Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Application of Bayesian spatial-temporal models for estimating unrecognized COVID-19 deaths in the United States

Yuzi Zhang\textsuperscript{a,}\textsuperscript{*}, Howard H. Chang\textsuperscript{a}, A. Danielle Iuliano\textsuperscript{b}, Carrie Reed\textsuperscript{b}

\textsuperscript{a}Department of Biostatistics and Bioinformatics, The Rollins School of Public Health of Emory University, 1518 Clifton Rd. N.E., Atlanta, GA 30322, USA

\textsuperscript{b}Influenza Division, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA

\textbf{A B S T R A C T}

In the United States, COVID-19 has become a leading cause of death since 2020. However, the number of COVID-19 deaths reported from death certificates is likely to represent an underestimate of the total deaths related to SARS-CoV-2 infections. Estimating those deaths not captured through death certificates is important to understanding the full burden of COVID-19 on mortality. In this work, we explored enhancements to an existing approach by employing Bayesian hierarchical models to estimate unrecognized deaths attributed to COVID-19 using weekly state-level COVID-19 viral surveillance and mortality data in the United States from March 2020 to April 2021. We demonstrated our model using those aged \geq 85 years who died. First, we used a spatial–temporal binomial regression model to estimate the percent of positive SARS-CoV-2 test results. A spatial–temporal negative-binomial model was then used to estimate unrecognized COVID-19 deaths by exploiting the spatial–temporal association between SARS-CoV-2 percent positive and all-cause mortality counts using an excess mortality approach. Computationally efficient Bayesian inference was accomplished via the Polya-Gamma representation of the binomial and negative-binomial models. Among those aged \geq 85 years, we estimated 58,200 (95\% CI: 51,300, 64,900) unrecognized COVID-19 deaths, which accounts for 26\% (95\% CI: 24\%, 29\%) of total deaths.

\* Corresponding author.

\textit{E-mail address:} yuzi.zhang@emory.edu (Y. Zhang).

https://doi.org/10.1016/j.spasta.2021.100584

2211-6753/© 2021 Elsevier B.V. All rights reserved.
COVID-19 deaths in this age group. Our modeling results suggest that COVID-19 mortality and the proportion of unrecognized deaths among deaths attributed to COVID-19 vary by time and across states.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

The Coronavirus Disease 19 (COVID-19) pandemic caused by the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has profoundly affected the world. It is reported that at least 771,000 people have died from COVID-19 in the United States as of November 23, 2021 (Centers for Disease Control and Prevention, 2021). This statistic is based on COVID-19 death data reported by state and county-level jurisdictions to the Centers for Disease Control and Prevention (CDC) and provisional death counts from the National Center for Health Statistics (NCHS). Currently, COVID-19 cases, hospitalizations, and deaths are nationally notifiable and captured by the National Notifiable Disease Surveillance System (NNDSS). However, as noted in previous studies, reporting is unlikely to capture all deaths related to SARS-CoV-2 infections (Woolf et al., 2020, 2021; Rossen et al., 2020; Rivera et al., 2020; Iuliano et al., 2021). Deaths may not be reported because of delays in reporting, failure to access medical care, false negative test results, or the cause of death being coded as a comorbid condition or a complication of COVID-19 (Vandoros, 2020). There is a pressing need to develop methods to estimate unrecognized deaths attributable to COVID-19. Accurate health burden estimates can guide allocation of public health resources, provide insights into spatial and temporal disease trends, and help in communicating effective prevention messages.

During the COVID-19 pandemic, an excess mortality approach has been widely adapted to estimate deaths attributable to COVID-19 (Woolf et al., 2020; Weinberger et al., 2020; Rivera et al., 2020; Rossen et al., 2020). This method has also been used for estimating deaths associated with other infectious diseases, including influenza, pneumonia, and hepatitis C virus infections (Cohen et al., 2010; Lin and Nichol, 2001; Neal, 2007). A critical component of the excess mortality approach is the estimation of baseline deaths (i.e., the number of deaths assuming there is no SARS-CoV-2 mortality). Total COVID-19 deaths are derived from comparing the estimated baseline deaths to observed deaths during the pandemic. In previous studies of COVID-19 deaths, various models have been used to estimate baseline deaths (Tatar et al., 2021; Vandoros, 2020). For example, Woolf et al. (2020), Weinberger et al. (2020), and Rivera et al. (2020) all used Poisson regression models with different covariates. More recently, a linear mixed model that accounts for temporal correlations in weekly mortality data was used in Belgium and the Netherlands (Verbeeck et al., 2021). Notably, the majority of previous studies estimated baseline deaths relying solely on pre-pandemic mortality data, making the assumption that historical mortality trends remain unchanged during the pandemic.

In a recent work by Iuliano et al. (2021), the number of baseline deaths is estimated by leveraging proxy data on infection activity, specifically the percent of positive SARS-CoV-2 test results. This approach models weekly all-cause deaths during the pandemic, after removing known COVID-19 deaths, as a function of lagged weekly percent positive. The main objective of our analysis was to extend the approach by Iuliano et al. (2021) in several important ways. First, we addressed the issue of small numbers of SARS-CoV-2 tests in the early phase of the outbreak by modeling positive test results using a Bayesian spatial–temporal binomial model to borrow information across geographical regions and epidemic weeks. Second, we evaluated the importance of propagating uncertainties in estimated percent positive in subsequent mortality analyses. Third, we examined the use of a spatial–temporal negative binomial model to estimate baseline mortality and unrecognized COVID-19 deaths; previous analyses often ignored potential spatial dependence. Finally, we applied the modeling framework to analyze weekly US COVID-19 deaths and surveillance data for the population ages ≥ 85 years in the 48 contiguous US states and the District of Columbia.
This age group was chosen to demonstrate this model because it had the largest COVID-19 mortality burden as measured by both total death counts and death rate previous estimated by Iuliano et al. (2021).

2. Data

All-cause and COVID-19 deaths, coded using International Classification of Diseases, 10th Edition (ICD-10), were extracted from the National Vital Statistics System (NVSS) for the population $\geq 85$ years of age between March 22, 2020 to April 25, 2021. Data were aggregated by epidemiologic week and by state, including the 48 contiguous states and the District of Columbia (DC). We used epidemiologic week defined by the National Notifiable Disease Surveillance System (NNDSS) for the purpose of reporting weekly disease incidence. We excluded Alaska and Hawaii in this analysis because these two states do not share spatial adjacency with other states. We adjusted for possible delays in reporting following the approach described elsewhere (Iuliano et al., 2021). Weekly state-specific SARS-CoV-2 surveillance data (i.e., number of tests and number of positive results) were obtained from multiple surveillance systems including the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE), Public Health Laboratory Interoperability Project (PHLIP), and Public Health Laboratory Information System 2 (PHLIS2) (Thompson et al., 2010) from March 8, 2020 to April 25, 2021. Two extra epidemiologic weeks of surveillance data were needed for SARS-CoV-2 testing to include the lagged percent positive in estimating unrecognized COVID-19 deaths. Census population data for 2019 were obtained from the United States Census Bureau (U.S. Census Bureau, 2019). In total, 1,134,371 all-cause deaths were extracted among those $\geq 85$ years of age, of which 162,086 (14.2%) were coded as deaths attributed to COVID-19 (ICD-10 U07.1). For SARS-CoV-2 surveillance data, a total of 2,286,472 tests were identified, of which 124,600 (5.4%) were SARS-CoV-2 positive. Summary statistics of state-level surveillance and mortality data are presented in Supplementary Table S1.

3. Methods

3.1. Spatial–temporal binomial models for positive tests

Let $N_{st}$ denote the number of positive SARS-CoV-2 test results for state $s$ during week $t$ with $s = 1, \ldots, S$ and $t = 1, \ldots, T$. We assume $N_{st} \sim \text{Binom}(D_{st}, p_{st})$, where $p_{st}$ is the percent of positive SARS-CoV-2 test results in decimal form and $D_{st}$ is the number of tests conducted. In subsequent excess mortality analyses, estimates of $p_{st}$ are used to reflect spatial–temporal variations in infection activity. We model $p_{st}$ on the logistic scale:

$$\log\left(\frac{p_{st}}{1-p_{st}}\right) = \alpha_0 + \theta(s, t),$$

where $\alpha_0$ is the intercept and $\theta(s, t)$ denotes a mean-zero spatial–temporal Gaussian process.

We consider three forms for $\theta(s, t)$. First, we assume an additive model

$$\theta(s, t) = v_s + \gamma_t,$$

where $v_s \overset{iid}{\sim} N(0, \sigma^2)$ denotes exchangeable state-specific random effects and $\gamma_t$ denotes week-specific random effects. Temporal dependence in $\gamma_t$ is modeled via a 1-dimensional proper conditional autoregressive model (CAR), $\gamma_t = (\gamma_1, \ldots, \gamma_T)^T \sim \text{CAR}(\rho_{\gamma}, \tau_{\gamma}^2)$. Specifically, the joint distribution of $\gamma$ is mean-zero multivariate Normal with covariance matrix given by $\tau_{\gamma}^2 (D_{\gamma} - \rho_{\gamma} W_{\gamma})^{-1}$, where $W_{\gamma}$ is a $T \times T$ symmetric matrix with the $(t, t')$ entry equal to 1 if week $t$ and $t'$ are adjacent, and 0 otherwise; and $D_{\gamma}$ is a $T \times T$ diagonal matrix with $(D_{\gamma})_{tt} = \sum_{i=1}^{T} (W_{\gamma})_{it}$. This is also known as a random walk model of order 1 and parameter $\rho_{\gamma}$ controls the degree of temporal dependence between weeks.

Second, we extend the above additive model by allowing state-specific random effects to also follow a CAR model, $v = (v_1, \ldots, v_S)^T \sim \text{CAR}(\rho_v, \tau_{v}^2)$. To induce spatial dependency in neighboring
Y. Zhang, H.H. Chang, A.D. Iuliano et al.  
Spatial Statistics 50 (2022) 100584

states, we set entries of the adjacency matrix \( W \) to be 1 if two states share a common boundary. Third, we consider a non-additive spatial–temporal process using a dynamic model for time-series of spatial processes (Banerjee et al., 2003):

\[
\theta(s, t) = \alpha \theta(s, t - 1) + \epsilon_t(s),
\]

where \( \alpha \) is the temporal autoregressive parameter and \( \epsilon_t = (\epsilon_{t1}, \ldots, \epsilon_{tS})^T \) are week-specific spatial residuals with \( \epsilon_t \sim \text{CAR}(\rho, \tau_2^2) \), which we model similarly using the same between-state adjacency matrix. Hereafter, we refer to the above three specifications of the spatial–temporal process \( \theta(s, t) \) as (A) exchangeable, (B) spatial, and (C) dynamic.

3.2. Spatial–temporal negative-binomial models for mortality

To estimate unrecognized COVID-19 deaths, our mortality models are developed using only the subset of all-cause deaths that are not classified on death certificates as being caused by COVID-19. Let \( Y_{st} \) denote the number of all-cause deaths with COVID-19 coded deaths removed for state \( s \) during week \( t \). Negative-binomial (NB) models are used to account for overdispersion often observed in death counts. To facilitate estimation and interpretation, we adapt the parametrization used in Pillow and Scott (2012) and Neelon (2019). Specifically, the NB model is parametrized as a hierarchical Poisson model and has the following form:

\[
Y_{st} | \lambda_{st} \sim \text{Poisson}(\lambda_{st}) \quad \lambda_{st} | \xi, \eta_{st} \sim \text{Gamma}(\xi, \exp(\eta_{st})), \tag{2}
\]

where \( \xi \) is the overdispersion parameter. The corresponding expectation and variance of \( Y_{st} \) are given by \( \mu_{st} = \xi \exp(\eta_{st}) \) and \( \mu_{st} + \frac{1}{\xi} \mu_{st}^2 \), respectively. Spatial–temporal variations \( Y_{st} \) are described by the log-linear mean model

\[
\log(\eta_{st}) = \log(\text{Pop}_{st}) + \beta_0 + \beta_1 p_{s,t-1} + \beta_2 p_{s,t-2} + \delta(s, t), \tag{3}
\]

where \( \text{Pop}_{st} \) is the at-risk population size and enters the model as an offset, \( \beta_0 \) is the intercept, \( p_{s,t-1} \) and \( p_{s,t-2} \) are 1-week and 2-week lagged percent of positive SARS-CoV-2 test results, and \( \delta(s, t) \) is a mean-zero spatial–temporal process. Similar to the spatial–temporal binomial test-positive models in Section 3.1, we consider three forms of \( \delta(s, t) \): (A) space–time additive with exchangeable state-specific random effects; (B) space–time additive with spatial state-specific CAR random effects; (C) dynamic spatial model.

Based on the excess mortality approach, the weekly number of unrecognized deaths attributed to SARS-CoV-2 transmission is defined as the difference between (1) expected deaths with observed SARS-CoV-2 circulation, and (2) counterfactual expected deaths assuming there is no SARS-CoV-2 circulation (i.e., setting \( p_{s,t-1} \) and \( p_{s,t-2} \) to 0). Hence for state \( s \) during week \( t \), the number of unrecognized deaths is defined as

\[
\xi \left( e^{\text{Pop}_{st}} - e^{\log(\text{Pop}_{st}) + \beta_0 + \delta(s, t)} \right). \tag{4}
\]

We note that the above quantity represents estimated COVID-19 deaths only among all-cause death that were not classified on death certificates as being caused by COVID-19. The total number of COVID-19 deaths can be obtained by adding estimated unrecognized COVID-19 deaths and COVID-19 reported deaths from death certificates.

3.3. Estimation and inference

We employed a two-stage Bayesian estimation procedure. First, percent positive \( p_{st} \) was estimated from the spatial–temporal binomial test-positive model, and these estimates were used as covariates in the second-stage NB mortality model. To account for uncertainties in \( p_{st} \) from the first-stage estimation, using multiple realizations of \( p_{st} \) from their joint posterior distribution, we fitted separate NB mortality model and combined posterior samples of parameters. This approach is similar to conducting Bayesian analysis with multiple imputation advocated by Gelman et al. (1995). Parameter estimation in the spatial–temporal models was carried out via Markov Chain Monte Carlo
**Table 1**

| Model        | Exchangeable\(^a\) | Spatial\(^b\) | Dynamic\(^c\) |
|--------------|---------------------|---------------|---------------|
| Binomial     | 59699               | 59619         | 18271         |
| NB model     |                     |               |               |
| imputation 1 | 26015               | 26040         | 27943         |
| imputation 2 | 25949               | 25995         | 27873         |
| imputation 3 | 25801               | 25899         | 27719         |
| imputation 4 | 25887               | 25877         | 27427         |
| imputation 5 | 25665               | 25577         | 27957         |

\(^a\)Space–time additive with exchangeable state-specific random effects and week-specific proper conditional autoregressive model (CAR) random effects.

\(^b\)Space–time additive with state-specific CAR random effects and week-specific CAR random effects.

\(^c\)Dynamic spatial model.

(MCMC) using Gibbs sampling and Metropolis–Hastings (MH) algorithms. Particularly, the high-dimensional spatial and temporal random effects were handled by efficient Gibbs algorithms using the Polya-Gamma (PG) representation of binomial and NB models (Pillow and Scott, 2012; Polson et al., 2013).

For the binomial test-positive models described in Section 3.1, all model parameters can be estimated using Gibbs sampling after specifying appropriate prior distributions and introducing PG latent random variables \(\omega_{1, st} \sim \text{PG}(D_{st}, \alpha_0 + \theta(s, t))\). Similarly, for the NB mortality model, we used latent variables of the form \(\omega_{2, st} \sim \text{PG}(y_{st} + \xi, \eta_{st})\). The following prior distributions were used. Normal \((0, 100)\) prior was assigned for all fixed effect regression coefficients \((\alpha_0, \beta_0, \beta_1, \beta_2)\); inverse-Gamma \((0.1, 0.1)\) priors were assigned for all variance parameters \((\sigma^2, \tau^2, \epsilon^2, \nu^2, \gamma^2)\); and discrete priors (i.e., 1000 equally spaced values between 0 and 1) were assigned for parameters \(\rho_e, \rho_v\), and \(\rho_p\) in CAR models. Details of the MCMC algorithm for fitting spatial–temporal binomial models and NB models are given in the Supplementary Materials.

We used the widely available information criterion (WAIC) to select the form of the spatial–temporal random effects in the test-positive model and the mortality model. WAIC is a measure of predictive accuracy where lower WAIC value is preferred. We follow the definition presented in Banerjee et al. (2003) where WAIC is defined on the deviance scale. Multiple realizations of \(p_{st}\) from the best test-positive model were used for selecting the best mortality model. All analyses were conducted using R 4.0.2 (R Core Team, 2021).

### 4. Results

WAIC from the fitted test-positive and mortality models with different forms of spatial–temporal processes are given in Table 1. Five sets of percent positive were imputed from their posterior distributions to fit the mortality model. We found that five imputations are sufficient to quantify related uncertainties. Specifically, in sensitivity analyses, increasing the number of imputations to 100 yielded similar 95% credible intervals (CIs) of predicted weekly deaths (≈ 0.1% difference in CI endpoints). For the test-positive model, the dynamic spatial–temporal random process which yielded the lowest WAIC is clearly preferred. This may reflect the importance of allowing the timings of when the epidemic started and peaked to vary across states. In contrast, for modeling death counts, the additive spatial and temporal random effects are preferred over the dynamic model, indicating that the temporal trend in baseline mortality rates is similar across states in the absence of SARS-CoV-2. We report COVID-19 death estimates from the spatial random effect specification (i.e., the Spatial model in Table 1) of the NB mortality model since three out of five imputed datasets yielded lower WAIC under this model.

Estimates of parameters included in the selected binomial test-positive and NB mortality models are presented in Tables S4 and S5. The temporal autoregressive parameter \(\alpha\) in the test-positive
model was estimated to be 0.94 (95% CI: 0.93, 0.95), which indicates that the temporal dependence of the spatial process is strong. For the NB mortality model, we observed positive estimates of parameters associated percent positive at 1-week and 2-week lags (0.56 and 0.22), these estimates suggest that weekly all-cause deaths with COVID-19 deaths removed increased as percent positive increased. Two parameters included in CAR models controlling temporal and spatial dependence were estimated to be 0.99, which indicates that observed mortality data exhibited temporal and spatial dependence. We also note that the estimated overdispersion parameter $\xi$ is large (87.51 95% CI: 54.05, 148.06), implying small over-dispersion. This is likely because the mortality model contains week and state-specific random effects that explained residuals well.

Fig. 1 displays the estimated percent positive from the dynamic test-positive model compared to a naive approach (i.e., maximum likelihood estimator [MLE]) for six selected states: Florida, Montana, Nebraska, New York, Oregon, and South Dakota. These six states were chosen to highlight between-state differences in the number of epidemic peaks and their timings, as well as model performance in states with different number of tests. In states with a large number of SARS-CoV-2 tests, estimates of percent positive from the two methods are nearly identical. For example, New York and Florida had, respectively, 136,324 and 80,007 tests conducted during the study time period (Fig. 1). However, for states where small number of tests were conducted (e.g., 4121 total tests in Nebraska) the dynamic test-positive model yielded smoother estimates along with narrower 95% CIs as compared to the naive approach. During early phases of the outbreak, the dynamic model
can also provide percent positive estimates in states when there were no reported tests (e.g., South Dakota). The estimated state-specific weekly percent positive of all 48 contiguous states and DC are provided in Supplementary Materials.

The total number of unrecognized deaths attributed to COVID-19 during study period among populations aged ≥ 85 years was estimated to be 58,200 (95% CI: 51,300, 64,900). This represents 26% of estimated total deaths attributable to COVID-19. The estimated COVID-19 death rate per 1000 population ranged from 18.0 to 44.4 across Health & Human Services (HHS) regions. The highest COVID-19 death rate (44.4, 95% CI: 43.4, 45.4) was observed in HHS Regions 2 (New Jersey and New York). As indicated in Table 2, although HHS Region 10 (Idaho, Oregon, and Washington) had the lowest COVID-19 death rate (18.0, 95% CI: 17.2, 18.8), this HHS Region suffered from the most under-counting; the estimated proportion of unrecognized COVID-19 deaths among all COVID-19 deaths is 0.35 (95% CI: 0.32, 0.38).

Temporal trends of reported and predicted all-cause deaths with COVID-19 coded deaths removed on the national and the state-level are presented in Fig. 2(a) and Fig. 3, respectively. Although there is considerable variation in weekly counts at the state-level, our mortality model captures the temporal trend of reported death counts on the national level very well. The baseline mortality trend assuming there is no SARS-CoV-2 circulation (shown in green) can also be interpreted as the weekly expected death count that is not explained by variations in percent positive. In Fig. 2(a), as indicated by the vertical distance between predicted mortality counts (shown in red) and estimated baseline counts, there are a large number of unrecognized COVID-19 deaths in the first four weeks of the study period, followed by a decrease on the national level. This is further illustrated by examining the proportion of unrecognized COVID-19 deaths among all COVID-19 deaths, which is a measure of the degree of under-counting as shown in Fig. 2(b). This decrease may be related to the increasing availability of diagnostic testing for SARS-CoV-2. In Fig. 2(b), we also see that the proportion of unrecognized deaths is smaller when the total number of COVID-19 deaths is higher. Since the

Table 2
Estimates of unrecognized COVID-19 deaths, COVID-19 death rate per 1000 population, and proportion of unrecognized COVID-19 deaths among estimated total COVID-19 deaths from March 22, 2020 to April 25, 2021 among populations aged ≥ 85 years.

| HHS Regiond | Reported COVID-19 Deaths among ≥ 85 years of agea | Estimated Number of Unrecognized COVID-19 Deaths among ≥ 85 years of age (95% CI)b | Estimated COVID-19 Death Rate among ≥ 85 years of age per 1000 population (95% CI)c | Estimated Proportion of Unrecognized COVID-19 Deaths among ≥ 85 years of age (95% CI)d |
|-------------|---------------------------------------------|--------------------------------------------------------------------------------|---------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| 1           | 11430                                      | 2900 (2600, 3300)                                                              | 40.5 (39.5, 41.5)                                                               | 0.20 (0.18, 0.22)                                                             |
| 2           | 23644                                      | 5800 (5100, 6500)                                                              | 44.4 (43.4, 45.5)                                                               | 0.20 (0.18, 0.22)                                                             |
| 3           | 18419                                      | 6600 (5800, 7400)                                                              | 36.9 (35.7, 38.1)                                                               | 0.26 (0.24, 0.29)                                                             |
| 4           | 25956                                      | 12800 (11400, 14400)                                                           | 27.9 (26.8, 28.9)                                                               | 0.33 (0.30, 0.36)                                                             |
| 5           | 29932                                      | 9600 (8500, 10700)                                                             | 35.5 (34.4, 36.5)                                                               | 0.24 (0.22, 0.26)                                                             |
| 6           | 16605                                      | 7600 (6600, 8500)                                                              | 35.8 (34.4, 37.1)                                                               | 0.31 (0.29, 0.34)                                                             |
| 7           | 8361                                       | 3100 (2700, 3500)                                                              | 35.8 (34.7, 37.0)                                                               | 0.27 (0.25, 0.29)                                                             |
| 8           | 4689                                       | 1100 (900, 1200)                                                               | 28.2 (27.6, 28.9)                                                               | 0.19 (0.17, 0.21)                                                             |
| 9           | 20127                                      | 7000 (6200, 7900)                                                              | 28.4 (27.5, 29.3)                                                               | 0.26 (0.23, 0.28)                                                             |
| 10          | 2923                                       | 1600 (1400, 1800)                                                              | 18.0 (17.2, 18.8)                                                               | 0.35 (0.32, 0.38)                                                             |
| Total       | 162085                                     | 58200 (51300, 64900)                                                           | 33.3 (32.3, 34.3)                                                               | 0.26 (0.24, 0.29)                                                             |

aReported deaths coded as COVID-19 from the National Vital Statistics System (NVSS).

bAll values were rounded to their nearest hundred.

cThe proportion of estimated unrecognized COVID-19 deaths among estimated total COVID-19 deaths.

dHHS Region 1 — CT, MA, ME, NH, RI, VT; HHS Region 2 — NJ, NY; HHS Region 3 — DC, DE, MD, PA, VA, WV; HHS Region 4 — AL, FL, GA, KY, MS, NC, SC, TN; HHS Region 5 — IL, IN, MI, MN, OH, WI; HHS Region 6 — AR, LA, NM, OK, TX; HHS Region 7 — IA, KS, MO, NE; HHS Region 8 — CO, MT, ND, SD, UT, WY; HHS Region 9 — AZ, CA, NV; HHS Region 10 — ID, OR, WA.

eAcross 48 contiguous states and DC.
Fig. 2. Time-series plot of (a) observed all-cause deaths with COVID-19 deaths removed (blue line), predicted all-cause deaths with COVID-19 deaths removed obtained from the selected negative-binomial mortality model (red line), and estimated expected deaths assuming there is no SARS-CoV-2 circulation (green line); (b) reported COVID-19 death rate per 1000 population (gray line), estimated COVID-19 death rate per 1000 population (black line), and estimated proportion of unrecognized COVID-19 deaths (blue line) across 48 contiguous states and DC from March 22, 2020 to April 25, 2021 among population aged ≥ 85 years. Colored areas represent the corresponding 95% CIs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The estimated COVID-19 death rate presented in Fig. 2(b) shows a clear seasonal pattern, two distinct peaks appeared in the end of April 2020 and the beginning of January 2021. Temporal patterns of COVID-19 death rate and proportion of unrecognized deaths among estimated deaths attributable to COVID-19 varied by state as shown in Fig. 4. For example, we estimated a spike appearing around November 2020 in Montana, South Dakota, and Nebraska for COVID-19 death rate. However, a spike was not observed in New York. We also found that in New York and South Dakota, the proportion of unrecognized deaths among estimated deaths attributable to COVID-19 is stable and relatively low (around 0.10) after October 2020. Another example is the sudden increase in Nebraska in April 2021 that was not observed in other states. Temporal trends in COVID-19 death rate and proportion of unrecognized deaths for individual states are available in the Supplementary Materials.

Spatial heterogeneity of the COVID-19 death rate and proportion of unrecognized COVID-19 deaths is evident in Fig. 5. During the study period, six states located in the northeast U.S. (Connecticut, Massachusetts, New Jersey, New York, Pennsylvania, and Rhode Island) had notably high estimated COVID-19 death rates, ranging from 42.6 to 50.4 per 1000 population, and among which Rhode Island had the highest COVID-19 death rate (50.4, 95% CI: 49.4, 51.5). High estimated COVID-19 death rates were also observed in two midwestern adjacent states (South Dakota and Nebraska). We estimated that states varied substantially in the degree of under-counting of deaths attributable to COVID-19 (Fig. 5(b)). For example, Oregon had the highest proportion of unrecognized COVID-19 deaths (0.46, 95% CI: 0.42, 0.50).

5. Discussion

In this paper, we describe a two-stage spatial-temporal modeling framework to estimate unrecognized deaths attributable to COVID-19. To model all-cause deaths not coded as COVID-19, we adapted a negative-binomial model that accounts for infection activity by including estimated
percent positive at 1-week and 2-week lags. The spatial–temporal binomial model improves the estimation of percent positive especially when there were few or no reported tests in the early phase of outbreak. Our modeling framework aims to incorporate the statistical uncertainty resulting from the first-stage estimation, allows for overdispersion in mortality counts, accounts for spatial–temporal dependence, and facilitates the inference of a variety of quantities of interest (e.g., overall COVID-19 death rate, proportion of unrecognized COVID-19 deaths, and baseline deaths assuming no circulation of COVID-19). Our modeling framework is also widely applicable for estimating the disease burden of other infectious diseases that suffer from under-counting issues (e.g., influenza and pneumonia).

We also fitted the Poisson log-linear random-intercept model used in Iuliano et al. (2021) to the current mortality data using estimates from the binomial test-positive model as covariates. During March 2020 to April 2021, the estimated total number of unrecognized deaths among those aged ≥ 85 years was 62,700 (95% CI: 58,600, 66,900). While the point estimate is comparable to our estimates of 58,200 (95% CI: 51,300, 64,900), our approach results in larger posterior interval width (13,600 versus 8,300). This is likely due to the incorporation of additional sources of uncertainty, such as estimation errors associated with the percent positive and spatial–temporal random effects. Also, Iuliano et al. (2021) assume a common temporal trend modeled using natural cubic splines with 5 degrees of freedom, while our approach models temporal trend more flexibly with a first-order random walk. However, for state-specific estimates, we found that our model sometimes
provides smaller interval widths when the number of COVID-19 deaths is small, likely due to the use of random effects to borrow information across locations.

By introducing latent Polya-Gamma variables and leveraging appealing properties of PG distributions, the computational cost associated with fully Bayesian inference for spatial–temporal binomial and negative-binomial models is greatly reduced. This will likely allow for subsequent analyses at finer spatial resolution (e.g., county-level) and with longer study period. Our analysis also only focused on those aged ≥ 85 years, a population with high burden of COVID-19, which may be due to the high prevalence of comorbid conditions such as diabetes, cardiovascular diseases, and renal impairment. Analysis among younger age groups will involve modeling SARS-CoV-2 percent positive with smaller sample sizes and lower death counts, where leveraging spatial–temporal dependence may be particularly helpful. Our experience in modeling death counts in states with small population (Figs. 3 and 4) also highlight the importance of quantifying uncertainties in estimates. On the other hand, the baseline mortality trends in younger age groups may exhibit weaker seasonality because of the lower rates of chronic diseases. This which may facilitate the statistical model to estimate residual temporal variation attributable to COVID-19 activity. Furthermore, in assessing the society and economic impacts of COVID-19, deaths in younger age groups are associated with larger years of potential life loss (YPLL), a commonly used for assessing premature deaths; though assumptions and uncertainties in values assigned to each age at death also play an important role in these calculations.
Fig. 5. A map of (a) estimated COVID-19 death rate per 1000 population; (b) estimated proportion of unrecognized COVID-19 deaths across 48 contiguous states with state borders from March 22, 2020 to April 25, 2021 among population aged ≥ 85 years.

There are several limitations of this analysis that should be acknowledged. First, the spatial–temporal binomial regression model for positive SARS-CoV-2 testing results only contained an intercept. Additional factors may be useful for predicting percent positive of SARS-CoV-2 test results among subpopulations, including sex, gender, race/ethnicity, medical condition, timing of
vaccination, circulation of the Delta variant, and use of boosters (de Lusignan et al., 2020; Fan et al., 2020). The percent positive can then be used to estimate COVID-19 disease burden among specific high-risk subpopulations. Second, we assumed a constant autoregressive parameter \( \alpha \) in the binomial regression. The model with state-specific \( \alpha \) may be worth further investigations when spatially-varying temporal dependence is considered. Third, we assumed the associations between weekly mortality and 1-week and 2-week lagged percent positive to be constant across space and time. We explored allowing for month or state-specific regression coefficients to potentially capture differences due to timing of the pandemic or spatial heterogeneity, but found no improvement in model fit as measured by WAIC under both models. It is also possible to consider more parsimonious spatially-varying coefficient models to allow state-specific effects of percent positive. Fourth, we found estimated state-specific mortality trends may not capture peaks in some states (Fig. 3). This may be due to the use of a common temporal trends across states, even though WAIC does not suggest the dynamic spatial–temporal random effect model is better. Other flexible methods to capture space–time interaction, such as temporal splines with spatially-varying coefficients, warrant further investigations.

We estimated that 26\% of deaths attributable to COVID-19 were unrecognized among those aged \( \geq 85 \) years during the pandemic up to April 2021. We note that estimating attributable deaths using an excess death approach may capture indirect deaths associated with the pandemic, such as those related to mental health, if the temporal patterns of these indirect causes are correlated with percent positive. Understanding factors that contributed to regional and temporal differences in under-counting can provide important knowledge in improving and designing disease burden surveillance systems. For example, the observed temporal decreases in proportion of unrecognized COVID-19 deaths may be due to increase in diagnostic capacity and experience. This is particularly important in the future when widespread access to testing is not sustainable.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention (CDC). The authors thank Neha Patel, Jeremy Reich, and Krista Kniss for their help in compiling the mortality and surveillance data. This work was supported in part by the National Institutes of Health, USA Award R01AI125842. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funder.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.spasta.2021.100584.

References

Banerjee, S., Carlin, B.P., Gelfand, A.E., 2003. Hierarchical Modeling and Analysis for Spatial Data. Chapman and Hall/CRC.
Centers for Disease Control and Prevention, 2021. Coronavirus disease 2019 (COVID-19). https://www.cdc.gov/coronavirus/2019-ncov/index.html (Accessed 23 November 2021).
Cohen, C., Simonsen, L., Kang, J.-W., Miller, M., McNerney, J., Blumberg, L., Schoub, B., Madhi, S.A., Viboud, C., 2010. Elevated influenza-related excess mortality in South African elderly individuals, 1998–2005. Clin. Infect. Dis. 51 (12), 1362–1369.
de Lusignan, S., Dorward, J., Correa, A., Jones, N., Akinyemi, O., Amirthalingam, G., Andrews, N., Byford, R., Dabrera, G., Elliot, A., et al., 2020. Risk factors for SARS-CoV-2 among patients in the oxford royal college of general practitioners research and surveillance centre primary care network: a cross-sectional study. Lancet Infect. Diseases 20 (9), 1034–1042.
Fan, V.S., Dominitz, J.A., Eastment, M., Locke, E., Green, P., Berry, K., O’Hare, A.M., Shah, J.A., Crothers, K., Ioannou, G.N., 2020. Risk factors for testing positive for SARS-CoV-2 in a national US healthcare system. Clin. Infect. Dis.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. Bayesian Data Analysis. Chapman and Hall/CRC.

Iuliano, A.D., Chang, H.H., Patel, N.N., Threlkel, R., Kniss, K., Reich, J., Steele, M., Hall, A.J., Fry, A.M., Reed, C., 2021. Estimating under-recognized covid-19 deaths, united states, March 2020-May 2021 using an excess mortality modelling approach. Lancet Reg. Health-America. 1, 100019.

Lin, J.C., Nichol, K.L., 2001. Excess mortality due to pneumonia or influenza during influenza seasons among persons with acquired immunodeficiency syndrome. Arch. Internal Med. 161 (3), 441–446.

Neal, K.R., 2007. Excess mortality rates in a cohort of patients infected with the hepatitis C virus: a prospective study. Gut 56 (8), 1098–1104.

Neelon, B., 2019. Bayesian zero-inflated negative binomial regression based on pòlya-gamma mixtures. Bayesian Anal. 14 (3), 829.

Pillow, J., Scott, J., 2012. Fully Bayesian inference for neural models with negative-binomial spiking. Adv. Neural Inf. Process. Syst. 25, 1898–1906.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

Rivera, R., Rosenbaum, J.E., Quispe, W., 2020. Excess mortality in the United States during the first three months of the COVID-19 pandemic. Epidemiol. Infect. 148.

Rossen, L.M., Branum, A.M., Ahmad, F.B., Sutton, P., Anderson, R.N., 2020. Excess deaths associated with COVID-19, by age and race and ethnicity—United States, January 26–October 3, 2020. Morb. Mortal. Wkly Rep. 69 (42), 1522.

Tatar, M., Habibdoust, A., Wilson, F.A., 2021. Analysis of excess deaths during the COVID-19 pandemic in the state of florida. Am J Public Health 111 (4), 704–707.

Thompson, M., Shay, D., Zhou, H., Bridges, C., Cheng, P., Burns, E., Breshe, J., Cox, N., et al., 2010. Estimates of deaths associated with seasonal influenza-United States, 1976–2007. Morb. Mortal. Wkly Rep. 59 (33), 1057–1062.

U.S. Census Bureau, 2019. State population by characteristics: 2010–2019. https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html (Accessed 10 August 2020).

Vandoros, S., 2020. Excess mortality during the Covid-19 pandemic: Early evidence from England and Wales. Soc. Sci. Med. 258, 113101.

Verbeke, J., Faes, C., Neyens, T., Hens, N., Verbeke, G., Deboosere, P., Molenberghs, G., 2021. A linear mixed model to estimate COVID-19-induced excess mortality. MedRxiv.

Weinberger, D.M., Chen, J., Cohen, T., Crawford, F.W., Mostashari, F., Olson, D., Pitzer, V.E., Reich, N.G., Russi, M., Simonsen, L., et al., 2020. Estimation of excess deaths associated with the COVID-19 pandemic in the United States, March to May 2020. Jama Int. Med. 180 (10), 1336–1344.

Woolf, S.H., Chapman, D.A., Sabo, R.T., Weinberger, D.M., Hill, L., 2020. Excess deaths from COVID-19 and other causes, March-April 2020. JAMA 324 (5), 510–513.

Woolf, S.H., Chapman, D.A., Sabo, R.T., Zimmerman, E.B., 2021. Excess deaths from COVID-19 and other causes in the US, March 1, 2020, to January 2, 2021. JAMA 325 (17), 1786–1789.