR).



Supplementary Figure 5. Area under the receiver operating characteristics curve (AUC) summaries across 100 iterations of the Bayesian model averaging procedure-based linear regression for each scenario.

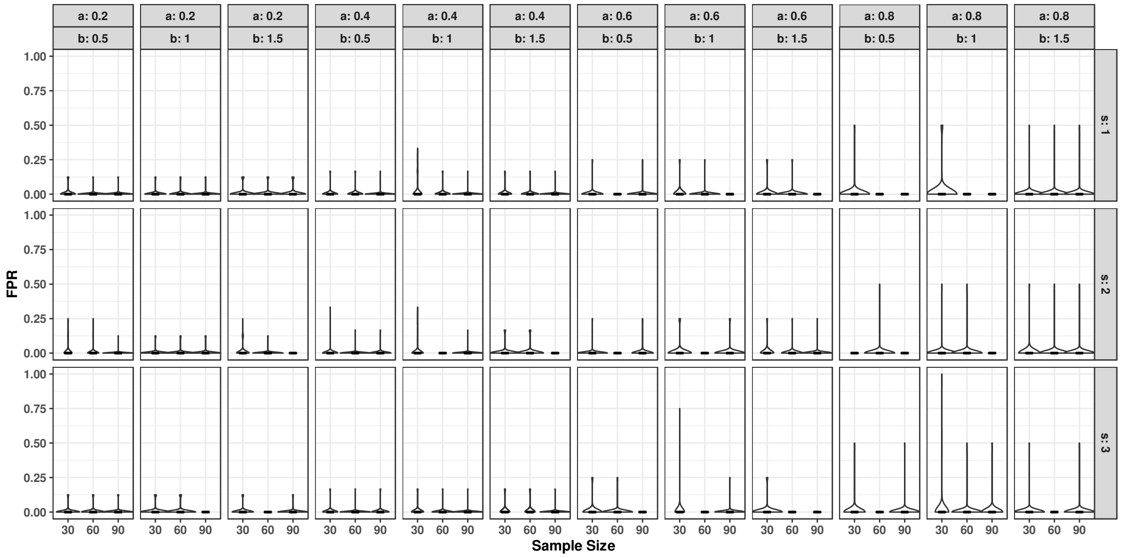
In the panels, a: proportion of true (non-zero) signals among 10 covariates, b: minimum absolute value for the non-zero coefficients, s: standard deviation of the noise distribution.

Bhattacharyya R, *et al. BMJ Open* 2022; 12:e056292. doi: 10.1136/bmjopen-2021-056292



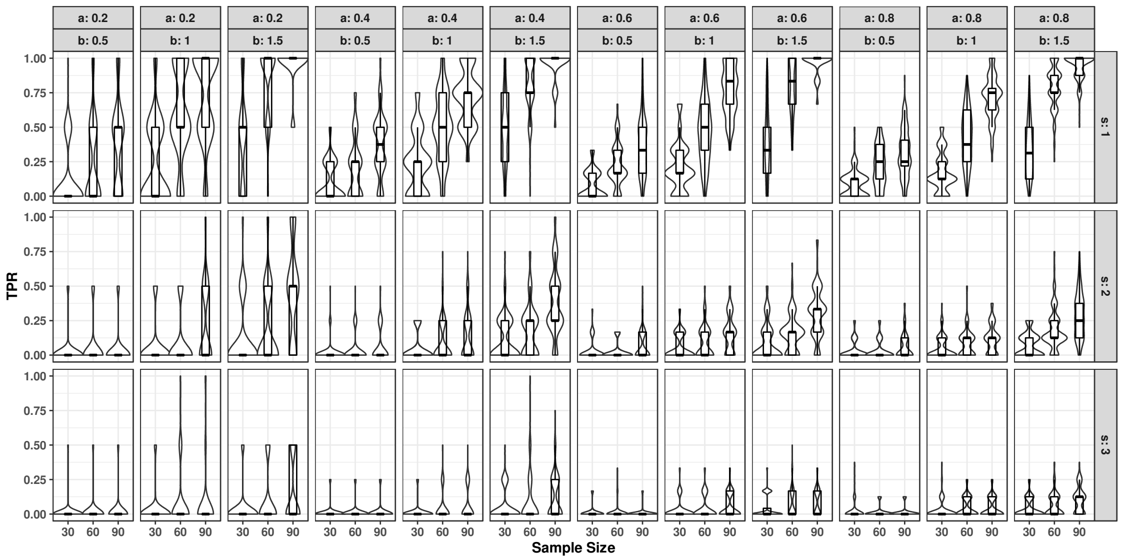
Supplementary Figure 6. False discovery rate (FDR) summaries across 100 iterations of the Bayesian model averaging procedure-based linear regression for each scenario.

In the panels, a: proportion of true (non-zero) signals among 10 covariates, b: minimum absolute value for the non-zero coefficients, s: standard deviation of the noise distribution.



Supplementary Figure 7. False positive rate (FPR) summaries across 100 iterations of the Bayesian model averaging procedure-based linear regression for each scenario.

In the panels, a: proportion of true (non-zero) signals among 10 covariates, b: minimum absolute value for the non-zero coefficients, s: standard deviation of the noise distribution.



Supplementary Figure 8. True positive rate (TPR) summaries across 100 iterations of the Bayesian model averaging procedure-based linear regression for each scenario.

In the panels, a: proportion of true (non-zero) signals among 10 covariates, b: minimum absolute value for the non-zero coefficients, s: standard deviation of the noise distribution.

3. SUPPLEMENTARY TABLES

Supplementary Table 1. Summary of indicators constituting the five themed vulnerability indices along with data sources.

Variable	Variable Description	Data Source
Socioeconomic and Demographic Factors (Theme 1)		
Population	Calculated as a number of population (district wise)	Census of India, 2011 https://censusindia.gov.in/2011-prov-results/data_files/orissa/Provisional Population Total Orissa-Book.pdf Population data: linearly projected population for 2019 using growth rate calculated for each district based on 2001 and 2011 census
Literacy Rate	Calculated as percentage of females who completed primary level of education	Census of India, 2011 https://censusindia.gov.in/2011-prov-results/data_files/orissa/Provisional Population Total Orissa-Book.pdf
Work Participation	Calculated as percentage of people who are working	
Per Capita NDDP		Economic Survey of Odisha, Directorate of Economic & Statistics, Odisha India 2011-12 http://desorissa.nic.in/pdf/lt_publication/Economic_Survey_2011_12.pdf
Housing and Hygiene Conditions (Theme 2)		
Households with improved Drinking Water Sources	Calculated as percentage of households reporting improved Drinking Water Sources	National Family Health Survey-4 2015-16 http://rchiips.org/nfhs/pdf/NFHS4/OR_FactSheet.pdf
Households using improved Sanitation facility	Calculated as percentage of households reporting improved Sanitation facility	
Households using clean fuel for cooking	Calculated as percentage of households reporting clean fuel for cooking	
Availability of Health Care (Theme 3)		
Availability of public hospitals (at district level)	Calculated as number of public hospitals (primary health center, sub-divisional and above) per 10000 population	Directorate of Health Services, Department of Health and Family Welfare, Government of ODISHA website https://health.odisha.gov.in/
General number of beds (at district level)	Number of beds available per 10000 population	
Preparedness of COVID (Theme 4)		
Total Beds Capacity (at district level)	Capacity of beds per 10000 population	COVID Dashboard Govt. Of Odisha https://statedashboard.odisha.gov.in/
Total ICU Beds (at district level)	ICU beds per 10000 population	
Temporary Medical Camps (at district level)	Temporary medical camps per 10000 population	
COVID Hospital Testing Centers (at district level)	Testing centers per 10000 population	
Epidemiological Factors (Theme 5)		
Total HIV Positive	Percentage of total HIV positive to total tested (Male + Female)	Health Management Information System (2019-20) https://hmis.nhp.gov.in/downloadfile?filepath=publications/Rural-Health-Statistics/RHS%202019-20.pdf
Plasmodium Vivax Test Positive	Percentage of plasmodium Vivax test positive to total blood smears examined	
Plasmodium Falciparum Test Positive	Percentage of plasmodium Falciparum test positive to total blood smears examined	
Infants Deaths due to Pneumonia	Percentage of deaths due to Pneumonia to total reported Infant deaths	
Hypertension	Percentage of Hypertension positive to total persons examined	National Family Health Survey-4 2015-16 http://rchiips.org/nfhs/pdf/NFHS4/OR_FactSheet.pdf
Cervix	Percentage of Cervix positive to total persons examined	
Breast	Percentage of Breast positive to total persons examined	
Oral cavity	Percentage of Oral cavity positive to total persons examined	
Obesity	Percentage of overweight/obese to total persons examined	

Bhattacharyya R, *et al. BMJ Open* 2022; 12:e056292. doi: 10.1136/bmjopen-2021-056292

Supplementary Table 2. Summary of sample size and parameter choices for the effective reproduction number estimation and subsequent regression procedures.

Parameter	Value Used	Justification
Sample Size and Related Choices		
n (Sample Size)	30	Data used across the 30 districts of Odisha.
M (Number of Themes)	5	Data available from across different surveys.
T (Number of Days)	350 (May 1, 2020 – Apr 15, 2021)	Initial few months prior to May saw very cases in most of the Odisha districts, resulting in inflated and unstable estimates of the effective reproduction number.
K (iR Window)	14 (Apr 2 – Apr 15, 2021)	K = 1, 7, 14, 30 were investigated. K = 14 yielded the most stable and close to normal distribution of the iR.
Choices for estimate_R Function in Epiestim R Package[6]		
Method	parametric_SI	Known and specifiable mean and sd of the serial interval distribution, as indicated in the subsequent rows.
Estimation Window	5 days	Chosen by investigating stability of the estimates and widths of the resulting confidence intervals across varying window lengths.
Serial Interval Prior Distribution	Gamma with mean 3.96 days and sd 4.75 days	Following recent studies on serial interval distribution of COVID-19 cases.[7]

Supplementary Table 3. Summary statistics from Bayesian model averaging-based multiple linear regression models (theme-specific and overall) for the outcome instantaneous R (iR).

Source Model	Covariate	Posterior Inclusion Probability	Coefficient Posterior Mean	Coefficient Posterior SD
Availability of Health Care	Public Hospitals	0.62	0.35	0.36
	General No. of Beds	0.23	0.03	0.21
	Sub-divisional Hospitals	0.16	0.00	0.13
COVID Preparedness	Temporary Medical Camps Rural	0.59	0.33	0.35
	Temporary Medical Camps Urban	0.37	0.15	0.25
	Total ICU Beds Capacity	0.30	0.12	0.27
	Total Beds Capacity	0.17	0.01	0.10
	COVID Hospital Test Centers	0.16	0.08	1.02
Epidemiological Factors	TB	0.75	0.39	0.29
	Obesity	0.72	-0.41	0.33
	Pneumonia	0.34	0.13	0.24
	P Vivax	0.24	-0.05	0.13
	P Falciparum	0.20	0.03	0.14
	Breast Cancer	0.19	-0.03	0.13
	Hypertension	0.19	-0.03	0.13
	Cervical Cancer	0.17	0.01	0.12
	HIV	0.16	0.01	0.09
	Oral Cancer	0.16	-0.01	0.10
Housing and Hygiene Conditions	Households Using Clean Fuel	0.59	0.37	0.42
	Households with Improved Drinking	0.36	0.13	0.23
	Households Using Improved Sanitation	0.26	-0.08	0.29
Socioeconomic and Demographic Factors	Population	0.66	-0.35	0.32
	Literacy Rate	0.48	0.23	0.30
	Work Participation	0.35	-0.14	0.28
	Per Capita NDDP	0.16	0.01	0.09
Overall Vulnerability	Housing and Hygiene Conditions	0.45	0.18	0.26
	COVID Preparedness	0.42	0.18	0.27
	Epidemiological Factors	0.42	0.17	0.26
	Availability of Health Care	0.35	0.13	0.23
	Socioeconomic and Demographic Factors	0.20	-0.04	0.15

Bhattacharyya R, *et al. BMJ Open* 2022; 12:e056292. doi: 10.1136/bmjopen-2021-056292

Supplementary Table 4. Summary statistics from Bayesian model averaging-based multiple linear regression models (theme-specific and overall) for the outcome variability in R (vR).

Source Model	Covariate	Posterior Inclusion Probability	Coefficient Posterior Mean	Coefficient Posterior SD
Availability of Health Care	Public Hospitals	0.41	1.36	2.16
	General No. of Beds	0.21	0.13	1.35
	Sub-divisional Hospitals	0.17	-0.17	1.08
COVID Preparedness	Temporary Medical Camps Urban	0.33	0.86	1.66
	Temporary Medical Camps Rural	0.26	0.58	1.46
	Total ICU Beds Capacity	0.20	0.40	1.39
	Total Beds Capacity	0.17	0.12	0.77
	COVID Hospital Test Centers	0.16	0.63	7.72
Epidemiological Factors	TB	0.80	3.29	2.20
	Pneumonia	0.55	2.03	2.30
	Obesity	0.47	-1.54	2.08
	Breast Cancer	0.42	-1.21	1.83
	P Falciparum	0.31	0.73	1.52
	P Vivax	0.28	-0.48	1.13
	Hypertension	0.23	-0.36	1.14
	HIV	0.21	0.25	0.89
	Cervical Cancer	0.18	0.08	0.95
	Oral Cancer	0.17	-0.07	0.80
Housing and Hygiene Conditions	Households Using Clean Fuel	0.40	1.56	2.68
	Households with Improved Drinking	0.29	0.69	1.49
	Households Using Improved Sanitation	0.23	-0.50	1.99
Socioeconomic and Demographic Factors	Epidemiological Factors	0.53	1.83	2.20
	COVID Preparedness	0.45	1.47	2.05
	Housing and Hygiene Conditions	0.30	0.69	1.44
	Availability of Health Care	0.23	0.44	1.22
Overall Vulnerability	Socioeconomic and Demographic Factors	0.20	-0.29	1.10
	Literacy Rate	0.55	2.19	2.48
	Work Participation	0.34	-1.02	2.06
	Population	0.34	-0.92	1.70
	Per Capita NDDP	0.15	-0.04	0.70

Bhattacharyya R, *et al. BMJ Open* 2022; 12:e056292. doi: 10.1136/bmjopen-2021-056292

Supplementary Table 5. Summary statistics from Bayesian model averaging-based simulation studies. n denotes the sample size, a denotes the proportion of true signals in the data generating mechanism, b denotes the minimum absolute value of non-zero coefficients, and s denotes the noise standard deviation. AUC: area under the receiver operating characteristics curve, FDR: false discovery rate, TPR: true positive rate, FPR: false positive rate. For FDR, TPR, and FPR, selections are determined using a 10% true FDR control. For each row, across 100 simulations, the first number denotes the median of the corresponding metric. The numbers in the parentheses denote respectively the corresponding first and third quartiles.

n	a	b	s	AUC	FDR	TPR	FPR
30	0.2	0.5	1	0.812 (0.672, 0.938)	0.000 (0.000, 0.004)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.2	0.5	2	0.750 (0.625, 0.875)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.2	0.5	3	0.688 (0.562, 0.766)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.2	1	1	0.938 (0.812, 1.000)	0.000 (0.000, 0.030)	0.000 (0.000, 0.500)	0.000 (0.000, 0.000)
30	0.2	1	2	0.750 (0.562, 0.875)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.2	1	3	0.750 (0.625, 0.875)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.2	1.5	1	1.000 (0.875, 1.000)	0.004 (0.000, 0.032)	0.500 (0.000, 0.500)	0.000 (0.000, 0.000)
30	0.2	1.5	2	0.750 (0.625, 0.938)	0.000 (0.000, 0.003)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.2	1.5	3	0.625 (0.562, 0.875)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.4	0.5	1	0.750 (0.625, 0.885)	0.000 (0.000, 0.022)	0.000 (0.000, 0.250)	0.000 (0.000, 0.000)
30	0.4	0.5	2	0.646 (0.542, 0.760)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.4	0.5	3	0.625 (0.542, 0.750)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.4	1	1	0.875 (0.792, 0.969)	0.016 (0.000, 0.039)	0.250 (0.000, 0.250)	0.000 (0.000, 0.000)
30	0.4	1	2	0.708 (0.625, 0.792)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.4	1	3	0.667 (0.583, 0.750)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.4	1.5	1	1.000 (0.917, 1.000)	0.033 (0.004, 0.051)	0.500 (0.250, 0.750)	0.000 (0.000, 0.000)
30	0.4	1.5	2	0.708 (0.625, 0.833)	0.000 (0.000, 0.021)	0.000 (0.000, 0.250)	0.000 (0.000, 0.000)
30	0.4	1.5	3	0.667 (0.583, 0.750)	0.000 (0.000, 0.003)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.6	0.5	1	0.750 (0.667, 0.833)	0.000 (0.000, 0.033)	0.000 (0.000, 0.167)	0.000 (0.000, 0.000)
30	0.6	0.5	2	0.667 (0.542, 0.750)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.6	0.5	3	0.625 (0.542, 0.750)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.6	1	1	0.833 (0.750, 0.917)	0.033 (0.002, 0.055)	0.167 (0.167, 0.333)	0.000 (0.000, 0.000)
30	0.6	1	2	0.708 (0.583, 0.792)	0.000 (0.000, 0.030)	0.000 (0.000, 0.167)	0.000 (0.000, 0.000)
30	0.6	1	3	0.667 (0.583, 0.750)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.6	1.5	1	0.958 (0.875, 1.000)	0.032 (0.015, 0.049)	0.333 (0.167, 0.500)	0.000 (0.000, 0.000)
30	0.6	1.5	2	0.750 (0.625, 0.875)	0.000 (0.000, 0.032)	0.000 (0.000, 0.167)	0.000 (0.000, 0.000)
30	0.6	1.5	3	0.667 (0.583, 0.750)	0.000 (0.000, 0.002)	0.000 (0.000, 0.042)	0.000 (0.000, 0.000)
30	0.8	0.5	1	0.750 (0.625, 0.875)	0.004 (0.000, 0.045)	0.125 (0.000, 0.125)	0.000 (0.000, 0.000)
30	0.8	0.5	2	0.688 (0.562, 0.812)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.8	0.5	3	0.750 (0.625, 0.812)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.8	1	1	0.812 (0.625, 0.938)	0.030 (0.007, 0.051)	0.125 (0.125, 0.250)	0.000 (0.000, 0.000)
30	0.8	1	2	0.688 (0.562, 0.812)	0.000 (0.000, 0.012)	0.000 (0.000, 0.125)	0.000 (0.000, 0.000)
30	0.8	1	3	0.625 (0.500, 0.750)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
30	0.8	1.5	1	0.875 (0.812, 1.000)	0.034 (0.022, 0.051)	0.312 (0.125, 0.500)	0.000 (0.000, 0.000)
30	0.8	1.5	2	0.750 (0.625, 0.875)	0.001 (0.000, 0.039)	0.000 (0.000, 0.125)	0.000 (0.000, 0.000)

30	0.8	1.5	3	0.688 (0.500, 0.875)	0.000 (0.000, 0.002)	0.000 (0.000, 0.125)	0.000 (0.000, 0.000)
60	0.2	0.5	1	0.938 (0.750, 1.000)	0.000 (0.000, 0.011)	0.000 (0.000, 0.500)	0.000 (0.000, 0.000)
60	0.2	0.5	2	0.719 (0.562, 0.812)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
60	0.2	0.5	3	0.688 (0.562, 0.812)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
60	0.2	1	1	1.000 (1.000, 1.000)	0.009 (0.000, 0.034)	0.500 (0.500, 1.000)	0.000 (0.000, 0.000)
60	0.2	1	2	0.812 (0.625, 0.938)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
60	0.2	1	3	0.750 (0.625, 0.875)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
60	0.2	1.5	1	1.000 (1.000, 1.000)	0.004 (0.001, 0.028)	1.000 (0.500, 1.000)	0.000 (0.000, 0.000)
60	0.2	1.5	2	0.938 (0.750, 1.000)	0.000 (0.000, 0.003)	0.000 (0.000, 0.500)	0.000 (0.000, 0.000)
60	0.2	1.5	3	0.812 (0.625, 0.938)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
60	0.4	0.5	1	0.875 (0.750, 0.958)	0.004 (0.000, 0.035)	0.250 (0.000, 0.250)	0.000 (0.000, 0.000)
60	0.4	0.5	2	0.708 (0.583, 0.833)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
60	0.4	0.5	3	0.667 (0.542, 0.750)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
60	0.4	1	1	1.000 (0.958, 1.000)	0.020 (0.005, 0.047)	0.500 (0.250, 0.750)	0.000 (0.000, 0.000)
60	0.4	1	2	0.792 (0.667, 0.875)	0.000 (0.000, 0.008)	0.000 (0.000, 0.250)	0.000 (0.000, 0.000)
60	0.4	1	3	0.667 (0.583, 0.750)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
60	0.4	1.5	1	1.000 (1.000, 1.000)	0.018 (0.006, 0.038)	0.750 (0.750, 1.000)	0.000 (0.000, 0.000)
60	0.4	1.5	2	0.875 (0.750, 0.958)	0.005 (0.000, 0.046)	0.250 (0.000, 0.250)	0.000 (0.000, 0.000)
60	0.4	1.5	3	0.750 (0.625, 0.875)	0.000 (0.000, 0.009)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
60	0.6	0.5	1	0.875 (0.781, 0.958)	0.028 (0.004, 0.051)	0.167 (0.167, 0.333)	0.000 (0.000, 0.000)
60	0.6	0.5	2	0.708 (0.583, 0.792)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
60	0.6	0.5	3	0.667 (0.542, 0.708)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
60	0.6	1	1	1.000 (0.917, 1.000)	0.034 (0.018, 0.049)	0.500 (0.333, 0.667)	0.000 (0.000, 0.000)
60	0.6	1	2	0.792 (0.708, 0.917)	0.000 (0.000, 0.039)	0.000 (0.000, 0.167)	0.000 (0.000, 0.000)
60	0.6	1	3	0.667 (0.583, 0.792)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
60	0.6	1.5	1	1.000 (1.000, 1.000)	0.021 (0.010, 0.043)	0.833 (0.667, 1.000)	0.000 (0.000, 0.000)
60	0.6	1.5	2	0.875 (0.792, 0.958)	0.025 (0.000, 0.052)	0.167 (0.000, 0.167)	0.000 (0.000, 0.000)
60	0.6	1.5	3	0.771 (0.667, 0.875)	0.000 (0.000, 0.032)	0.000 (0.000, 0.167)	0.000 (0.000, 0.000)
60	0.8	0.5	1	0.875 (0.750, 0.938)	0.032 (0.006, 0.050)	0.250 (0.125, 0.375)	0.000 (0.000, 0.000)
60	0.8	0.5	2	0.688 (0.609, 0.828)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
60	0.8	0.5	3	0.719 (0.562, 0.828)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
60	0.8	1	1	1.000 (0.938, 1.000)	0.039 (0.018, 0.055)	0.375 (0.250, 0.625)	0.000 (0.000, 0.000)
60	0.8	1	2	0.750 (0.625, 0.875)	0.003 (0.000, 0.036)	0.125 (0.000, 0.125)	0.000 (0.000, 0.000)
60	0.8	1	3	0.719 (0.625, 0.812)	0.000 (0.000, 0.016)	0.000 (0.000, 0.125)	0.000 (0.000, 0.000)
60	0.8	1.5	1	1.000 (1.000, 1.000)	0.031 (0.016, 0.046)	0.750 (0.750, 0.875)	0.000 (0.000, 0.000)
60	0.8	1.5	2	0.875 (0.750, 1.000)	0.034 (0.009, 0.054)	0.125 (0.125, 0.250)	0.000 (0.000, 0.000)
60	0.8	1.5	3	0.750 (0.625, 0.875)	0.000 (0.000, 0.019)	0.000 (0.000, 0.125)	0.000 (0.000, 0.000)
90	0.2	0.5	1	1.000 (0.812, 1.000)	0.001 (0.000, 0.016)	0.500 (0.000, 0.500)	0.000 (0.000, 0.000)
90	0.2	0.5	2	0.750 (0.609, 0.875)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
90	0.2	0.5	3	0.719 (0.562, 0.828)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
90	0.2	1	1	1.000 (1.000, 1.000)	0.006 (0.000, 0.033)	1.000 (0.500, 1.000)	0.000 (0.000, 0.000)
90	0.2	1	2	0.875 (0.750, 1.000)	0.000 (0.000, 0.033)	0.000 (0.000, 0.500)	0.000 (0.000, 0.000)
90	0.2	1	3	0.812 (0.625, 0.938)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
90	0.2	1.5	1	1.000 (1.000, 1.000)	0.001 (0.000, 0.009)	1.000 (1.000, 1.000)	0.000 (0.000, 0.000)
90	0.2	1.5	2	1.000 (0.938, 1.000)	0.007 (0.000, 0.034)	0.500 (0.000, 0.500)	0.000 (0.000, 0.000)
90	0.2	1.5	3	0.875 (0.750, 1.000)	0.000 (0.000, 0.002)	0.000 (0.000, 0.500)	0.000 (0.000, 0.000)
90	0.4	0.5	1	0.917 (0.875, 1.000)	0.014 (0.002, 0.040)	0.375 (0.250, 0.500)	0.000 (0.000, 0.000)
90	0.4	0.5	2	0.750 (0.625, 0.833)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
90	0.4	0.5	3	0.667 (0.583, 0.750)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
90	0.4	1	1	1.000 (1.000, 1.000)	0.020 (0.006, 0.044)	0.750 (0.500, 0.750)	0.000 (0.000, 0.000)
90	0.4	1	2	0.833 (0.708, 0.917)	0.000 (0.000, 0.018)	0.000 (0.000, 0.250)	0.000 (0.000, 0.000)
90	0.4	1	3	0.750 (0.656, 0.833)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
90	0.4	1.5	1	1.000 (1.000, 1.000)	0.004 (0.001, 0.020)	1.000 (1.000, 1.000)	0.000 (0.000, 0.000)
90	0.4	1.5	2	1.000 (0.917, 1.000)	0.023 (0.003, 0.048)	0.250 (0.250, 0.500)	0.000 (0.000, 0.000)
90	0.4	1.5	3	0.792 (0.708, 0.885)	0.000 (0.000, 0.027)	0.000 (0.000, 0.250)	0.000 (0.000, 0.000)
90	0.6	0.5	1	0.917 (0.865, 1.000)	0.026 (0.007, 0.048)	0.333 (0.167, 0.500)	0.000 (0.000, 0.000)
90	0.6	0.5	2	0.708 (0.583, 0.792)	0.000 (0.000, 0.021)	0.000 (0.000, 0.167)	0.000 (0.000, 0.000)
90	0.6	0.5	3	0.708 (0.583, 0.792)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
90	0.6	1	1	1.000 (1.000, 1.000)	0.029 (0.014, 0.050)	0.833 (0.667, 1.000)	0.000 (0.000, 0.000)

90	0.6	1	2	0.875 (0.781, 0.917)	0.028 (0.000, 0.052)	0.167 (0.000, 0.167)	0.000 (0.000, 0.000)
90	0.6	1	3	0.729 (0.625, 0.833)	0.000 (0.000, 0.008)	0.000 (0.000, 0.167)	0.000 (0.000, 0.000)
90	0.6	1.5	1	1.000 (1.000, 1.000)	0.008 (0.002, 0.028)	1.000 (1.000, 1.000)	0.000 (0.000, 0.000)
90	0.6	1.5	2	0.917 (0.833, 1.000)	0.032 (0.006, 0.049)	0.333 (0.167, 0.333)	0.000 (0.000, 0.000)
90	0.6	1.5	3	0.792 (0.708, 0.917)	0.000 (0.000, 0.032)	0.000 (0.000, 0.167)	0.000 (0.000, 0.000)
90	0.8	0.5	1	0.875 (0.812, 1.000)	0.032 (0.011, 0.050)	0.250 (0.219, 0.406)	0.000 (0.000, 0.000)
90	0.8	0.5	2	0.688 (0.625, 0.875)	0.000 (0.000, 0.032)	0.000 (0.000, 0.125)	0.000 (0.000, 0.000)
90	0.8	0.5	3	0.688 (0.562, 0.750)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)	0.000 (0.000, 0.000)
90	0.8	1	1	1.000 (1.000, 1.000)	0.028 (0.015, 0.045)	0.750 (0.625, 0.781)	0.000 (0.000, 0.000)
90	0.8	1	2	0.875 (0.750, 0.938)	0.024 (0.000, 0.046)	0.125 (0.000, 0.125)	0.000 (0.000, 0.000)
90	0.8	1	3	0.688 (0.562, 0.812)	0.000 (0.000, 0.028)	0.000 (0.000, 0.125)	0.000 (0.000, 0.000)
90	0.8	1.5	1	1.000 (1.000, 1.000)	0.010 (0.001, 0.028)	1.000 (0.875, 1.000)	0.000 (0.000, 0.000)
90	0.8	1.5	2	0.938 (0.875, 1.000)	0.036 (0.023, 0.054)	0.250 (0.125, 0.375)	0.000 (0.000, 0.000)
90	0.8	1.5	3	0.875 (0.688, 0.938)	0.023 (0.000, 0.049)	0.125 (0.000, 0.125)	0.000 (0.000, 0.000)

Bhattacharyya R, *et al. BMJ Open* 2022; 12:e056292. doi: 10.1136/bmjopen-2021-056292

REFERENCES

1. International Institute for Population Sciences (IIPS) and ICF. National Family Health Survey (NFHS-4), 2015-16: India Mumbai: IIPS; 2017 [Available from: https://dhsprogram.com/pubs/pdf/FR339/FR339.pdf accessed May 10 2022.

2. Acharya R, Porwal A. A vulnerability index for the management of and response to the COVID-19 epidemic in India: an ecological study. *The Lancet Global health* 2020;8(9):e1142-e51. doi: 10.1016/S2214-109X(20)30300-4 [published Online First: 2020/07/16]

3. Pearson K. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 1901;2(11):559-72. doi: 10.1080/14786440109462720

4. Zeugner S, Feldkircher M. Bayesian Model Averaging Employing Fixed and Flexible Priors: TheBMSPackage forR. *Journal of Statistical Software* 2015;68(4) doi: 10.18637/jss.v068.i04

5. Baladandayuthapani V, Ji Y, Talluri R, et al. Bayesian random segmentation models to identify shared copy number aberrations for array CGH data. *Journal of the american statistical association* 2010;105(492):1358-75.

6. Cori A, Ferguson NM, Fraser C, et al. A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology* 2013;178(9):1505-12. doi: 10.1093/aje/kwt133 [published Online First: 2013/09/15]

7. Du Z, Xu X, Wu Y, et al. Serial interval of COVID-19 among publicly reported confirmed cases. *Emerging infectious diseases* 2020;26(6):1341.