Si *et al. Population Health Metrics* (2026) 24:15
https://doi.org/10.1186/s12963-026-00456-7

Population Health Metrics

RESEARCH Open Access

Multilevel regression and poststratification interface: an application to track community-level COVID-19 viral transmission

Yajuan Si[1*], Toan Tran[1], Jonah Gabry[2], Mitzi Morris[2] and Andrew Gelman[3]

Abstract

Purpose Public health surveillance systems require high-quality data to represent the population. In the absence of comprehensive or random testing throughout the COVID-19 pandemic, we have developed a proxy method for synthetic random sampling to estimate the actual community-level viral incidence, based on viral testing of patients who are asymptomatic and present for elective procedures within a hospital system.

Methods The approach collects routine testing data on SARS-CoV-2 exposure among outpatients and performs statistical adjustments of sample representation using multilevel regression and poststratification (MRP), a procedure that adjusts for nonrepresentativeness of the sample and yields stable small group estimates. We extend MRP to accommodate time-varying data and granular geography.

Results We have developed an open-source, user-friendly MRP interface for public implementation of the Bayesian analysis workflow. We illustrate the MRP interface with an application to track community-level COVID-19 viral transmission in Michigan. We present the estimated infection rate over time for the targeted population and across demographic and geographic subpopulations.

Conclusion The interface provides timely, substantive insights into population health trends and serves as a valuable surveillance tool for future epidemic preparedness. Beyond monitoring COVID-19, the MRP interface can analyze a wide range of health and social science data, making it broadly applicable to diverse research areas with reproducibility and scientific rigor.

Keywords Bayesian workflow, Bias correction, Subgroup estimation, Surveillance system, Reproducibility

*Correspondence:
Yajuan Si
yajuan@umich.edu
[1]Institute for Social Research, University of Michigan, 426 Thompson St., Ann Arbor 48104, MI, USA
[2]Institute for Social and Economic Research and Policy, Columbia University, 1255 Amsterdam Ave., New York 10027, NY, USA
[3]Departments of Statistics and Political Science, Columbia University, 1255 Amsterdam Ave., New York 10027, NY, USA

Introduction

Early and accurate knowledge of incidence and trends of transmission within communities is crucial for monitoring a pandemic and supporting policymakers in assessing the effects of restrictive measures on individual and community behaviors. However, without universal screening or random testing, government policy and healthcare implementation responses have relied on testing people who were symptomatic or presumed exposed, with policies guided by officially reported positivity rates and counts of positive cases in the community. These data are biased in concept and flawed in practice. It is essential to establish an operational surveillance system that allows prompt assessments of mitigation efforts and future predictions of clinical burdens. This would trigger an effective healthcare response and inform other epidemics. Since universal random testing is not always feasible, we must develop alternative methods that offer similar advantages. An effective proxy measure should be able to detect increases in viral incidence before these trends become clinically relevant. It should also identify decreases in incidence, allowing for the safe suspension of mitigation strategies and supporting economic and social recovery. Ideally, the data collection and analysis procedure would be practical at the community level and applicable nationwide, with the capacity to focus on burdens within specific demographics. It must also be reliable, statistically valid, cost-effective, and automatic, ensuring that it can be deployed promptly in future pandemics.

We present a novel Bayesian workflow for multilevel regression and poststratification [MRP, 1], which is built on a user-friendly open-source interface, and demonstrate the application to track community-level COVID-19 viral transmission. Previous work has developed the foundation of such a proxy metric for COVID-19 tracking data collection and statistical adjustment of demographic representation [2, 3]. The approach collects electronic health records (EHRs) on routine viral testing of patients who present for elective procedures within a hospital system and performs MRP to estimate actual viral trends. The findings in a diverse urban-suburban-rural setting in Indiana show that this model predicts the clinical burden of SARS-CoV-2 earlier and more accurately than the currently accepted metrics. In contrast, the official testing data fail to inform the surge of clinical burdens.

Our new development enhances traditional MRP by introducing two novel methodological extensions: estimation varying across time and at granular geographic levels. Current applications of MRP predominantly rely on cross-sectional data collected at a single time point, either through probability sample surveys [e.g., 4, 5] or nonprobability samples [e.g., 6, 7]. In contrast, this paper extends MRP to accommodate time-varying data, enabling the tracking of trends. Moreover, due to sample size constraints, existing MRP estimates are typically at the national or state level. Our interface overcomes these limitations by supporting analyses at multiple geographic resolutions, including ZIP codes, counties, and states. The interface is capable of modeling both binary and continuous outcomes. As a demonstration, we apply the interface to time-varying, ZIP-code-level data to monitor demographic and county-level COVID-19 viral transmission trends in Michigan. We generate granular, timely and reproducible findings to direct resource allocation, guide surge prediction and inform rapid public health responses and evidence-based policy making.

As a prediction approach to modeling the outcome measures with individual-level and group-level predictors, MRP has become increasingly popular for subgroup estimation. Originally applied to estimate state-level public opinions from sociodemographic subgroups using sample surveys, MRP has two key components: (1) multilevel regression for small group estimation by setting up a predictive model with a large number of covariates and regularizing with Bayesian prior specifications, and (2) poststratification to adjust for selection bias and correct for imbalances in the sample composition from the target population. Flexible modeling of study outcomes can capture complex data structures conditional on poststratification cells, which are determined by the cross-tabulation of categorical auxiliary variables and calibrate the sample discrepancy with population control information. Besides applications in social sciences, especially in election forecasting [e.g., 6, 8], MRP has also shown promise in public health research [e.g., 4, 5, 7, 9, 10] show that the key to the success of MRP in applications is the inclusion of highly predictive covariates, and [11] discuss estimation approaches when the population distribution of the poststratification variables is unknown.

The new workflow improves the MRP method with an application of tracking community-level transmission—across geographic areas and demographic subgroups—to monitor the epidemic over time using the developed interface as an operational surveillance system. Leveraging the statistical programming language Stan [12] to conduct Bayesian computation and model estimation, specifically with the R package `cmdstanr` [13], the user-friendly interface promotes methodological transparency, reproducibility, and standardization for public health and social science research. The interface is accessible via both web-based and local platforms using the R package `shinymrp` [14], offering flexible tools to facilitate analyses for diverse users. By integrating data preparation,

model fitting, diagnostic checks, and graphical reporting into a comprehensive end-to-end pipeline, the interface enables researchers and practitioners to standardize MRP analyses across studies and datasets. This feature substantially lowers technical barriers, allowing reproducible application of advanced statistical methods and fostering comparability in population health research. We illustrate model use and interface capabilities with a real-world application tracking community-level COVID-19 viral transmission in Michigan. Health agencies can use subgroup-specific incidence estimates to allocate testing resources and direct targeted interventions to communities with greater vulnerability. Tracking epidemiological trends in real time through routine hospital data provides actionable guidance for surge prediction and rapid public health responses. While our primary application centers on community-level COVID-19 monitoring, the MRP workflow developed here is broadly applicable across a range of public health surveillance contexts.

The main contributions of this paper include: 1) introducing the MRP computational interface; 2) extending MRP models with time-varying and granular geographic data; 3) applying the metric to track COVID-19 viral transmission in Michigan. We describe the data source and MRP methods in Sect. "Methods". Section "Bayesian workflow with MRP" presents the workflow of statistical analyses, from data preprocessing, descriptive summaries, model fitting and diagnostics, to result presentation and validation. Section "Discussion" concludes with discussions and potential directions for future work.

Methods

Data

Our COVID-19 tracking approach collects electronic health records of prospective surgical (and other invasive procedure) patients who are asymptomatic and have tested for acute SARS-CoV-2 infection before performing the procedure [2, 3]. Upon the reopening of elective medical and surgical procedures after the initial COVID-19 outbreak in early 2020, all preoperative patients were uniformly required—per the American Society of Anesthesiology guidelines—to undergo surgical risk evaluation and testing for acute SARS-CoV-2 infection before any such procedures. This policy was implemented nationwide across the U.S. All elective patients were presumed asymptomatic because any individual reporting symptoms or recent exposure to the virus would have their procedure either canceled or deferred. Using a standardized protocol, all preoperative patients underwent polymerase chain reaction (PCR) testing for viral RNA four days prior to their scheduled procedure, with tests administered by health system staff and samples analyzed using the same system. This PCR testing protocol was maintained consistently throughout the study period. Additionally, a subset of patients, for whom preoperative blood testing was clinically indicated based on age, health status, or surgery type, were also screened for the presence of immunoglobulin G (IgG) to the SARS-CoV-2 nucleocapsid protein (IgG N), beginning May 1, 2020.

In collaboration with hospital database managers and in compliance with HIPAA privacy regulations, we collected EHR data including PCR test results, test dates, sex, age, race, and five-digit ZIP codes. The group represents broad age, racial, and socioeconomic diversity, with its only explicit correlation to disease status being the selection for elective surgical procedures and absence of symptoms or known exposure. We assume that, within any demographic and geographic stratum, the ratio of asymptomatic to symptomatic SARS-CoV-2 infections remains constant. Accordingly, the incidence of asymptomatic infections should proportionally reflect community-wide viral incidence and can serve as a proxy for true incidence trends, though this ratio may change with the emergence of new viral variants and the level of acquired immunity over time. To the extent that healthcare use or other factors affect the selection and ratio, we expect much of this variation to be addressed through our model adjustments. We discuss the potential violation of these assumptions in Sect. "Discussion". MRP adjusts the demographic (sex, age, and race) and geographic (five-digit ZIP code) distributions to the target population. The target population is defined as U.S. residents dwelling in the catchment area of the collected ZIP codes. We link the input patient EHR data with ZIP codes to census tract measures in the American Community Survey (ACS), the largest household survey of the U.S. population [15], and use the ACS aggregated summaries of sociodemographic and socioeconomic characteristics as geographic predictors at the ZIP level.

Previous work has treated PCR test sensitivity and specificity as unknown parameters, incorporating information from prior studies and accounting for estimation uncertainty in final MRP estimates [2, 3, 16], which we recommend for use when their values are undetermined. The interface allows for specifying different sensitivity and specificity values; for our demonstration, we presume 70% clinical sensitivity and 100% specificity, consistent with the previous setting during the same study period (March 2020–October 2022) [2, 3, 16, 17].

We track viral infections on a weekly basis. Below, we first introduce the conventional MRP framework for cross-sectional data, followed by extensions to accommodate time-varying data features.

MRP for cross-sectional data

MRP first fits a multilevel regression model to predict the outcome measure as a function of a set of factors, then poststratifies the categorical factors so that their distributions match those of the target population. We use a binary outcome of interest as an example. Let $y_i(=0/1)$ be the binary response for individual *i*, with $y_i = 1$ indicating the positive response. We employ a logistic regression with varying effects for age, race, and ZIP code, where the ZIP-code-level variation is further explained by the ZIP-code-level predictors.

$$\Pr(y_i = 1) = \text{logit}^{-1}(\beta_1 + \beta_2 \text{male}_i + \alpha^{\text{age}}_{\text{a[i]}} + \alpha^{\text{race}}_{\text{r[i]}} + \alpha^{\text{ZIP}}_{\text{s[i]}}), \quad (1)$$

where male_i is an indicator for men, $\alpha^{\text{age}}_{\text{a}}$ is the age effect, with a value of $a[i]$ for subject *i*, on the log-odds function of the probability of having a positive response, $\alpha^{\text{race}}_{\text{r}}$ is the racial effect, and $\alpha^{\text{ZIP}}_{\text{s}}$ is the ZIP-code-level effect. In the Bayesian framework, we assign hierarchical priors to varying intercepts as a default setting:

$$\begin{aligned} \alpha^{\text{age}} &\sim \text{ normal } (0, \sigma^{\text{age}}), \ \ \sigma^{\text{age}} \sim \text{ normal}_+(0, 2.5) \\ \alpha^{\text{race}} &\sim \text{ normal } (0, \sigma^{\text{race}}), \ \ \sigma^{\text{race}} \sim \text{ normal}_+(0, 2.5). \end{aligned} \quad (2)$$

Here $\text{normal}_+(0, 2.5)$ represents a half-normal distribution with the mean 0 and standard deviation 2.5 restricted to positive values. As we have ZIP-code-level predictors $\vec{Z}^{\text{ZIP}}_s$, we need to build another model in which $\alpha^{\text{ZIP}}_{\text{s}}$ is the outcome of a linear regression with ZIP-code-level predictors:

$$\begin{aligned} \alpha^{\text{ZIP}}_{\text{s}} = \vec{\alpha}\vec{Z}^{\text{ZIP}}_s + e_s, \ \ e_s \sim \text{ normal } (0, \sigma^{\text{ZIP}}), \\ \sigma^{\text{ZIP}} \sim \text{ normal}_+(0, 2.5), \end{aligned} \quad (3)$$

where e_s is a ZIP-code-level random error.

Alternative prior distributions, including structured priors for high-order interaction terms [5] and spatial prior accounting for geospatial correlation, can also be specified. We use the default setting in the interface with normal priors given in (2) and (3) as examples and discuss extensions to other settings in Sect. "Discussion".

Because (1) assumes that the people in the same poststratification cell share the same response probability, we can replace the microdata with cellwise aggregates and employ a binomial model for the sum of the responses in cell *j* as $y^*_j \sim \text{binomial}(n_j, \theta_j)$, where n_j is the sample cell size and $\theta_j = \text{logit}^{-1}(\beta_1 + \beta_2 \text{male}_j + \alpha^{\text{age}}_{\text{a[j]}} + \alpha^{\text{race}}_{\text{r[j]}} + \alpha^{\text{ZIP}}_{\text{s[j]}})$ using the cellwise effects of all factors. The input data for analysis can thus be either microdata or cellwise aggregates.

To generate overall population or subgroup estimates, we combine model predictions within the poststratification cells—in the contingency table of sex, age, race, and ZIP—weighted by the population cell frequencies N_j, which are derived from the linked ACS data in our application. Additionally, we may choose custom poststratification data for specific target populations (e.g., a different country, rather than the U.S.). If we write the expected outcome in cell *j* based on model (1) as $\hat{\theta}_j$ in cell *j*, the population average from MRP is then:

$$\hat{\theta}^{\text{pop}} = \frac{\sum_j N_j \hat{\theta}_j}{\sum_j N_j}.$$

The MRP estimator for county *c* aggregates over covered cells *j* in that county as,

$$\hat{\theta}^{\text{pop}}_s = \frac{\sum_{j \in \text{county c}} N_j \hat{\theta}_j}{\sum_{j \in \text{county c}} N_j}.$$

We implement Bayesian inference for the estimates, where the variance estimates and 95% credible intervals are computed based on the posterior samples.

MRP for time-varying data

As an example of time-varying data, we model weekly PCR testing results. Here, MRP proceeds in two steps: (1) fit a multilevel model to the testing data for incidence incorporating time and covariates, and (2) poststratify using the population distribution of the adjustment variables: sex, age, race, and ZIP codes, where we assume the population distribution is the same during the study period. Hence, the poststratification cell is defined by the cross-tabulation of sex, age, race, ZIP code, and indicators of time in weeks based on the test result dates.

We denote the PCR test result for individual *i* as y_i, where $y_i = 1$ indicates a positive result and $y_i = 0$ indicates negative. Similarly, with poststratification cells, we assume that people in the same cell have the same infection rate and can directly model cellwise summaries. We obtain aggregated counts as the number of tests n_j and the number of positive cases y^*_j in cell *j*. Let $p_j = \Pr(y_{j[i]} = 1)$ be the probability that person *i* in cell *j* tests positive. We account for the PCR testing sensitivity and specificity, where the positivity p_j is a function of the test sensitivity δ, specificity γ, and the true incidence π_j for people in cell *j*:

$$p_j = (1 - \gamma)(1 - \pi_j) + \delta \pi_j. \quad (4)$$

We fit a binomial model for y_j^*, $y_j^* \sim \text{binomial}(n_j, p_j)$ with a logistic regression for π_j with covariates—sex, age, race, ZIP codes, and time in weeks—to allow time-varying incidence in the multilevel model.

$$\text{logit}(\pi_j) = \beta_1 + \beta_2 \text{male}_j + \alpha^{\text{age}}_{\text{a}[j]} + \alpha^{\text{race}}_{\text{r}[j]} + \alpha^{\text{ZIP}}_{\text{s}[j]} + \alpha^{\text{time}}_{\text{t}[j]}, \quad (5)$$

where male_j is an indicator for men; $\text{a}[j]$, $\text{r}[j]$, and $\text{s}[j]$ represent age, race, and ZIP levels; and $\text{t}[j]$ denotes the time in weeks when the test result is collected for cell *j*. We include ZIP-code-level predictors $\vec{Z}_s^{\text{ZIP}}$ for ZIP code *s*,

$$\alpha_s^{\text{ZIP}} = \vec{\alpha}\vec{Z}_s^{\text{ZIP}} + e_s.$$

We assign the same priors in (2) and (3) to varying intercepts and error terms e_s. As to time-varying effects, we assume $\alpha_{\text{t}}^{\text{time}} \sim \text{normal}\,(0, \sigma^{\text{time}})$, with a weakly informative hyperprior, $\sigma^{\text{time}} \sim \text{normal}_+(0, 5)$.

As an example, we assign normal priors to the ZIP-code-level and time-varying effects. The interface leverages Stan's modeling capabilities to allow alternative prior choices and can be extended with advanced modeling [18], such as spatial priors for ZIP-code-level effects or time series priors (e.g., first-order autoregressive) for temporal effects. Alternative outcome models (e.g., negative binomial) can be specified to accommodate overdispersion. In our COVID-19 application, we did not find substantial differences in the examined group estimates with various outcome model and prior specifications, so we presented the results based on a binomial model with normal priors. We elaborate further on model extensions in Sect. "Discussion".

Using the estimated incidence $\hat{\pi}_j$ based on the Bayesian model in (5), we adjust for selection bias by applying the sociodemographic distributions in the community with population cell counts N_j based on the ACS, yielding population-level weekly incidence estimates:

$$\hat{\pi}_t = \frac{\sum_{j \in \text{ week } t} N_j \hat{\pi}_j}{\sum_{j \in \text{ week } t} N_j},$$

which can be restricted to specific subgroups or regions of interest, as another key property of MRP is to yield robust estimates for small groups. We obtain the Bayesian credible intervals from the posterior samples for inference.

Bayesian workflow with MRP

The interface implements an end-to-end Bayesian MRP workflow of statistical analyses, from data preprocessing, descriptive summaries, model fitting, diagnostics, to presentation of results, following the principles of [19]. For illustration, we apply this process to COVID-19 tracking in Michigan and validate the findings in comparison with other studies.

Data preprocessing

The input data can be either individual patient test records or aggregated summaries at the poststratification cell level. The data cleaning and linking process is illustrated by the flowchart in Fig. 1. This workflow automatically imputes missing predictor values using observed frequency distributions, converts categorical variables to factors, and standardizes continuous predictors at the ZIP code level. The MRP integrates three data sources: (1) PCR test results, (2) poststratification cell counts from the ACS, and (3) ZIP-code-level predictors linked from the ACS.

The PCR test records from hospitals include key demographic, geographic, and temporal measures: sex, race, age, five-digit ZIP code, PCR test result, and result date, for the modeling. Balancing operational feasibility,

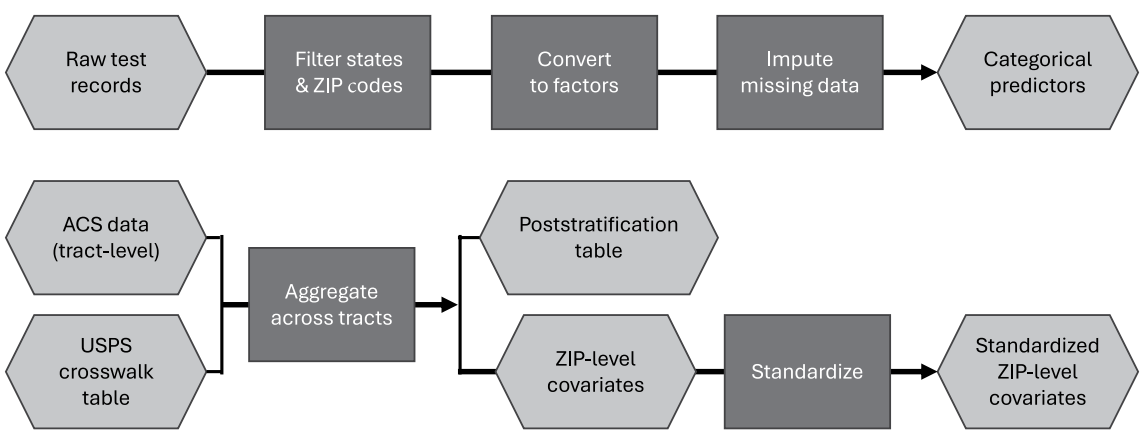

Fig. 1 Data preprocessing flowchart in the interface

timeliness and accuracy, patient records are linked to the 2021 five-year ACS dataset by residential ZIP codes, using the R package `tidycensus` [20]. This linkage serves two purposes: (1) defining the target population as people living in the ZIP codes' catchment area and deriving population counts for poststratification cells, and (2) incorporating area-level predictors of viral infection to adjust for geographic variation. While the ACS reports geography at the levels of census tracts, counties, and states, ZIP codes are defined by the U.S. Postal Service (USPS). We use the ZIP code crosswalk table released by the U.S. Department of Housing and Urban Development and USPS to link ZIP codes to census tracts [21] and calculate the ZIP-code-level measures by aggregating all available tract-level measures weighted by tract population counts. We select the county with the most-overlapping residential addresses for one ZIP code as the ZIP-linked county. The catchment area covered by the list of residence address ZIP codes provided by the Michigan Medicine patients can cover multiple states, beyond Michigan. We filter the data geographically by first removing ZIP codes with five or fewer records and then states that constitute less than 1% of the remaining data. We construct poststratification cells by cross-classifying sex, race, age, and ZIP code and obtain the population counts for these cells from weighted ACS sample distributions in the relevant catchment area. These counts are assumed to remain constant throughout the study period (2020–2022).

The geographic predictors include both individual-level variables (such as education, employment, and income) and tract-level variables (including urbanicity and the Area Deprivation Index [ADI, 22]. These are aggregated to ZIP codes as follows: (1) urbanicity: the percentage of covered census tracts classified as urban, weighted by tract population; (2) college: the percentage of residents with an Associate's degree or higher; (3) poverty: the percentage of residents with incomes below the poverty level in the past year; (4) employment: the percentage of the civilian labor force that is employed; (5) income: the population-weighted average of tract-level median household incomes over the past 12 months; and (6) ADI: the population-weighted average of tract-level ADI values across covered census tracts.

Descriptive statistics

We examine descriptive statistics of observed positivity across time and counties, demographics based on individual records, and characteristics of the covered geographic areas. The observed viral infection shows variation across time, geography, and demographic groups.

Figure 2 presents the highest value among weekly positivities and collected sample sizes across counties, exemplifying the large geographic variation. Most people are from the four counties in Southeast Michigan, where the medical center is located. However, the sample catchment area covers 94 counties. The test positivity among

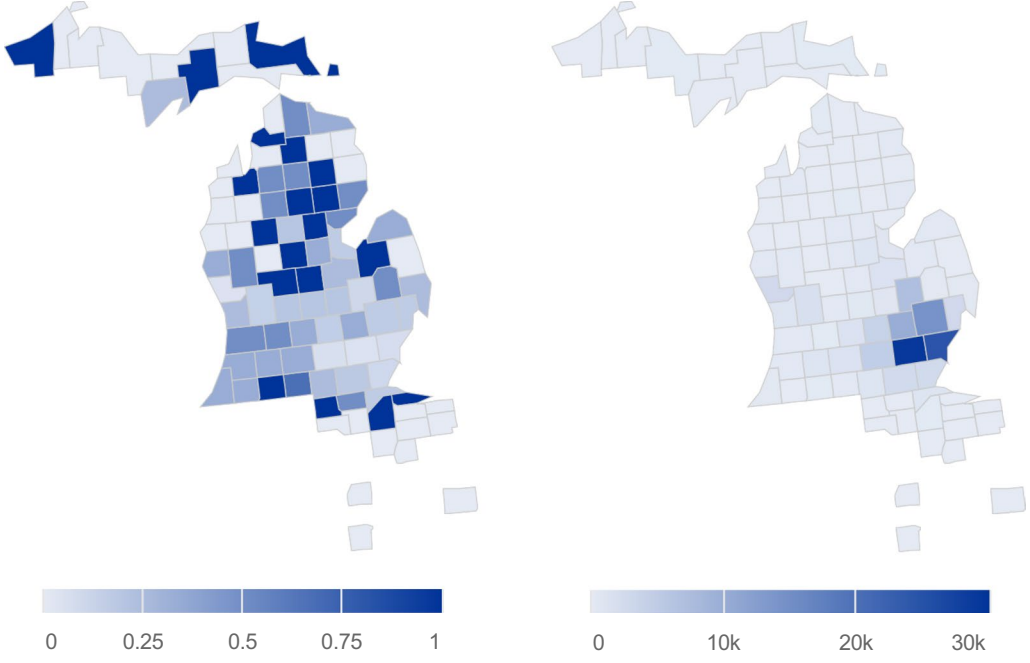

Fig. 2 Highest values among weekly positive response rates (left) and available sample sizes (right) across 94 counties in the catchment area, exemplifying the large geographic variation with sparse data

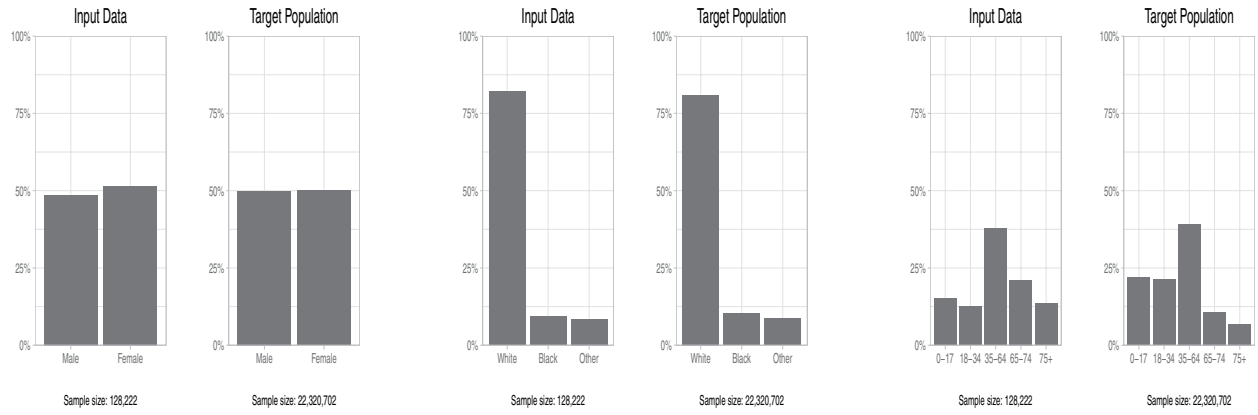

Fig. 3 Comparisons of demographic distributions between the input data of hospital patients and the target population approximated by linked American Community Survey data in the catchment area. The hospital patients over-represent female, White, and older people

asymptomatic patients is often lower than 1%, but greater variability in counties with a small number of tests results in higher than 80% positivity in some cases. The geographically adjacent areas may not share similar peak values.

Figure 3 compares the sex, race, and age distributions between the hospital patients ($n = 128{,}222$) and the population residing in the catchment area ($N = 22{,}320{,}702$). The hospital patients have larger proportions of female, White, and older people than the population, and this sample discrepancy will be adjusted by the poststratification step in MRP.

Figure 9 in the Appendix presents the distributions of geographic characteristics. The catchment area of the hospital patients' residence covers 848 ZIP codes and has a broad and diverse representation in terms of urban/rural areas, area deprivation status, higher education attainment, employment rate, income, and poverty. We expect that socioeconomic measures at the ZIP level would affect individual behaviors and health and be related to viral transmission. The geographic characteristics would explain the spatial variation. The poststratification uses the population counts by ZIP code but does not adjust the geographic characteristics.

Model fitting

We fit the model via Bayesian computation with Markov chain Monte Carlo in Stan and obtain model fit summaries and convergence assessments of the posterior sample of model parameters. We give an example of output from model (5) in Section B of the Appendix. With the interface, we can also specify and fit different models with various choices of individual/geographic covariates and varying/non-varying effects under different prior distributions.

Model diagnostics

We compare different models and present model diagnostic results. We employ the approximate leave-one-out cross-validation (LOO-CV) implemented in the R package `loo` [23] and posterior predictive checking [PPC; 24]. The LOO-CV assesses the posterior predictive performance of Bayesian models and compares different models on expected log predictive density (elpd) for new data.

We have compared the three models with different mean structure and variance specifications, with included predictors listed in Table 1. Model A includes coefficients for sex and geographic predictors and varying effects of age, race, time in weeks, and ZIP code. Model B adds high-order interactions, between race and college attainment status, to Model A. Model C removes the ZIP-varying effects from Model A.

Table 2 gives the LOO-CV outputs on the model comparison. The difference, elpd_diff, will be positive if the expected predictive accuracy for Model B or Model C is higher than that for Model A. The negative elpd_diff values show that Model A has the best predictive performance. The se_diff values support whether the improvement of Model A is substantial. A rule of thumb is to check whether the interval (elpd_diff- 2∗se_diff, elpd_diff + 2∗se_diff) covers the value 0. Hence, we select Model A for inference.

If the model fits the observed data well and preserves the correlation structure, we expect the model to generate replicated data of the observations that mimic the raw values. The weekly replicates use the observed number of tests and estimated positivity based on models (4) and (5) corresponding to each week to generate synthetic counts of positive cases and then the synthetic positive response rates. The PPC in Fig. 4 examines the weekly positivity and compares the raw values to 10 sets of replicates based on posterior predictive samples from Model A.

Table 1 Predictors included in each model. A checkmark indicates inclusion

Model	Male	Age	Race	Time	ADI	College	Employment	Income	Poverty	Urbanicity	Race × College	ZIP-varying effect
A	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
B	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
C	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		

Table 2 Model comparisons with the leave-one-out cross-validation

	elpd_diff	se_diff
Model A	0	0
Model B	-2.05	0.86
Model C	-4.75	3.20

The two columns show the expected log predictive density difference and its standard error, in each case comparing to Model A. The negative elpd_diff values show that Model A has the best predictive performance. The se_diff values support whether the improvement of Model A is substantial

When the number of tests is small, the generated replicates present large variability. Across time, the replicates are close to the observations, showing that the model fits the data well without red flags of failing to capture important structure.

Estimation results

Based on the selected Model A, we present the estimated infection incidence over time for the target population and demographic and geographic subpopulations.

Figure 5 shows the estimated viral transmission rate by week in the catchment area. We observe spikes in November 2020, January 2022, and August 2022. The MRP estimates are generally higher than the raw positivity, mainly because of the test sensitivity, where 70% of infections are tested positive. MRP matches the sample demographics with those in the population.

MRP stabilizes small group estimates and adjusts for the sample discrepancy within each group. Figure 6 presents the estimated incidence for White, Black, and other race categories. Whites tend to have lower infection rates than Black and other racial groups, even though most weekly differences are small and not significant. The model does not include time trends varying across racial groups, i.e., without racial moderation effects. The estimated trends are close to paralleling with similar spike and flat periods. Examining racial differences by week, we expect the differences to be small because of weekly small numbers of tests. When we calculate the cumulative incidences through the study period, Whites are less likely to be infected than Black and other racial groups, which is consistent with the literature findings [25]. The same observation of trends applies to the sex and age group estimates, given in Figs. 10 and 11 of the Appendix.

We select one week that observes a spike of infection, 01/31/2022–02/06/2022, and present the county-level estimated incidence with standard error (SE) values in Fig. 7. The collected test records across counties are sparse, where 32 out of 94 counties do not have any tests and 28 counties have only one test during the selected week. The county with the largest number of tests (230) is where the health system is located. The observed

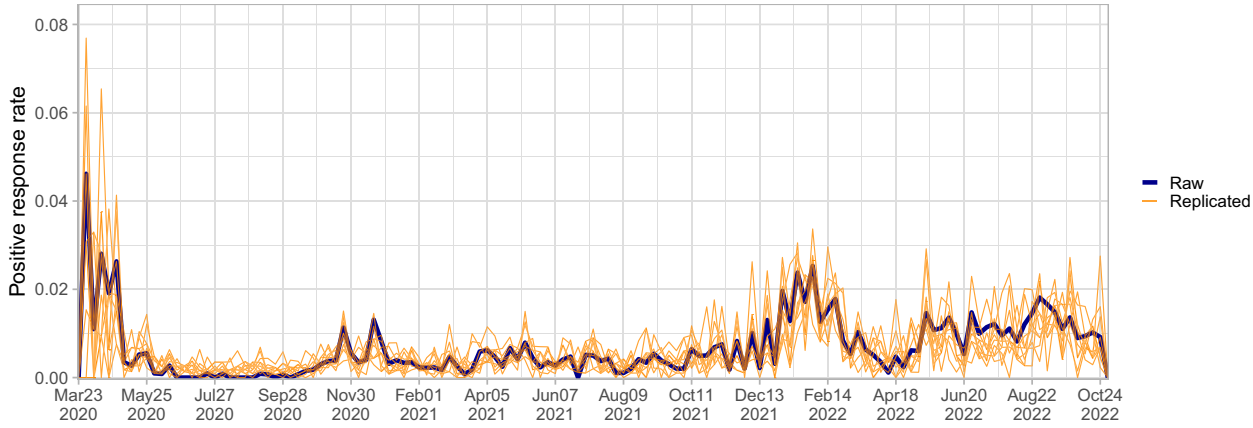

Fig. 4 Posterior predictive check comparing replicated positive response rates generated from the estimated model to raw rates by week. The replicates are close to the raw data, showing that the model fits the data well

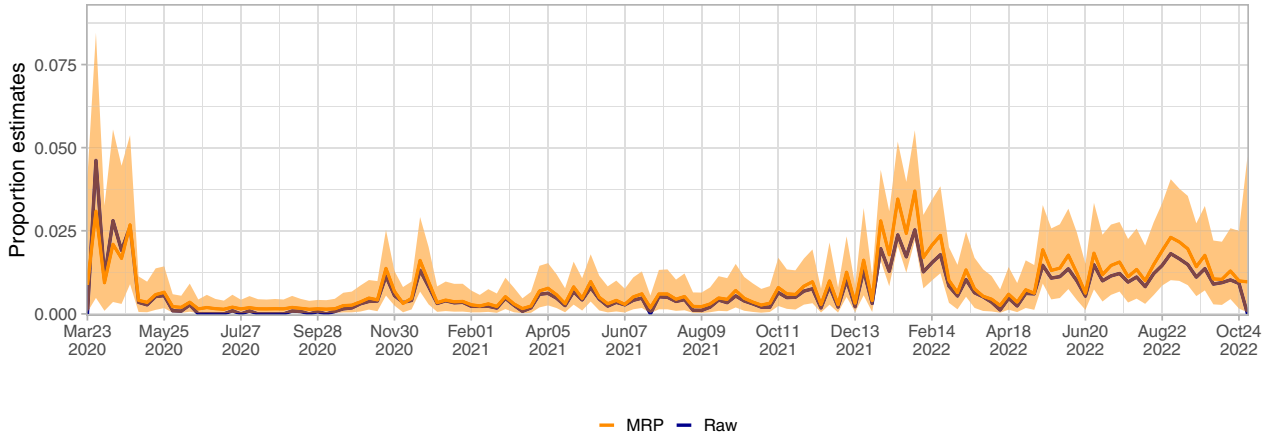

Fig. 5 Estimated weekly incidence in the community based on the multilevel regression and poststratification (MRP) in comparison with raw values. The shaded areas represent 95% credible intervals. The MRP estimates match the sample demographics with those in the population and are generally higher than the raw positivity, mainly because of the test sensitivity

positivity values are unreliable, and 52 out of 62 collected values are zero. Among the top five counties that report the largest numbers of tests and any positive cases during the studied week, the demographic distributions of tested patients are generally similar, with an over-representation of female, White, and older people compared to the ACS data. The MRP estimates are available for all 94 counties based on the predictions with the ACS data. The model fit summaries in the Appendix A show that the estimated coefficient of urbanicity (defined as the percentage of covered census tracts classified as urban, weighted by tract population) is -0.10 with the 95% interval of $(-0.21, 0.02)$, which shows that the urban areas tend to have lower infection rates than the non-urban areas. The multilevel model smooths county-level incidence estimates with a range of 0.1%–5.6%, with a median value of 0.6%. The SE values of the 94 county-wise incidence estimates are between 0.001 and 0.037, and the variation generally increases with the estimated incidence.

Validation and comparison

Our surveillance tool leverages routine hospital testing of asymptomatic patients to provide an early indicator of community disease presence, serving a similar function to wastewater monitoring for SARS-CoV-2 in public sewer systems. By tracking trends over time, both tools can detect increases in SARS-CoV-2 prevalence, thereby alerting health agencies to potential surges in cases and an increased clinical burden. We have compared our results to data from the Michigan Wastewater Dashboard for COVID-19 surveillance [26] and to weekly percentages of Emergency Department (ED) visits diagnosed as COVID-19 in Michigan, as reported by the CDC COVID Data Tracker [27]. Figure 8 demonstrates that

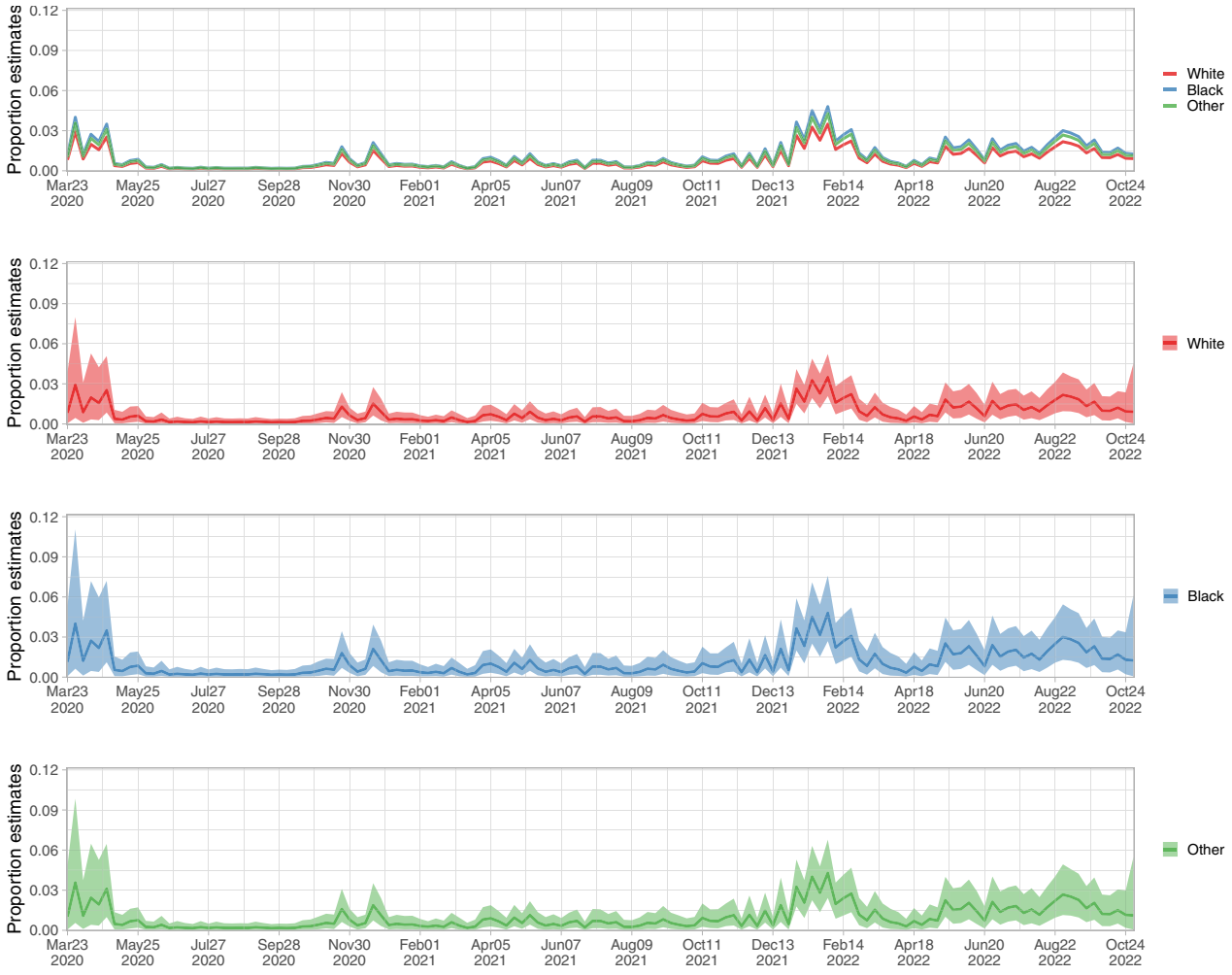

Fig. 6 Estimated weekly incidence across racial groups based on the multilevel regression and poststratification. Whites tend to have lower infection rates than Black and other racial groups, even though most weekly differences are small. The shaded areas represent 95% credible intervals

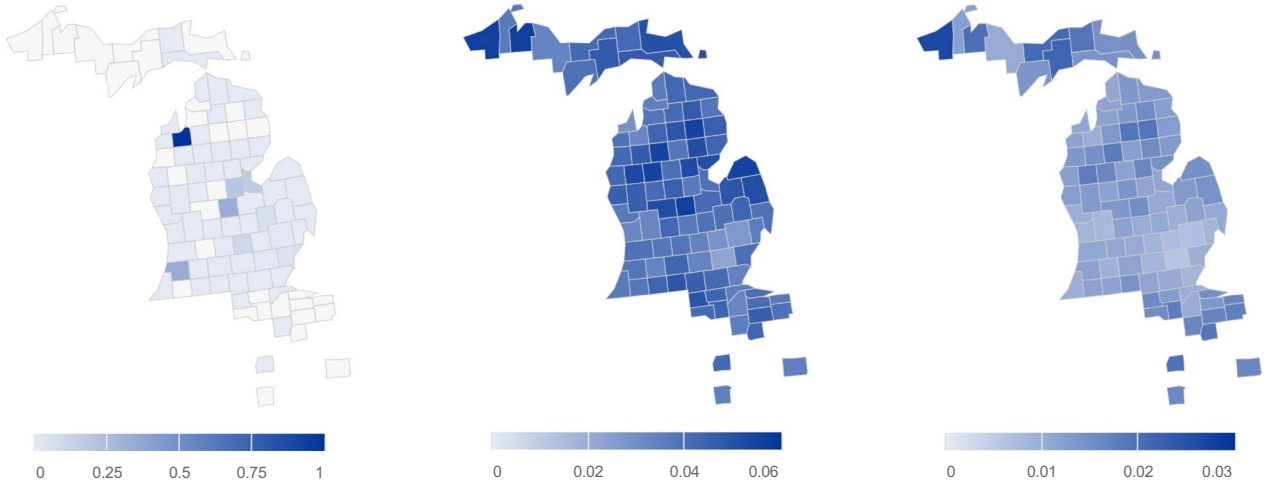

Fig. 7 Observed and estimated county-level incidence with standard errors during the week of 01/31/2022–02/06/2022 in the catchment area. The model estimates reduce between-county variation, and their standard errors generally increase with the estimated incidence values

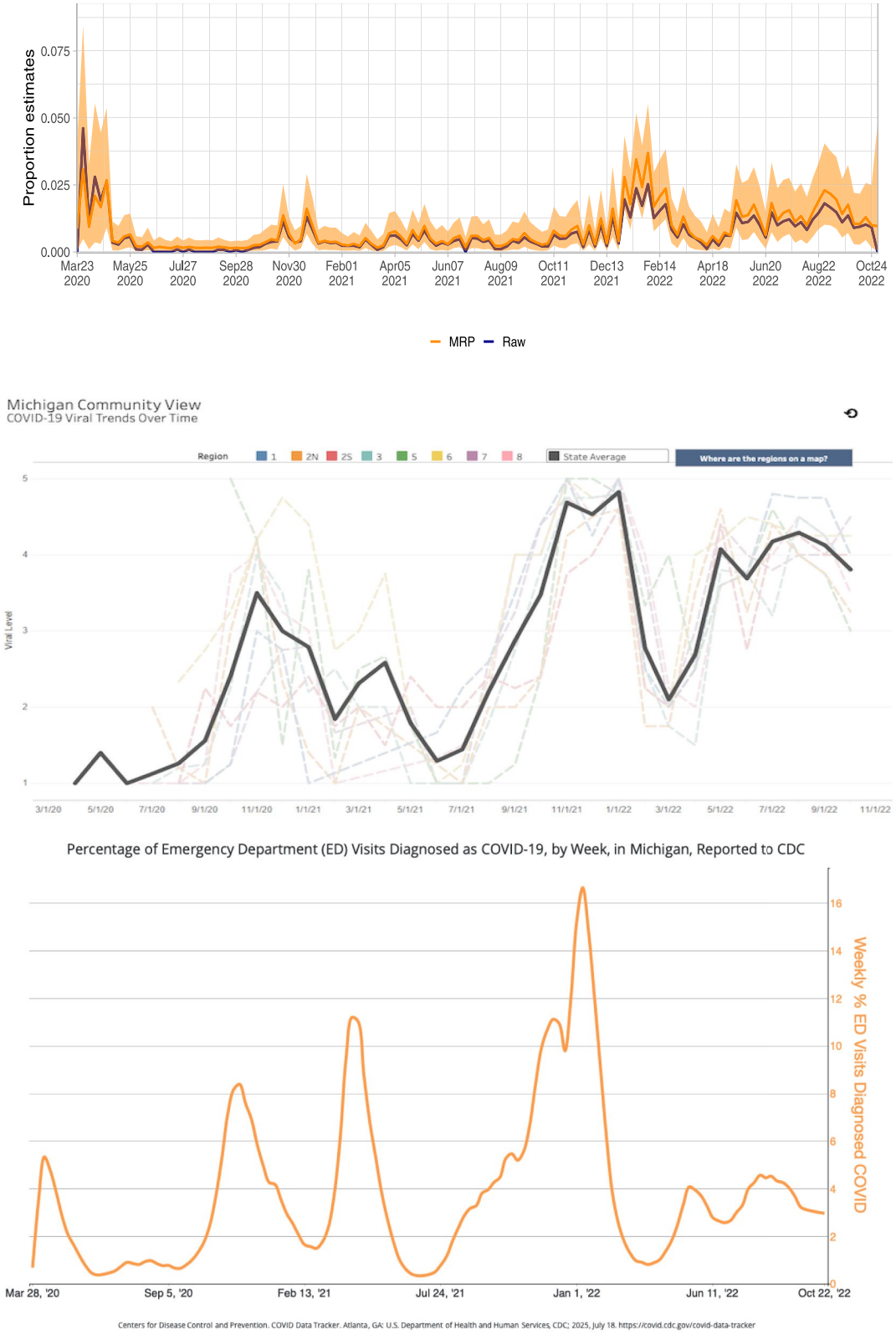

Fig. 8 Comparison of COVID-19 tracking trends estimated by applying multilevel regression and poststratification (MRP) to Michigan hospital test data, the Michigan wastewater monitored COVID-19 surveillance and weekly percentages of Emergency Department (ED) visits diagnosed as COVID-19 in Michigan, showing an alignment of transmission trends and early warning of clinical burdens

our estimated trends between March 22, 2020, and October 24, 2022, closely align with wastewater-based surveillance for SARS-CoV-2 shed into Michigan's public sewer systems. Both surveillance methods capture the spikes in November 2020 and January 2022, as well as the downward trend since August 2022. Notably, our surveillance approach can anticipate increases in ED visit numbers reported to the COVID Data Tracker by approximately one to two weeks. This validation supports findings previously reported in Indiana [2, 3].

Our results based on a representation adjustment of routine hospital test records serve a synthetic proxy for random sampling. When available, random-sample testing surveys provide valuable benchmarking data and should be leveraged to calibrate other data sources to ensure population representativeness [28, 29]. However, increasing nonresponse rates in these surveys necessitate nonresponse bias adjustments [30, 31]. For example, [32] have applied a similar method to MRP and adjusted for nonresponse bias in a randomized study of COVID-19 testing in Indiana, the response rate of which is 23.6%. Notably, the trends in calibrated new infection numbers reported by [28] are consistent with those seen in the MRP-adjusted hospital test incidence monitoring [2] between March 2020 and March 2021 in Indiana, particularly regarding the capture of infection spikes.

Discussion

With generalizability as the goal, the MRP method adjusts for selection bias and stabilizes small group estimates. We extend MRP to both cross-sectional and longitudinal data and facilitate subgroup trend analyses over time with a user-friendly interface. A key contribution of our MRP interface is its capacity to standardize statistical workflows and facilitate reproducibility. By integrating all analytic steps—from data preprocessing and model fitting to diagnostics and result visualization—into a comprehensive pipeline, the platform makes it straightforward for others to replicate findings and validate results on new datasets. The interface is built using open-source tools (Stan, R, Shiny) and is publicly available, supporting transparency and broad accessibility. By making MRP methods widely available without substantial coding requirements, the interface empowers a wider range of users to implement advanced statistical methods and strengthens training in reproducible population health analytics. These advantages position the interface as a foundational resource for standardized use and methodological rigor in future epidemic surveillance and related scientific studies.

The interface tracks the COVID-19 epidemic and delivers substantive findings in time, which is demonstrated using the Michigan health system data. The findings based on the MRP workflow have implications for public health monitoring and policy decision-making. By correcting sample selection bias and generating stable estimates for small demographic and geographic subgroups, the approach can identify populations at higher risk of infection or with lower access to healthcare resources. Health agencies can use subgroup-specific incidence estimates to allocate testing resources and direct targeted communication to communities with greater vulnerability. Tracking epidemiological trends using routine hospital data provides actionable guidance for predicting surges, anticipating clinical burden and responding swiftly to local outbreaks. Furthermore, the analytic pipelines support transparency and comparability in health reporting, which is vital for evidence-based policymaking and resource allocation.

Our approach has several key assumptions. First, we assume sample selection is ignorable, conditional on adjusted demographics and geography. The only factor determining the sample inclusion is the selection for elective surgical procedures. Elective surgery patients may differ from the broader community in unmeasured ways, such as healthcare access or overall health status. However, we need high-quality population data on healthcare use and health measures to adjust these potentially confounding factors. Second, we assume a constant ratio of asymptomatic to symptomatic SARS-CoV-2 infections within any demographic and geographic stratum. We use the estimated incidence based on asymptomatic test records to track the infection trend, but not the magnitude. The ratios may change values with new viral variants and immunity levels that are naturally-acquired or vaccine-induced. We have conducted sensitivity analyses by including the time indicators when the new viral was first detected in Michigan in the model and found that the estimated trends are similar, though the incidence rates have slightly shifted. We have also applied the MRP adjustment to viral IgG testing data of the same group of asymptomatic patients and validated the method using verified clinical metrics of viral and symptomatic disease incidence to show the expected biological correlation of these entities with the timing, rate, and magnitude of seroprevalence [3]. Third, the model-based adjustment is subject to model misspecification. We use a Bayesian binomial model, and our subgroup estimates are robust across different outcome models. It is possible that the model fails to capture some data structure. We suggest users conduct thorough model diagnostics, such as the PPC and LOO-CV in our paper, and result validation. Additionally, the interface's current focus is infection incidence estimation by subgroup, but it is extensible to

other epidemiological parameters (e.g., effective reproduction number, infection fatality ratio), which would require integrating individual-level test data with aggregate case and mortality data using hierarchical Bayesian frameworks. Despite these limitations, the flexibility of the MRP approach allows its use with a variety of sampling methods and data sources, supporting broader applicability beyond the specific settings tested so far. Though post-epidemic testing is currently paused, the interface has broad applicability for other disease monitoring and data analyses with population representation.

With respect to our current sampling method for COVID-19 viral tracking, in accordance with accepted American Society of Anesthesiology standards, all preoperative patients in the hospital system are subjected to surgical risk evaluation. Hence, routine hospital testing is already implemented uniformly across the U.S., increasing the operational feasibility of our proposed surveillance system. Previous work used data from a community hospital in Indiana [2], and the results show differences between the states. As shown in our county-level estimates in Michigan, geographic variation can be substantial. Expanding to data from more states will further enhance national generalizability. Empowering users of varying backgrounds to analyze their own localized data effectively, the workflow can be directly adopted by local, state, and national health departments to ensure methodological consistency.

Originally developed in response to COVID-19, the MRP interface provides a foundation for broader epidemic surveillance and diverse applications in health and social science research. It accommodates time-varying and cross-sectional data, continuous and binary outcomes, and supports subgroup analyses across demographic and geographic domains. Users can specify models, priors, and poststratification data, and analyze probability sample surveys, non-probability samples, and multiple data sources. Future enhancements will address complex spatiotemporal structures (e.g., integrating real-time individual and aggregate testing data with autoregressive structures across transmission neighborhoods), custom prior distributions, and poststratification with incomplete population data. The MRP workflow is broadly applicable across a range of public health surveillance contexts. For example, recent measles and pertussis outbreaks in various parts of the U.S. underscore the need for flexible, reliable surveillance systems that can rapidly estimate disease prevalence and identify at-risk populations [33, 34]. By facilitating subgroup estimation and adjusting for selection bias, our MRP interface could be readily adopted for outbreak response and ongoing surveillance of many infectious diseases, where sample representativeness and rapid trend analysis are critical. The potential to use routinely collected hospital data, school outbreaks, or sentinel surveillance and combine them with population-level information makes our framework adaptable beyond COVID-19, reinforcing its generalizability for future epidemic preparedness and monitoring of emerging health threats.

In summary, the MRP interface advances statistical methods for epidemic surveillance by combining methodological rigor, computational accessibility, and practical impact. As public health agencies move toward more data-driven and adaptive response models, tools that enable granular, timely, and reproducible estimation will become increasingly essential. By establishing a foundation for standardized population health analytics, our approach aims to support ongoing efforts in epidemic preparedness, resource optimization, and evidence-based policy.

Appendix A Summary of geographic characteristics

Figure 9 gives the distributions of geographic characteristics.

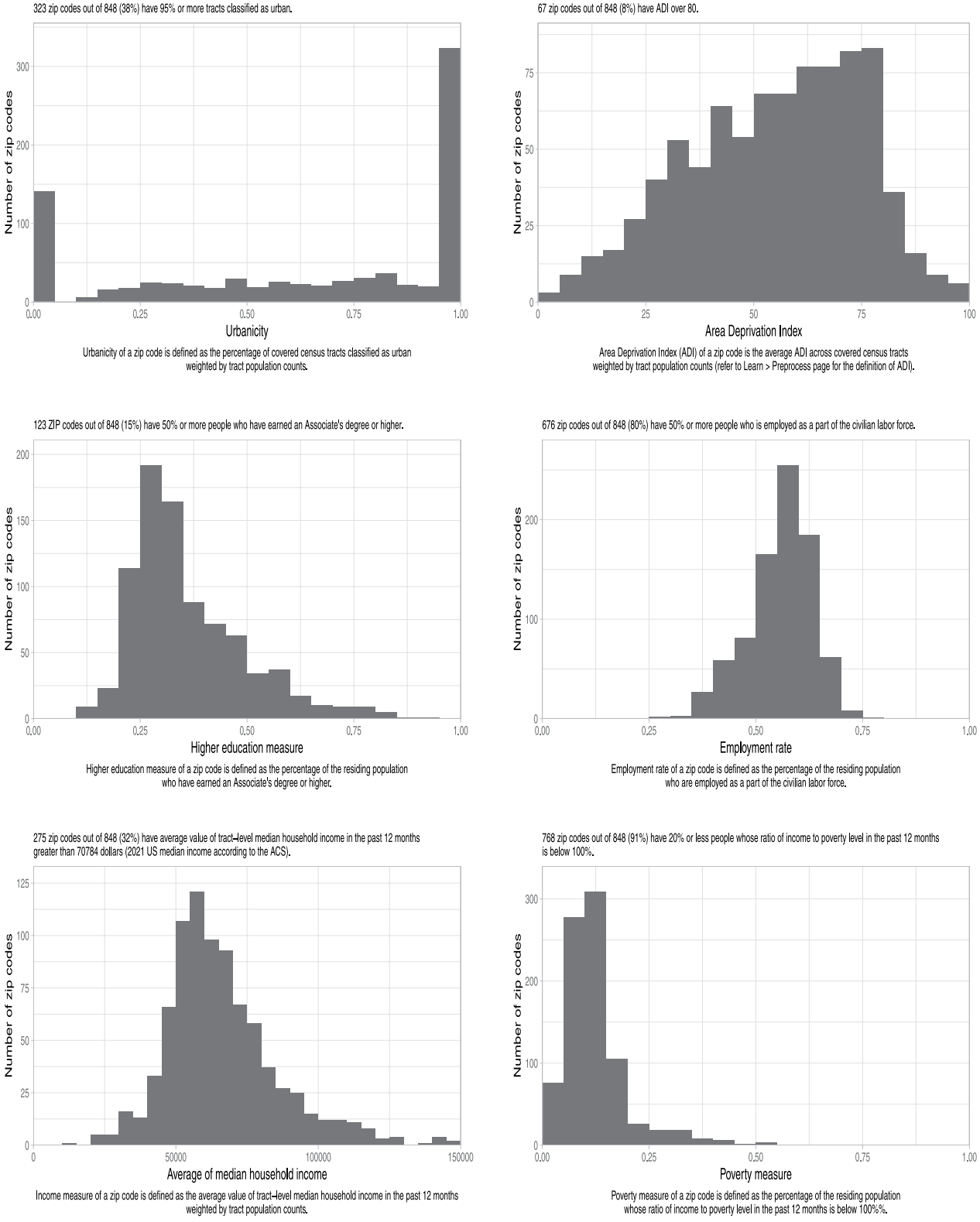

Fig. 9 Distributions of geographic characteristics based on the linked American Community Survey data in the catchment area

Appendix B Model fitting results

The Stan fit summaries of Model A are displayed in Table 3.

A binomial model with a logit function of the positive response rate. Samples are generated using 4 chains with 2500 post-warmup iterations each.

Table 3 Model A fit summaries

Non-varying effects							
	Estimate	**Est.Error**	**l-95%**	**u-95%**	**R-hat**	**Bulk_ESS**	**Tail_ESS**
Intercept	-5.45	0.45	-6.37	-4.50	1.00	1629	1498
Sex.male	0.19	0.10	0.00	0.38	1.00	10062	5779
Urbanicity	-0.10	0.06	-0.21	0.02	1.00	6723	5479
College	-0.16	0.12	-0.40	0.08	1.00	3681	5508
Employment	0.03	0.08	-0.13	0.19	1.00	6066	6221
Poverty	-0.07	0.10	-0.27	0.12	1.00	4295	5208
Income	-0.11	0.15	-0.40	0.18	1.00	4385	4825
ADI	0.07	0.10	-0.13	0.28	1.00	5272	1902
Standard deviations of varying effects							
	Estimate	**Est.Error**	**l-95%**	**u-95%**	**R-hat**	**Bulk_ESS**	**Tail_ESS**
Race (intercept)	0.45	0.43	0.04	1.72	1.00	1941	1404
Age (intercept)	0.63	0.38	0.24	1.64	1.00	1843	3246
Time (intercept)	1.13	0.12	0.92	1.37	1.00	1882	4113
ZIP (intercept)	0.36	0.11	0.11	0.57	1.01	989	596

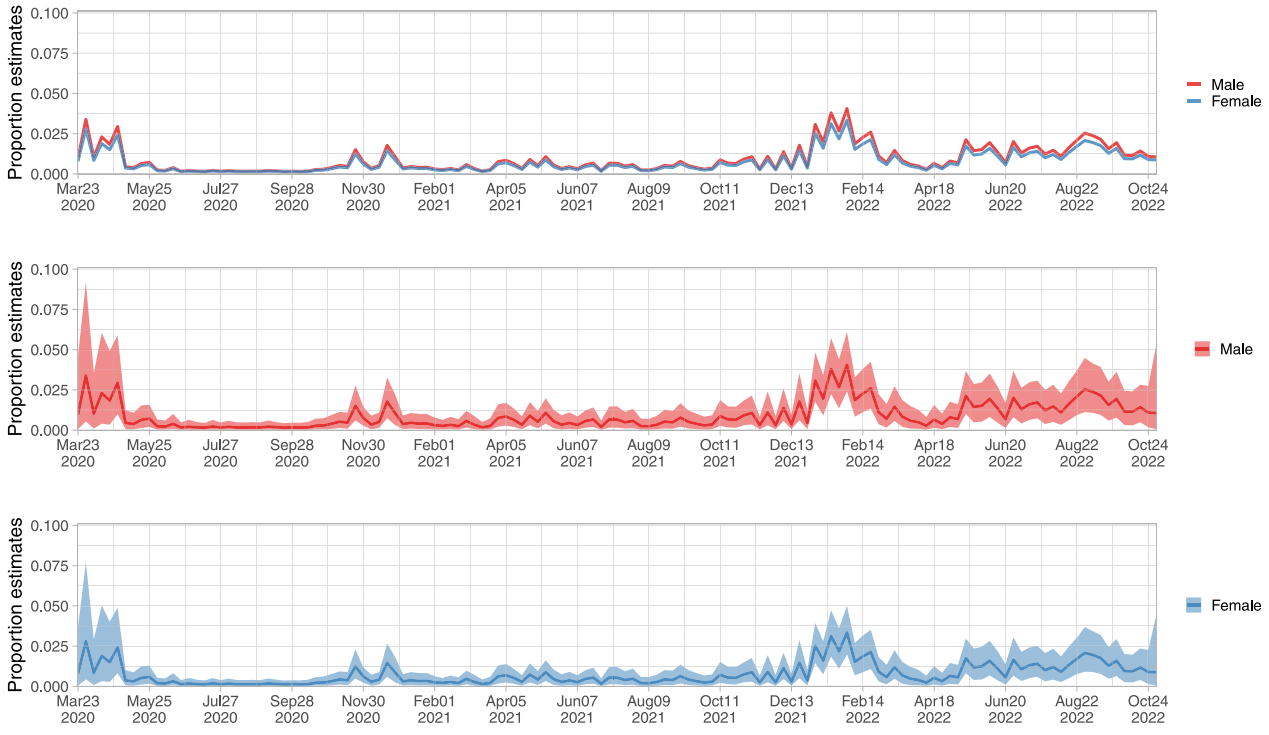

Fig. 10 Estimated weekly incidence by sex based on the multilevel regression and poststratification. Females tend to have lower infection rates than males with small weekly differences. The shaded areas represent 95% credible intervals

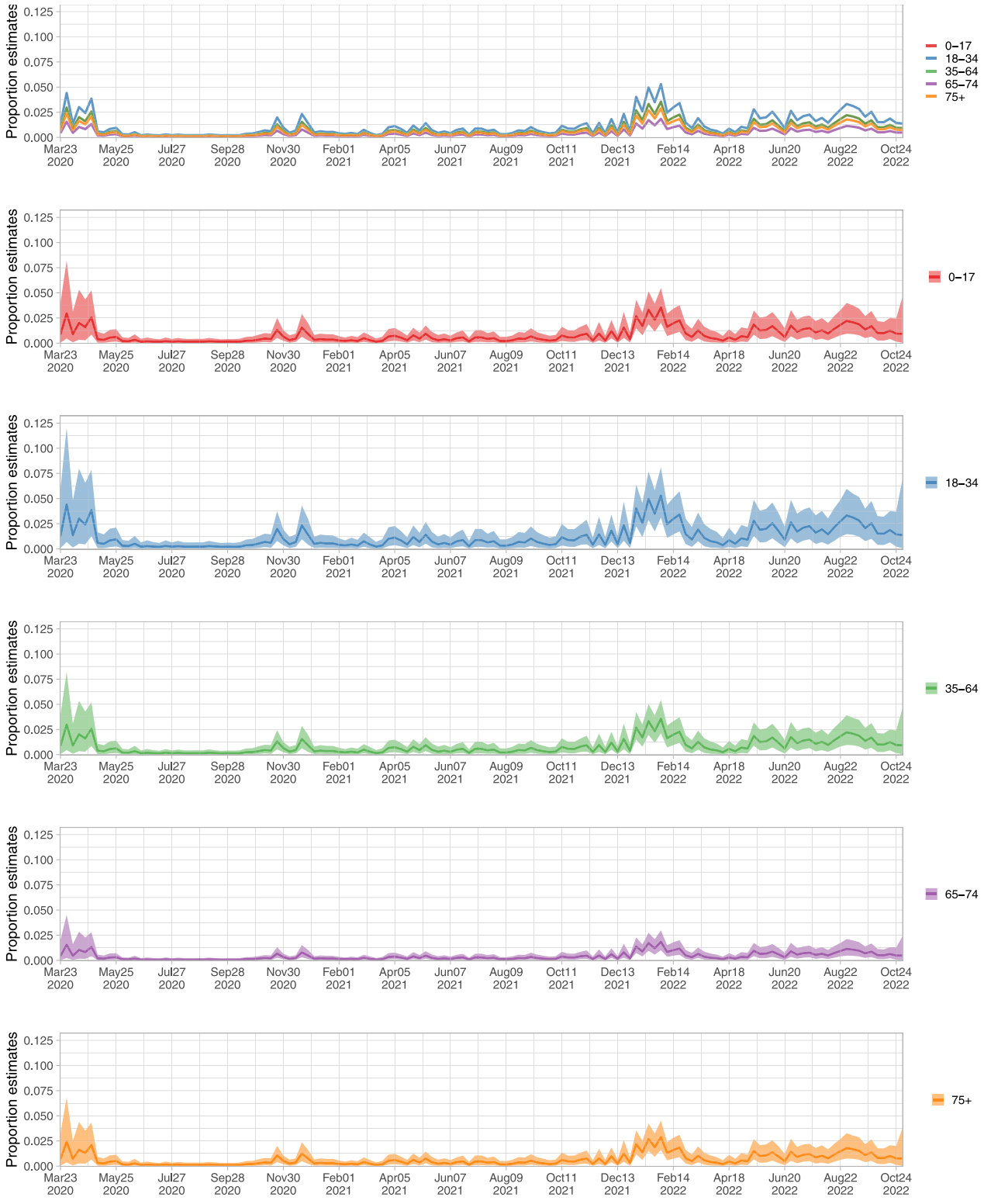

Fig. 11 Estimated weekly incidence by age group based on the multilevel regression and poststratification. Young adults tend to have lower infection rates than elders with small differences during most weeks. The shaded areas represent 95% credible intervals

Appendix C Model estimates

Figures 10 and 11 present the weekly incidence estimates by sex and age groups, respectively.

Acknowledgements
This work is supported by the National Institutes of Health through grant U01MD017867.

Author contributions
Y.S. wrote the main manuscript. T. T. conducted analyses and prepared figures. All authors reviewed the manuscript.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Competing interests
The authors declare no competing interests.

Received: 9 May 2024 / Accepted: 14 January 2026
Published online: 03 February 2026

References

1. Gelman A, Little TC. Poststratification into many categories using hierarchical logistic regression. Surv Methodol. 1997;23:127–35.
2. Covello L, Gelman A, Si Y, Wang S. Routine hospital-based SARS-CoV-2 testing outperforms state-based data in predicting clinical burden. Epidemiology. 2021;32(6):792–9.
3. Si Y, Covello L, Wang S, Covello T, Gelman A. Beyond vaccination rates: a synthetic random proxy metric of total SARS-CoV-2 immunity seroprevalence in the community. Epidemiology. 2022;33(4):457–64.
4. Zhang X, Holt JB, Yun S, Lu H, Greenlund KJ, Croft JB. Validation of multilevel regression and poststratification methodology for small area estimation of health indicators from the behavioral risk factor surveillance system. Am J Epidemiol. 2015;182(2):127–37.
5. Si Y, Trangucci R, Gabry JS, Gelman A. Bayesian hierarchical weighting adjustment and survey inference. Surv Methodol. 2020;46(2):181–214.
6. Wang W, Rothschild D, Goel S, Gelman A. Forecasting elections with non-representative polls. Int J Forecast. 2015;31(3):980–91.
7. Si Y. On the use of auxiliary variables in multilevel regression and poststratification. Stat Sci. 2025;40(2):272–88.
8. Lauderdale BE, Bailey D, Blumenau J, Rivers D. Model-based pre-election polling for national and sub-national outcomes in the US and UK. Int J Forecast. 2020;36:399–413.
9. Downes M, Gurrin LC, English DR, Pirkis J, Currier D, Spittal MJ, et al. Multilevel regression and poststratification: a modeling approach to estimating population quantities from highly selected survey samples. Am J Epidemiol. 2018;187(8):1780–90.
10. Downes M, Carlin J. Multilevel regression and poststratification versus survey sample weighting for estimating population quantities in large population health studies. Am J Epidemiol. 2020;189(7):717–25.
11. Li K, Si Y. Embedded multilevel regression and poststratification: model-based inference with incomplete poststratifier information. Stat Med. 2024;43(2):256–78.
12. Stan Development Team: Stan: A C++ library for probability and sampling. http://mc-stan.org 2024.
13. Gabry J, Češnovar R, Johnson A, Bronder S. cmdstanr: R Interface to cmdStan. https://mc-stan.org/cmdstanr R package version 0.8.1, https://discourse.mc-stan.org 2024.
14. Tran T, Gabry J, Si Y. shinymrp: Interface for Multilevel Regression and Poststratification. https://doi.org/10.32614/CRAN.package.shinymrp 2025.
15. U. S. Bureau of the Census: The American Community Survey. https://www.census.gov/programs-surveys/acs 2025.
16. Gelman A, Carpenter B. Bayesian analysis of tests with unknown specificity and sensitivity. J R Stat Soc Ser C Appl Stat. 2020;69(5):1269–83.
17. Bendavid E, Mulaney B, Sood N, Shah S, Bromley-Dulfano R, Lai C, et al. Covid-19 antibody seroprevalence in Santa Clara County. California International Journal of Epidemiology. 2021;50(2):410–9.
18. Si Y, Pillai NS, Gelman A. Bayesian nonparametric weighted sampling inference. Bayesian Anal. 2015;10(3):605–25.
19. Gelman A, Vehtari A, Simpson D, Margossian CC, Carpenter B, Yao Y, Kennedy L, Gabry J, Bürkner P-C, Modrák M. Bayesian workflow. https://arxiv.org/abs/2011.01808 2020.
20. Walker K, Herman M, Eberwein K, Walker MK. R Package 'tidycensus' Version 1.7.3. https://cran.r-project.org/web/packages/tidycensus/index.html 2025.
21. U.S. Department of Housing and Urban Development: HUD–USPS ZIP Code Crosswalk Files: ZIP–Tract. https://www.huduser.gov/portal/datasets/usps_crosswalk.html 2023.
22. Kind AJ, Buckingham WR. Making neighborhood-disadvantage metrics accessible–the Neighborhood Atlas. N Engl J Med. 2018;378(26):2456.
23. Vehtari A, Gabry J, Magnusson M, Yao Y, Bürkner P-C, Paananen T, Gelman A. loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. R package version 2.7.0, https://mc-stan.org/loo/ 2024.
24. Gelman A, Meng XL, Stern H. Posterior predictive assessment of model fitness via realized discrepancies. Stat Sin. 1996;6:733–807.
25. Magesh S, John D, Li WT, Li Y, Mattingly-app A, Jain S, et al. Disparities in COVID-19 outcomes by race, ethnicity, and socioeconomic status: a systematic review and meta-analysis. JAMA Netw Open. 2021;4(11):2134147. https://doi.org/10.1001/jamanetworkopen.2021.34147.
26. Michigan Department of Health & Human Services: Michigan Wastewater Dashboard for COVID-19 surveillance. https://www.michigan.gov/coronavirus/stats/wastewater-surveillance/wastewater-surveillance-for-covid-19/dashboard 2025.
27. Centers for Disease Control and Prevention: Trends in United States COVID-19 Deaths, Emergency Department (ED) Visits, and Test Positivity by Geographic Area. https://covid.cdc.gov/covid-data-tracker/#trends_select_7dayeddiagnosed_26 2025.
28. Irons NJ, Raftery AE. Estimating SARS-CoV-2 infections from deaths, confirmed cases, tests, and random surveys. Proc Natl Acad Sci U S A. 2021;118(31):2103272118.
29. Menachemi N. Population point prevalence of SARS-CoV-2 infection based on a statewide random sample–Indiana, April 25–29, 2020. MMWR Morb Mortal Wkly Rep. 2020;69(29):960–4.
30. Si Y, Little RJA, Mo Y, Sedransk N. A case study of nonresponse bias analysis in educational assessment surveys. J Educ Behav Stat. 2023;48(3):271–95.
31. Si Y, Little RJA, Mo Y, Sedransk N. Nonresponse bias analysis in longitudinal studies: a comparative review with an application to the Early Childhood Longitudinal Study. Int Stat Rev. 2024;92(3):383–405.
32. Yiannoutsos CT, Halverson PK, Menachemi N. Bayesian estimation of SARS-CoV-2 prevalence in Indiana by random testing. Proc Natl Acad Sci U S A. 2021;118(5):2013906118.
33. Centers for Disease Control and Prevention: Measles Cases and Outbreaks. https://www.cdc.gov/measles/cases-outbreaks.html 2025.
34. Centers for Disease Control and Prevention: About Trends in Whooping Cough Disease. https://www.cdc.gov/pertussis/outbreaks.html 2025.

Publisher's Note