Infectious Disease Modelling 9 (2024) 70–83

Contents lists available at ScienceDirect

Infectious Disease Modelling

journal homepage: www.keaipublishing.com/idm

Reconstruction of incidence reporting rate for SARS-CoV-2 Delta variant of COVID-19 pandemic in the US☆

Alexandra Smirnova *, Mona Baroonian

Department of Mathematics & Statistics, Georgia State University, Atlanta, USA

ARTICLE INFO

Article history:
Received 30 September 2023
Received in revised form 3 December 2023
Accepted 3 December 2023
Available online 9 December 2023
Handling Editor: Dr Daihai He

Keywords:
Epidemiology
Incidence reporting rate
Compartmental model
Transmission dynamic
Regularization algorithm

ABSTRACT

In recent years, advanced regularization techniques have emerged as a powerful tool aimed at stable estimation of infectious disease parameters that are crucial for future projections, prevention, and control. Unlike other system parameters, i.e., incubation and recovery rates, the case reporting rate, Ψ, and the time-dependent effective reproduction number, $\mathcal{R}_e(t)$, are directly influenced by a large number of factors making it impossible to pre-estimate these parameters in any meaningful way. In this study, we propose a novel iteratively-regularized trust-region optimization algorithm, combined with $S_uS_vI_uI_vRD$ compartmental model, for stable reconstruction of Ψ and $\mathcal{R}_e(t)$ from reported epidemic data on vaccination percentages, incidence cases, and daily deaths. The innovative regularization procedure exploits (and takes full advantage of) a unique structure of the Jacobian and Hessian approximation for the nonlinear observation operator. The proposed inversion method is thoroughly tested with synthetic and real SARS-CoV-2 Delta variant data for different regions in the United States of America from July 9, 2021, to November 25, 2021. Our study shows that case reporting rate during the Delta wave of COVID-19 pandemic in the US is between 12% and 37%, with most states being in the range from 15% to 25%. This confirms earlier accounts on considerable under-reporting of COVID-19 cases due to the impact of "silent spreaders" and the limitations of testing.

1. Introduction

Compartmental models have been widely used to analyze complex biological systems with multiple interacting sections, particularly those involving the spread of infectious diseases (Fenichel et al., 2011). In these models of disease transmission, a well-mixed population is segregated into distinct groups or compartments in accordance with their particular disease status, such as susceptibility, infection, and recovery. The transmission dynamics of the disease within and between these compartments are then modeled through the implementation of mathematical equations thus helping scholars and decision-makers to mitigate and control emerging epidemics. Over the years, compartmental models have been extensively utilized to study the transmission of infectious diseases such as HIV, influenza, cholera, COVID-19, and others (He, Peng, & Sun, 2020;

☆ Supported by NSF award 2011622 (DMS Computational Mathematics).
* Corresponding author.
E-mail addresses: asmirnova@gsu.edu (A. Smirnova), mbaroonian1@student.gsu.edu (M. Baroonian).
Peer review under responsibility of KeAi Communications Co., Ltd.

https://doi.org/10.1016/j.idm.2023.12.001

Kudryashov, Chmykhov, & Vigdorowitsch, 2021; Smirnova et al., 2020, 2022). Their ability to quantify the diffusion of diseases and the related intricate systems was instrumental in guiding high-stakes recommendations on public health policies and interventions.

When COVID-19 spread all over the world (Li et al., 2020), compartmental models came to light as a valuable tool in forecasting of future incidence cases, especially under diverse circumstances involving immunization campaigns, social distancing measures, and the emergence of novel virus strains (Patalon et al., 2022; Smirnova et al., 2022). At the onset of COVID-19, when little was know about the disease transmission, a simple *SI* compartmental model (Kosmidis & Macheras, 2020) was proposed to explain the dynamics of the virus. It assumed that individuals can only be susceptible (S) or infectious (I), and did not account for recovery or immunity. While this model was appropriate for investigating the primary spread of an infectious ailment, it was not sufficient for emulating the prolonged dynamics of the outbreak.

As COVID-19 evolved, *SIS* (Michael Otunuga, 2021) and *SIR* (Kudryashov et al., 2021) models were employed to study the transmission of the disease. The *SIS* is a simpler version of the *SIR* model, where individuals can be infected multiple times and not develop immunity after recovery. In (Michael Otunuga, 2021), the time-dependent probability distribution for the number of infected individuals was introduced, and the effects of noise intensity on changes in transmission and recovery rates were investigated. In (Betti, Luigi Bragazzi, Heffernan, Kong, & Raad, 2021), a modified epidemiological compartmental *SIR* model, aimed at the integration of non-pharmaceutical (physical/social distancing) and pharmaceutical (immunization) public health control measures, was proposed and fitted to the cumulative COVID-19 data for the province of Ontario, Canada, from September to December, 2020. Different vaccine roll-out strategies were simulated until 75% of the population was vaccinated, including a no-vaccination scenario.

More complex COVID-19 models, such as the *SEIR* model (He et al., 2020; Tang et al., 2020), included additional compartments to account for the latent period of the disease and the possibility of transmission before symptoms appear. In these models, the population has been divided into four compartments. The exposed compartment (E) represented individuals that have been infected but were not yet infectious. The methodology proposed in (Calafiore, Novara, & Possieri, 2020) adopted *SIRD* model and time-varying parameters in order to capture alterations in the epidemic behavior which stemmed from containment measures, antiviral treatments, and other relevant factors. The authors employed sparse identification techniques to deduce time-varying parameters from epidemic data. In *SIDARTHE* model (Giordano, Franco, et al., 2020), the infection progression included 8 different stages, susceptible (S), infected (I), diagnosed (D), ailing (A), recognized (R), threatened (T), healed (H) and extinct (E), which gave rise to a careful study of possible scenarios of epidemic spread under various control strategies.

An important part of research on COVID-19 was the addition of a Quarantine (Q) compartment to the standard *SEIR* model (Sharma, Volpert, & Banerjee, 2020). Individuals in the Quarantine compartment were isolated from the rest of the population following their exposure to COVID-19. The resulting *SEIQR* framework allowed to facilitate the appraisal of distinct quarantine strategies such as obligatory segregation, contact tracing, and self-imposed quarantine. To better forecast healthcare requirements, in (Leontitsis et al., 2021), the Infected compartment was further divided into (1) asymptomatic, (2) isolated and (3) hospitalized, thus giving rise to *SEAHIR* (Susceptible-Exposed-Asymptomatic-Hospitalized-Isolated-Removed) COVID-19 model.

A crucial aspect of COVID-19 transmission is the existence of inconspicuous disseminators (Aguilar, Samuel Faust, Westafer, & Gutierrez, 2020; Mizumoto, Kagaya, Alexander, & Chowell, 2020). In (Moghadas et al., 2020), it has been suggested that unobtrusive transmission, comprised of both presymptomatic and asymptomatic occurrences, may be the primary cause of epidemics. That is, the majority of incidence cases can be attributed to silent transmission resulting from a combination of the presymptomatic phase and asymptomatic infections. Consequently, even if all symptomatic cases are secluded, a widespread epidemic may still transpire. In (Moghadas et al., 2020), the effect of segregating asymptomatic infections in combination with symptomatic instances has been measured and it has been estimated that segregating more than one-third of asymptomatic infections is vital to restrict a future outbreak to under 1% of the population.

In another study (Subramanian, He, & Pascual, 2021), using *SEPIAR* (Susceptible–Exposed–Presymptomatic-Infectious–Asymptomatic-Recovered) model that incorporated daily RT-PCR testing information fit to New York City data at the early stage of COVID-19, the authors concluded that the proportion of symptomatic cases may be as low as 13–18%, bringing the overall reproduction number to $3.2 - 4.4$. For Omicron variant, depending on the population and testing methods in different surveys of Omicron-positive individuals, the estimates for the proportion of asymptomatic infections varied widely (Ma et al., 2021; Oran & Eric, 2021; Shang et al., 2022), ranging from 20% to 80%, with the most reliable approximation suggesting that probably $40 - 45$% of Omicron-infected individuals were asymptomatic.

Given the significance of quantifying the impact of "silent spreaders" for control and prevention, in this paper, we combine $S_uS_vI_uI_vRD$ compartmental model (Luo et al., 2023) with a novel problem-oriented regularized optimization algorithm aimed at stable estimation of reporting rates, Ψ, and time-dependent effective reproduction numbers, $\mathcal{R}_e(t)$, from COVID-19 data on vaccination percentages, incidence cases, and daily deaths. The proposed technique is thoroughly tested with synthetic and real SARS-CoV-2 Delta variant data for different regions in the United States of America from July 9, 2021, to November 25, 2021.

The paper is organized as follows. In Section 2, we introduce $S_uS_vI_uI_vRD$, which stands for Susceptible unvaccinated-Susceptible vaccinated-Infected unvaccinated-Infected vaccinated-Recovered-Deceased, compartmental model (Luo et al., 2023) for COVID-19. In section 3, our innovative nonlinear constrained minimization problem and a new computational method for the reconstruction of Ψ and $\mathcal{R}_e(t)$ are presented together with some unique features of Jacobian and Hessian

approximations for the nonlinear observation operator. Numerical experiments with synthetic and real data for COVID-19 pandemic are discussed in Sections 4 and 5, respectively. The summary of our findings and future research plans are outlined in Section 6.

2. Model for disease transmission

The $S_uS_vI_uI_vRD$ model (Luo et al., 2023), derived in collaboration with Dr. Ruiyan Luo, the Department of Population Health Sciences, GSU School of Public Health, incorporates the development and widespread distribution of COVID-19 vaccines prior to the start of SARS-CoV-2 Delta variant (Agossou, Nicodème Atchadé, & Moussa Djibril, 2021; Angeli, Neofotistos, Mattheakis, & Kaxiras, 2022; Gustavo Barbosa LibotteFran Sérgio Lobato et al., 2020). The model accounts for varying disease dynamics, such as different transmission, recovery and death rates, within vaccinated and unvaccinated individuals by taking into the consideration the vaccination status of both susceptible and infected individuals. Specifically, the model includes susceptible unvaccinated (S_u), susceptible vaccinated (S_v), infected unvaccinated (I_u), infected vaccinated (I_v), recovered (R), and deceased (D) compartments (Luo et al., 2023). In the model, N denotes the population size at the beginning of the study period, that is, $N = S_u(0) + S_v(0) + I_u(0) + I_v(0) + R(0) + D(0)$. We assume that $D(t)$ is the cumulative number of individuals deceased due to COVID-19 from the start of the study period up until time t, and all other changes in the population (due to birth, immigration, death of causes rather than COVID-19, etc.) are balanced out. Thus, $N - D(t)$ stands for the total living population of the region at time t.

$$\frac{\partial S_u}{\partial t} = -\beta(t)\frac{S_u(t)}{N - D(t)}(I_u(t) + I_v(t)) - pS_u(t) + \delta_r R(t) + \delta_v S_v(t) \tag{2.1}$$

$$\frac{\partial S_v}{\partial t} = pS_u(t) - (1-\alpha)\beta(t)\frac{S_v(t)}{N - D(t)}(I_u(t) + I_v(t)) - \delta_v S_v(t) \tag{2.2}$$

$$\frac{\partial I_u}{\partial t} = \beta(t)\frac{S_u(t)}{N - D(t)}(I_u(t) + I_v(t)) - (\gamma_{u,r} + \gamma_{u,d})I_u(t) \tag{2.3}$$

$$\frac{\partial I_v}{\partial t} = (1-\alpha)\beta(t)\frac{S_v(t)}{N - D(t)}(I_u(t) + I_v(t)) - (\gamma_{v,r} + \gamma_{v,d})I_v(t) \tag{2.4}$$

$$\frac{\partial R}{\partial t} = \gamma_{u,r}I_u(t) + \gamma_{v,r}I_v(t) - \delta_r R(t) \tag{2.5}$$

$$\frac{\partial D}{\partial t} = \gamma_{u,d}I_u(t) + \gamma_{v,d}I_v(t) \tag{2.6}$$

The meaning of all variables and parameters in ODE system (2.1)–(2.6) is given in Tables 1 and 2, respectively. Susceptible unvaccinated individuals get vaccinated at a rate p, and become infected at a time dependent transmission rate, $\beta(t)$. It is understood in the model that susceptible vaccinated individuals become infected at a lower rate, $(1-\alpha)\beta(t)$, where $0 < \alpha < 1$, is a measure of vaccine efficacy. The smaller values of α correspond to less efficacy. Following (Johnson, 2022; Luo et al., 2023), we assume that infected unvaccinated humans recover at a lower rate and have a higher risk of hospitalization and death as compared to vaccinated humans. In model (2.1)–(2.6), infected unvaccinated and vaccinated individuals recover at rates $\gamma_{u,r}$ and $\gamma_{v,r}$, and die at rates $\gamma_{u,d}$ and $\gamma_{v,d}$, respectively. The loss of immunity is taken into account by considering the movement back to susceptible unvaccinated class from susceptible vaccinated and recovered classes at rates δ_v and δ_r.

While not explicitly present in the ODE system, the effective reproduction number, $\mathcal{R}_e(t)$, measures the contagiousness or transmissibility of an infectious disease. It represents the average number of secondary infections caused by each infected individual in a population at a specific point in time, in view of various factors such as the disease's inherent transmissibility, population characteristics, and control measures in place. Therefore, stable reconstruction of the effective reproduction number, $\mathcal{R}_e(t)$, and its underlying transmission rate, $\beta(t)$, is of paramount importance (Giordano, Blanchini, et al., 2020; Roosa et al., 2020; Thompson, 2020). Estimating these parameters allows for the real-time analysis of the effectiveness of control and prevention and enables accurate forecasting of future incidence cases (Bjørnstad, Finkenstädt, & Grenfell, 2002; Chowell,

Table 1
Variables for SVIRD compartmental model (2.1)–(2.6).

Variable	Meaning
$S_u(t)$	Number of susceptible unvaccinated individuals
$S_v(t)$	Number of susceptible vaccinated individuals
$I_u(t)$	Number of infectious unvaccinated individuals
$I_v(t)$	Number of infectious vaccinated individuals
$R(t)$	Number of recovered individuals
$D(t)$	Number of deceased individuals

Table 2
Parameter values for Delta variant of COVID-19, July 9 to November 25, 2021.

Parameter	Meaning	Value	Units	Source
$\beta(t)$	Transmission rate			
ψ	Incidence reporting parameter			
α	Vaccine dose efficacy	0.8		(Lewis et al., 2022; Lopez Bernal et al., 2021; Luo et al., 2023)
$\gamma_{u,r}$	Recovery rate of unvaccinated	0.0995	day^{-1}	Luo et al. (2023)
$\gamma_{v,r}$	Recovery rate of vaccinated	0.09996	day^{-1}	Luo et al. (2023)
$\gamma_{u,d}$	Fatality rate for unvaccinated	0.00027	day^{-1}	Luo et al. (2023)
$\gamma_{v,d}$	Fatality rate for vaccinated	0.000021	day^{-1}	Luo et al. (2023)
δ_v	Loss of immunity for vaccinated	0.011	day^{-1}	(Doria-Rose et al., 2021; Luo et al., 2023)
δ_r	Loss of immunity due to COVID-19	0	day^{-1}	Luo et al. (2023)

Viboud, Simonsen, & Moghadas, 2016). Whereas other system parameters, i.e., death and recovery rates, are less dependent on public health policies, the effective reproduction number and the transmission rate of the disease are directly influenced by mitigation measures.

Using the next-generation matrix (Diekmann, Heesterbeek, & Metz, 1990; Tang et al., 2020; van den Driessche & Watmough, 2002), one computes the effective reproduction number, $\mathcal{R}_e(t)$, for compartmental model (2.1)–(2.6) as (Luo et al., 2023)

$$\mathcal{R}_e(t) = \frac{\beta(t)}{N - D(t)} \left(\frac{S_u(t)}{\gamma_{u,r} + \gamma_{u,d}} + \frac{(1-\alpha) S_v(t)}{\gamma_{v,r} + \gamma_{v,d}} \right). \tag{2.7}$$

The function $\mathcal{R}_e(t)$ is a key metric for assessing the spread of infectious diseases and estimating the potential impact of an outbreak. It helps public health officials and researchers to understand how quickly a disease is spreading and whether control measures are effective in reducing the transmission.

When $\mathcal{R}_e(t)$ is greater than 1, it indicates that each infected individual, on average, is transmitting the disease to more than one other person, leading to an exponential growth of the outbreak. Conversely, when $\mathcal{R}_e(t)$ is less than 1, it suggests that the transmission is declining, as each infected person is infecting fewer than one other person, leading to a decrease in the overall number of cases. Monitoring and estimating $\mathcal{R}_e(t)$ is crucial for guiding public health interventions. If $\mathcal{R}_e(t)$ is above 1, it may signal the need for implementing additional control measures such as social distancing, mask mandates, contact tracing, vaccination campaigns, or lockdowns to bring the transmission under control. On the other hand, if $\mathcal{R}_e(t)$ is consistently below 1, it indicates that the disease is being effectively contained, although vigilance is still required to prevent potential resurgences. The effective reproduction number, $\mathcal{R}_e(t)$, is not a fixed value but can change over time as conditions evolve, interventions are implemented, or new variants of the pathogen emerge. Therefore, ongoing monitoring and adaptation of control strategies are necessary to manage the spread of infectious diseases effectively.

3. Nonlinear minimization problem

Due to a considerable number of asymptomatic cases (Ma et al., 2021; Shang et al., 2022), we assume that incidence data, η, is underreported, and the true number of incidence cases is η/Ψ, $0 < \Psi < 1$. Thus, our goal is to optimize the model with respect to the transmission rate, $\beta = \beta(t)$, and the reporting rate, Ψ. To that end, we discretize the transmission rate, $\beta(t)$, by projecting it onto a finite subspace spanned by base functions $P_1(t), P_2(t), \ldots, P_m(t)$. This gives rise to the discrete analog, $\tilde{\beta}[\theta]$, in the form

$$\tilde{\beta}[\theta](t) = \sum_{j=1}^{m} \theta_j P_j(t). \tag{3.1}$$

If one solves ODE system (2.1)–(2.6) with $\beta = \tilde{\beta}[\theta]$, one obtains state variables, $\tilde{S}_u[\theta]$, $\tilde{S}_v[\theta]$, $\tilde{I}_u[\theta]$, $\tilde{I}_v[\theta]$, $\tilde{R}[\theta]$, and $\tilde{D}[\theta]$, as functions of the expansion coefficients, θ, thus reducing the original problem to the unconstrained minimization. It follows from equations (2.1) and (2).2) that the daily number of newly infected people is equal to $\tilde{\beta}[\theta] \frac{(\tilde{I}_u[\theta] + \tilde{I}_v[\theta])}{N - \tilde{D}[\theta]} \left(\tilde{S}_u[\theta] + (1-\alpha)\tilde{S}_v[\theta] \right)$ while, according to (2.6), the daily number of individuals deceased due to COVID-19 is equal to $\gamma_{u,d}\tilde{I}_u[\theta] + \gamma_{v,d}\tilde{I}_v[\theta]$. Therefore, the observation operators for incidence cases, η, and daily new deaths, σ, are defined, respectively, as

$$\Phi[\theta] := \tilde{\beta}[\theta] \frac{(\tilde{I}_u[\theta] + \tilde{I}_v[\theta])}{N - \tilde{D}[\theta]} \left(\tilde{S}_u[\theta] + (1-\alpha)\tilde{S}_v[\theta] \right) \quad \text{and} \quad \Omega[\theta] := \gamma_{u,d}\tilde{I}_u[\theta] + \gamma_{v,d}\tilde{I}_v[\theta]. \tag{3.2}$$

Hence one arrives at the following nonlinear least squares problem (NLSP)

$$\min_{\theta,\psi}\Lambda[\theta,\psi], \quad \text{where } \Lambda[\theta,\psi]=\frac{\lambda_1}{2}\|\Phi[\theta]-\psi\eta\|^2+\frac{\lambda_2}{2}\|\Omega[\theta]-\sigma\|^2, \quad \lambda_1,\lambda_2>0. \tag{3.3}$$

In (3.3), the reporting index, ψ, is the reciprocal of the reporting rate, Ψ. The purpose of λ_1 and λ_2 is to balance the two residuals, since $\psi\eta$ and σ are of different orders of magnitude. By the first order necessary condition (FONC),

$$\frac{\partial}{\partial\psi}\Lambda[\theta,\psi]=0, \quad \text{that is,} \quad -\eta^T\Phi[\theta]+\psi\|\eta\|^2=0 \quad \text{and} \quad \psi=\frac{\eta^T\Phi[\theta]}{\|\eta\|^2}. \tag{3.4}$$

This yields NLSP in the form

$$\min_{\theta}\left\{\frac{\lambda_1}{2}\left|\left|\Phi\left[\theta\right]-\frac{\eta\eta^T\Phi[\theta]}{\|\eta\|^2}\right|\right|^2+\frac{\lambda_2}{2}||\Omega[\theta]-\sigma||^2\right\}. \tag{3.5}$$

Let $\nu:=\frac{\eta}{\|\eta\|}$ be the normalized incidence data. Then we cast our optimization problem as

$$\min_{\theta}\left\{\frac{\lambda_1}{2}\left|\left|\left(I-\nu\nu^T\right)\Phi\left[\theta\right]\right|\right|^2+\frac{\lambda_2}{2}||\Omega[\theta]-\sigma||^2\right\}, \quad \nu=\frac{\eta}{\|\eta\|}. \tag{3.6}$$

In (3.6), I is the identity matrix in $\mathbb{R}^{n\times n}$, where n is the number of data points, and $I-\nu\nu^T$ is an orthogonal projection. By the FONC,

$$\lambda_1\left(\left(I-\nu\nu^T\right)\Phi'[\theta]\right)^T(I-\nu\nu^T)\Phi[\theta]+\lambda_2\Omega'^T[\theta](\Omega[\theta]-\sigma)=0. \tag{3.7}$$

Since $P^2=P=P^T$ for $P:=I-\nu\nu^T$, one has $\lambda_1\Phi'^T[\theta]P\Phi[\theta]+\lambda_2\Omega'^T[\theta](\Omega[\theta]-\sigma)=0$. This yields the following Hessian operator:

$$H[\theta]:=\lambda_1\Phi''^T[\theta]P\Phi[\theta]+\lambda_1\Phi'^T[\theta]P\Phi'[\theta]+\lambda_2\Omega''^T[\theta](\Omega[\theta]-\sigma)+\lambda_2\Omega'^T[\theta]\Omega'[\theta]. \tag{3.8}$$

Upon dropping the terms that contain the second derivative operator applied to a residual that is expected to decrease as iterations advance, one arrives at the Hessian approximation in the form

$$H[\theta]\approx\lambda_1\Phi'^T[\theta]P\Phi'[\theta]+\lambda_2\Omega'^T[\theta]\Omega'[\theta]. \tag{3.9}$$

This Hessian approximation is nonnegative definite. Indeed, for any $h\in\mathbb{R}^n$, one has

$$\lambda_1h^T\Phi'^TP\Phi'h+\lambda_2h^T\Omega'^T\Omega'h=\lambda_1h^T\Phi'^TP^TP\Phi'h+\lambda_2h^T\Omega'^T\Omega'h=\lambda_1(P\Phi'h)^TP\Phi'h+\lambda_2(\Omega'h)^T\Omega'h=\lambda_1\|P\Phi'h\|^2+\lambda_2\|\Omega'h\|^2\geq0.$$

To solve minimization problem (3.6) numerically, we use projected iteratively regularized Gauss-Newton algorithm (Bakushinsky & Kokurin, 2004; Kaltenbacher, Neubauer, & Scherzer, 2008; Smirnova & Bakushinsky, 2020):

$$P:=I-\frac{\eta\eta^T}{\|\eta\|^2}, \quad \Phi_k:=\Phi[\theta_k], \quad \Omega_k:=\Omega[\theta_k], \quad \Phi_k':=\Phi'[\theta_k], \quad \Omega_k':=\Omega'[\theta_k], \tag{3.11}$$

$$\left\{\left(\lambda_1\Phi_k'^TP\Phi_k'+\lambda_2\Omega_k'^T\Omega_k'+\tau_kI\right)z_k=-\left\{\lambda_1\Phi_k'^TP\Phi_k+\lambda_2\Omega_k'^T(\Omega_k-\sigma)+\tau_k(\theta_k-\bar{\theta})\right\} \quad \theta_{k+1}=\theta_k+z_k, \right. \tag{3.12}$$

where $\{\tau_k\}$ is the regularization sequence and $\bar{\theta}$ is the reference value for the unknown parameter, θ. The standard assumptions on the sequence $\{\tau_k\}$,

$$\tau_k\geq\tau_{k+1}>0, \quad \sup_{k=0,1,2,\dots}\sqrt{\frac{\tau_k}{\tau_{k+1}}}<\infty, \quad \text{and} \quad \lim_{k\to\infty}\tau_k=0, \tag{3.13}$$

guarantee that, given the right choice of initial approximations, the regularization is sufficient to ensure stability, but not excessive (Bakushinsky & Kokurin, 2004; Kaltenbacher et al., 2008). Clearly, a closed form solution to ODE system (2.1)–(2.6) is not available. Therefore, system (2.1)–(2.6) needs to be solved numerically for the current value of θ_k at every step of iterative process (3.12).

4. Numerical experiments with synthetic data

In this section we use our proposed methodology for stable estimation of unknown incidence reporting rate, Ψ, and time-dependent effective reproduction number, $\mathcal{R}_e(t)$, from synthetic data sets, η and σ, replicating a seasonal virus outbreak. We start with synthetic data to ensure that our results are reliable in a sense that we can reconstruct the same reporting rate, Ψ, and reproduction number, $\mathcal{R}_e(t)$, that were employed to generate incidence data on daily new cases and deaths before random noise has been added to simulate "real" measurements. Our ability to reconstruct the original disease parameters would indicate that numerical algorithm (3.12) is stable with respect to noise in our input data and that the values of Ψ and $\mathcal{R}_e(t)$ recovered from real data make sense.

The first step of our experiment with synthetic data is to select a "model" transmission rate, $\beta(t)$, and then to use this $\beta(t)$ along with pre-estimated parameters listed in Table 2 for solving forward problem (2.1) - (2.6). Once the forward problem has been solved, one can select a "model" reporting rate, Ψ, and evaluate observation operators (3.2), Φ and Ω, for incidence cases, η, and daily new deaths, σ, in order to get clean data on a given time interval $[a, b]$. Then, to replicate noise contaminated data in a real-life setting, one adds random Gaussian noise with zero mean to the clean epidemic data, as shown in Fig. 1. Since real incidence cases and deaths must be positive, uniform noise is added if data values become negative at any point.

In Fig. 2, one can see the original effective reproduction number, $\mathcal{R}_e(t)$, generated by "model" transmission rate, $\beta(t)$ (solid red curves). This transmission rate, $\beta(t)$, mimics a tangible scenario when mitigation measures are very efficient short-term but not sustainable in a long run. As the result, the epidemic wave gets successfully under control following the downward trend in both $\mathcal{R}_e(t)$ and $\beta(t)$ only to bounce back up in response to the new ascending stage in disease transmission. This kind of behavior for the effective reproduction number, $\mathcal{R}_e(t)$, and its underlying transmission rate, $\beta(t)$, is also possible if the second epidemic wave corresponds to a new strain of virus, which calls for different mitigation measures.

Given noisy synthetic data for incidence cases and daily new deaths, we employ regularized projected Gauss-Newton algorithm (3.12) to reconstruct the unknown expansion coefficients, θ, for the time-dependent transmission rate, $\beta(t)$. Given θ_j, $j = 1, 2, \ldots, m$, one calculates the reporting index, $\psi = 1/\Psi$, and the effective reproduction number, $\mathcal{R}_e(t)$, using identities (3.4) and (2.7), (3.1), respectively.

In order to quantify uncertainty in the extracted ψ and $\mathcal{R}_e(t)$, one refits the model (using parallel programming - *parfor* option in Matlab R2023a with ode23s ODE solver) to $M = 100$ additional data sets for incidence cases and daily deaths assuming Poisson error structure. The resulting M best-fit parameter sets are used to construct the histograms for each expansion coefficient, θ_j, $j = 1, 2, \ldots, m$, as illustrated in Fig. 3. In our numerical simulations with both synthetic and real data, shifted Legendre polynomials are used to project transmission rate, $\beta(t)$, onto a finite dimensional subspace with $m = 10$ (Smirnova et al., 2022).

To ensure an unbiased choice of initial guess for $\beta(t)$, which is expected to take values between zero and one, the data fitting procedure is initiated with $[\theta_1, \theta_2, \ldots, \theta_m]^T = [0.5, 0, \ldots, 0]^T$ at every bootstrap iteration. This choice of priors for θ_j, $j = 1, 2, \ldots, m$, corresponds to the constant value of $\beta_0(t) = 0.5$.

Due to severe ill-posedness of the parameter estimation inverse problem, with no regularization, iterative scheme (3.12) aimed at the reconstruction of θ_j, $j = 1, 2, \ldots, m$, for the transmission rate, $\beta(t)$, shown in Fig. 2 turns out to be divergent. However, the process can be stabilized with a broad range of initial damping factors, τ_0, as long as they are consistent with the rate of decay of the regularization sequence, $\{\tau_k\}$. In our experiment, we selected $\tau_0 = 10^{-6}$ and the regularization sequence, $\tau_k = 10^{-6}/(k+1)^3$. In general, one needs to have a higher rate of decay for $\{\tau_k\}$ with large initial penalties, τ_0, and a less rapid decay for smaller values of τ_0.

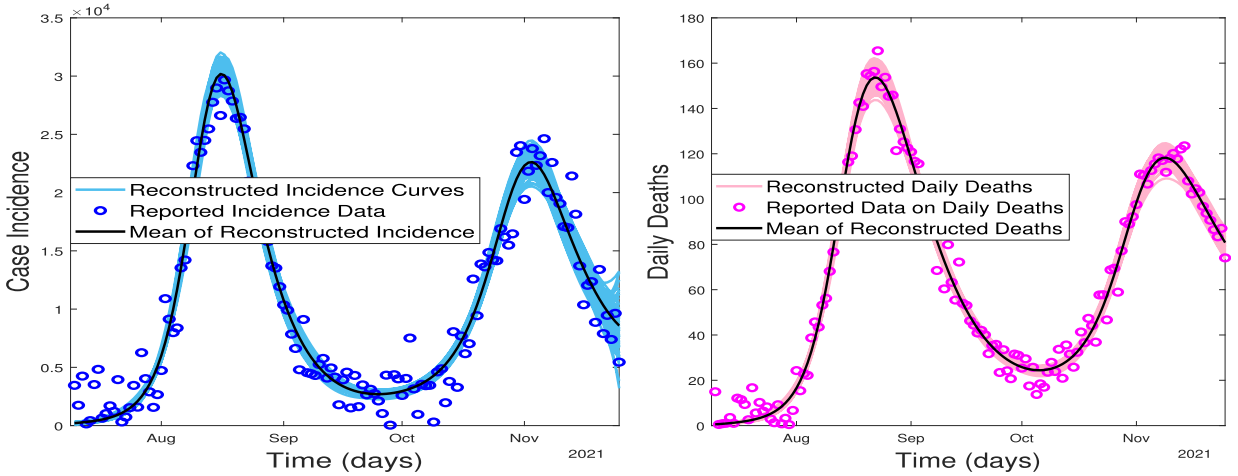

Fig. 1. Synthetic noisy data and data fit for daily new cases (left) and deaths (right).

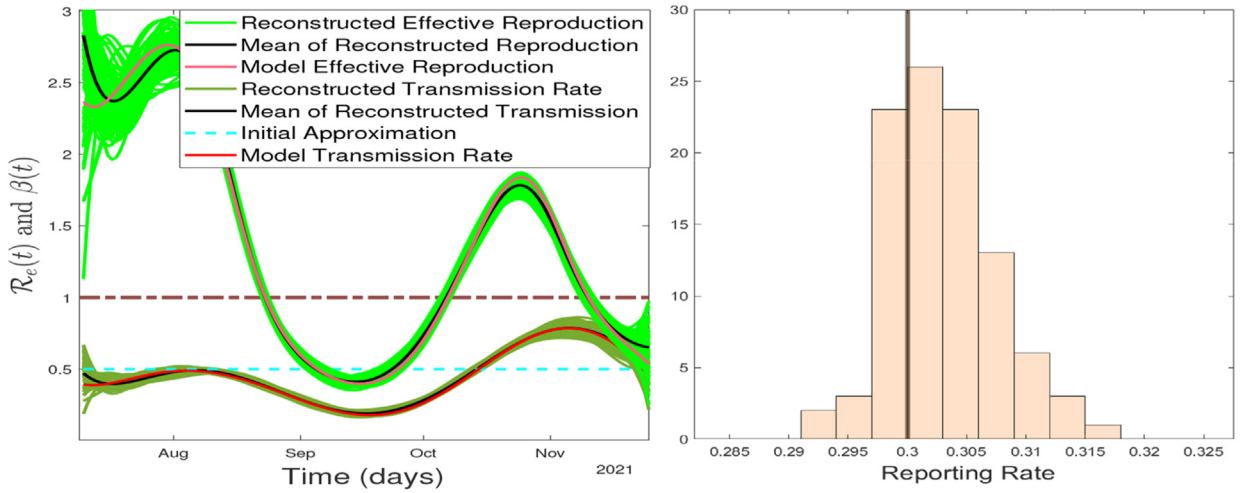

Fig. 2. Reconstructed and true values of disease transmission rate, $\beta(t)$, and effective reproduction number, $\mathcal{R}_e(t)$, (left) along with reconstructed and true incidence reporting rate, $\Psi = 1/\psi$, (right) given noisy synthetic data. The red dotted line shows the desired upper bound for $\mathcal{R}_e(t)$ that would prevent sustained spread of the infection.

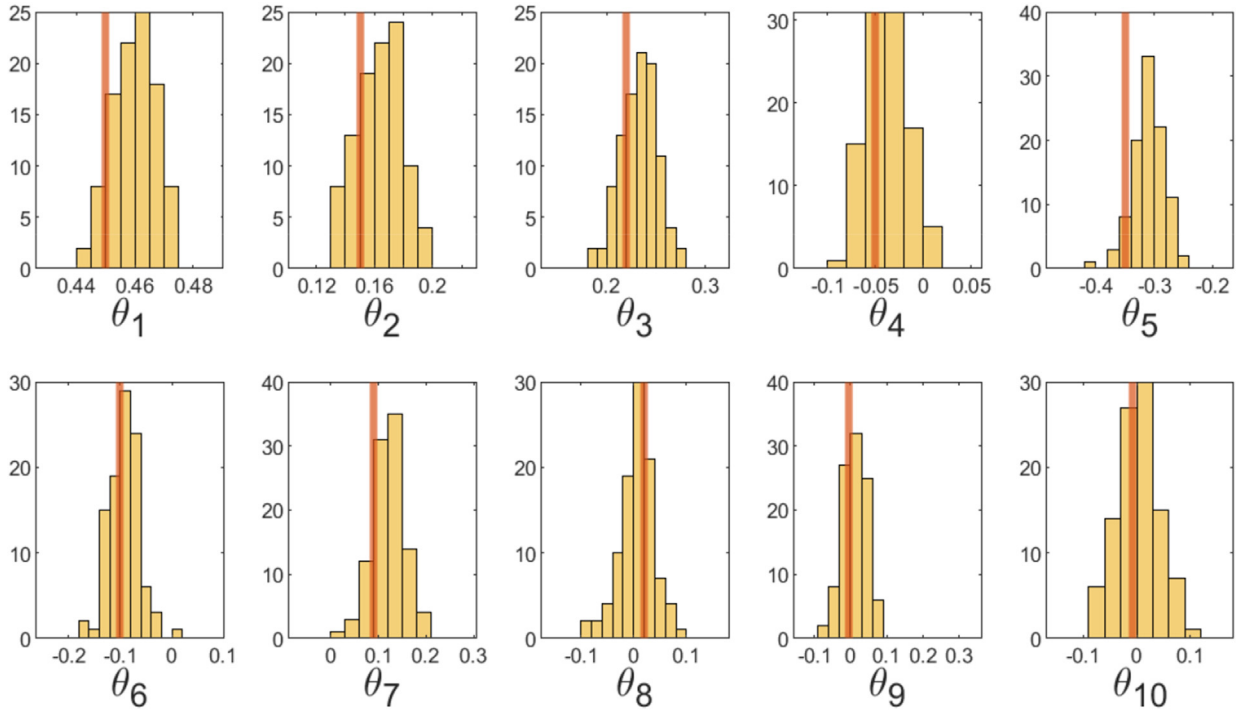

Fig. 3. Reconstructed Legendre expansion coefficients, θ, and true values of θ for model transmission rate, $\beta(t)$, given noisy synthetic data.

The stopping time for iterations (3.12) is set to $k = 5$ and is determined by the goodness of fit to both data sets, Φ and Ω. For the best reconstruction results, one has to avoid over-smoothing and over-fitting, while still adding sufficient regularization to the cost functional in order to keep the process together until convergence.

The values of pre-estimated epidemic parameters in ODE system (2.1)–(2.6) for synthetic and real data are given in Table 2. Pre-estimated initial values of the functions S_u, S_v, I_u, I_v, R, and D used in parameter estimation from synthetic incidence cases and daily new deaths, are given in Table 3. For noisy synthetic data, we take $p = 0.00086$ day^{-1}, which somewhat mimics the reporting rate in the states of Georgia and Texas during he same period of time (Luo et al., 2023). Table 3 also contains pre-estimated values of $S_u(0)$, $S_v(0)$, $I_u(0)$, $I_v(0)$, $R(0)$, and $D(0)$, for Delaware and Maryland (United States Census Bureau; Centers

Table 3
Initial values of state variables for numerical experiments with synthetic and real data (SARS-CoV-2 Delta variant of COVID-19 Pandemic) (United States Census Bureau).

Initial condition	Synthetic Data	Delaware	Maryland
N	4000000	992114	6173205
$S_u(0)$	1997500	496300	2727503
$S_v(0)$	2000000	495651	3445221
$I_u(0)$	2000	80	207
$I_v(0)$	500	83	274
$R(0)$	0	0	0
$D(0)$	0	0	0

for Disease Control and Preventiona; Centers for Disease Control and Preventionb), acquired for numerical simulations with real data on COVID-19 cases and deaths in these two states (see Figs. 4–7).

An important part of parameter estimation is the choice of weights, λ_1 and λ_2, which allows to carefully balance the two fidelity terms in (3.6) and to ensure the model is successfully fitted to data sets on daily new cases and deaths, η and σ, respectively. Without the use of these parameters, since incidence cases and deaths are not of the same order of magnitude, the misfit in daily new deaths is perceived as part of noise in incidence data, and the process is less sensitive to daily new deaths as compared to new incidence cases.

In our numerical study with synthetic data, we used $\lambda_1 = 10^{-12}$ and $\lambda_2 = 10^{-6}$ to balance the data fit and to achieve the most aggressive convergence rate of iteratively regularized algorithm (3.12). Overall, Figs. 1–3 indicate that in our experiments with synthetic data, the uncertainty in the reconstructed reporting rate, Ψ, and the effective reproduction number, $\mathcal{R}_e(t)$, is consistent with the level of noise in incidence data, and the estimation of all unknown parameters is very stable. The reconstructed bundles for $\mathcal{R}_e(t)$ and $\beta(t)$ completely cover the "model" values, and the reconstructed mean values are almost identical to "model" values except for some inevitable oscillations at the end points (Fig. 2 (left)).

The histogram for the recovered reporting rate, Ψ, shows that the "model" value of Ψ, which corresponds to 30% reporting rate, is within 95% confidence interval (CI) for the estimated parameter (Fig. 2 (right)).

Fig. 3 shows that all true values of Legendre coefficients for the transmission rate, $\beta(t)$, are within 95% CI for the reconstructed values. The data fit in Fig. 1 illustrates strong correlation between observation operators, Φ and Ω, produced from the reconstructed transmission rate, $\beta(t)$, and reporting index, ψ, and the data sets, η and σ.

Given noisy incidence synthetic data, iteratively regularized algorithm (3.12) constrained by compartmental model (2.1)–(2.6) estimates the case reporting rate at 30.3%, that is, $\Psi = 0.303$ (95%*CI*: [0.296, 0.312]). For comparison, we have also reconstructed the case reporting rate from cumulative data, which corresponds to the same incidence. The reporting rate approximated from cumulative data is $\Psi = 0.299$ (95%*CI*: [0.298, 0.301]), and the exact value, 0.3, is contained in both CIs. Interestingly, the trend for incidence reporting rate to be slightly higher than cumulative reporting rate continues as we move on to experiments with real data presented in the next section.

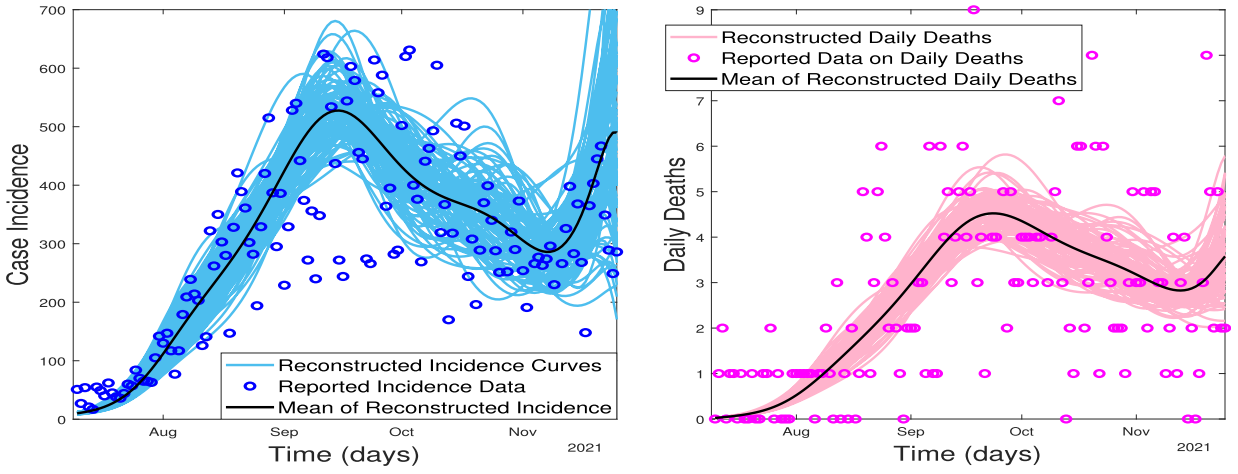

Fig. 4. Real data and data fit for the state of Delaware for incidence cases (left) and daily new deaths (right).

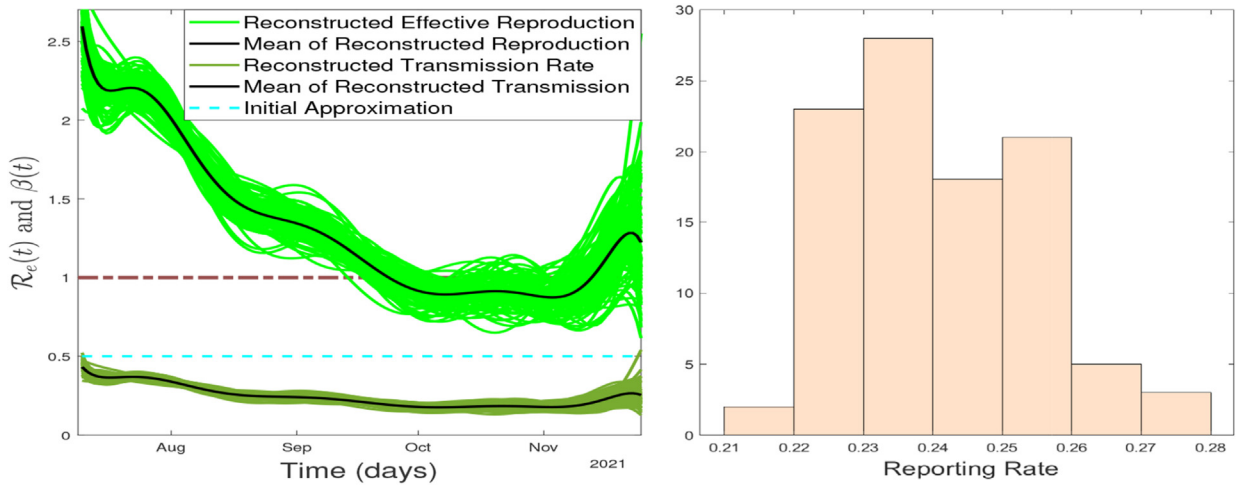

Fig. 5. Reconstructed effective reproduction number, $\mathcal{R}_e(t)$, and disease transmission rate, $\beta(t)$, (left) along with the reconstructed reporting rate, $\Psi = 1/\psi$, (right) for the state of Delaware. The red dotted line shows the desired upper bound for $\mathcal{R}_e(t)$ that would prevent sustained spread of the infection.

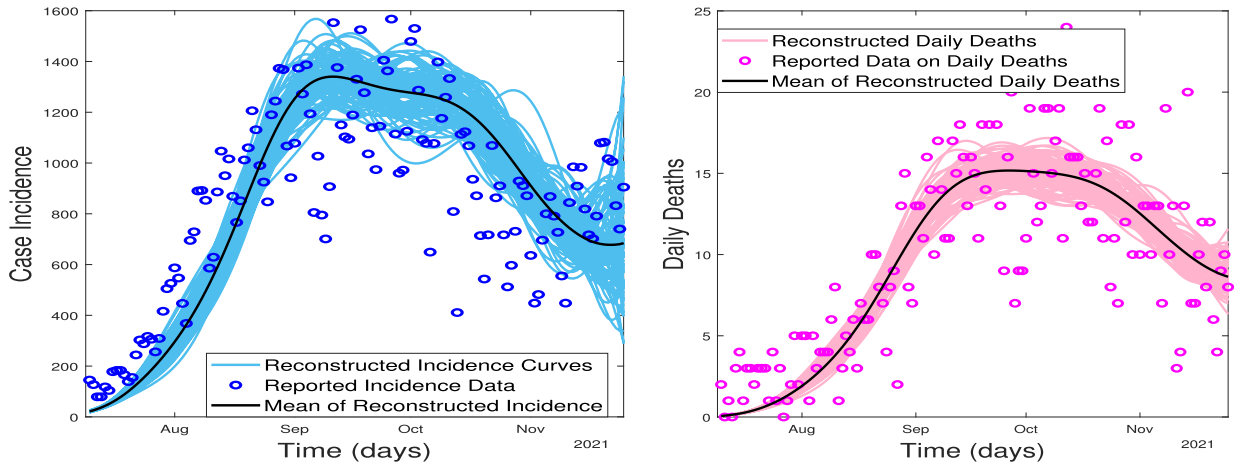

Fig. 6. Real data and data fit for the state of Maryland for incidence cases (left) and daily new deaths (right).

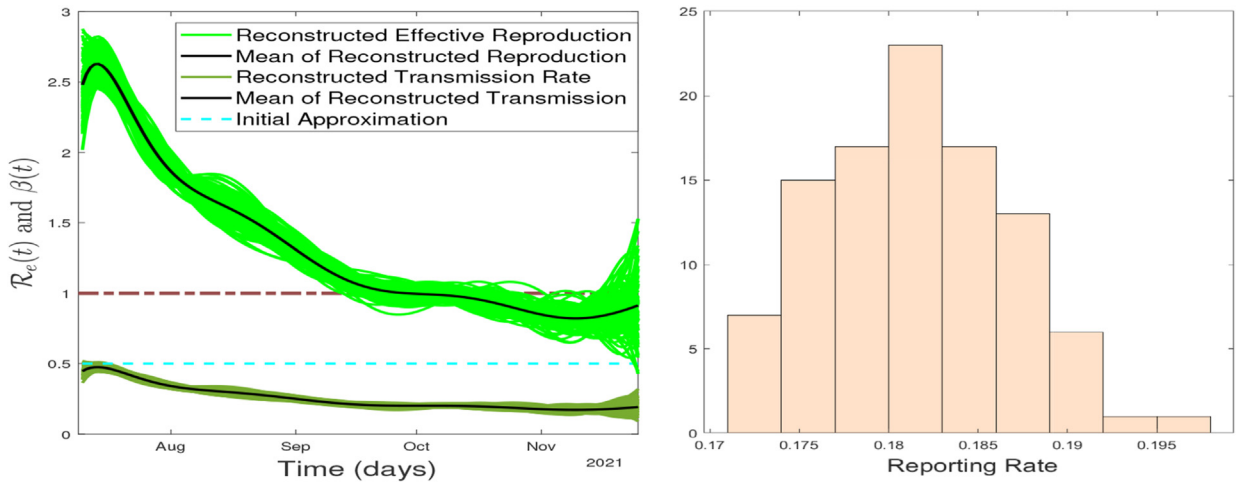

Fig. 7. Reconstructed effective reproduction number, $\mathcal{R}_e(t)$, and disease transmission rate, $\beta(t)$, (left) along with the reconstructed reporting rate, $\Psi = 1/\psi$, (right) for the state of Maryland. The red dotted line shows the desired upper bound for $\mathcal{R}_e(t)$ that would prevent sustained spread of the infection.

5. Simulations with real data

Following our experiments with synthetic data, in this section we use iteratively regularized projected Gauss-Newton algorithm (3.12) together with $S_uS_vI_uI_vRD$ model (2.1)–(2.6) to reconstruct incidence reporting rates, Ψ, and time-dependent effective reproduction numbers, $\mathcal{R}_e(t)$, from real data on incidence cases and new daily deaths, η and σ, on SARS-CoV-2 Delta variant of COVID-19 in different locations of the United States between July 9 and November 25, 2021 (CDC, 2022; Centers for Disease Control and Preventionb).

We procured our datasets from the Centers for Disease Control and Prevention (CDC) (Centers for Disease Control and Preventionb), ensuring its comprehensiveness by incorporating data from all states within the United States. Specifically, we used data on new incidence cases and daily new deaths from (CDC) (Centers for Disease Control and Preventionb) for the states of Alabama, Arizona, Arkansas, Colorado, Delaware, Florida, Georgia, Idaho, Kentucky, Maryland, Michigan, New Jersey, New York, North Carolina, Texas, Washington, and Pennsylvania. However, with some datasets, we encountered certain obstacles, such as incongruous data documentation and restricted data accessibility. To overcome these challenges, we employed a 7-day moving average. This method, averaging data from three days before and after a specific day, helped to smooth out daily variations and bring to light more meaningful trends that might have been obscured in the raw data (due to no reporting or under-reporting over the weekends and holidays, for example). The 7-day average of incidence cases and deaths from CDC (Centers for Disease Control and Preventionb) was used for Alaska, Connecticut, Hawaii, Illinois, Indiana, Iowa, Kansas, Louisiana, Maine, Massachusetts, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Mexico, North Dakota, Oklahoma, Oregon, Rhode Island, South Carolina, Tennessee, Utah, Vermont, West Virginia, Wisconsin, and Wyoming. For the states of Ohio, South Dakota, and Virginia we incorporated 7-day average death data from Johns Hopkins University (JHU) (Johns Hopkins University) since it provided more consistent reporting compared to CDC (Centers for Disease Control and Preventionb). In the case of Ohio, the 7-day average death data from JHU (Johns Hopkins University) was combined with data on new incidence cases from CDC (Centers for Disease Control and Preventionb). For South Dakota and Virginia, the 7-day average death data from JHU (Johns Hopkins University) was used together with the 7-day incidence average from CDC (Centers for Disease Control and Preventionb).

While between July 9 and November 25, 2021, COVID-19 incidence curve emerged as an aggregation of multiple sub-epidemic waves, the Delta strain was dominant and generated most of the cases (CDC, 2022). Newly developed Pfizer-BioNTech, Moderna, and Johnson & Johnson COVID-19 vaccines were distributed to the US general population starting from early 2021 (Centers for Disease Control and Preventiona), and the vaccination campaign had a major impact on the progression of the pandemic during this stage (Lewis et al., 2022; Lopez Bernal et al., 2021; Luo et al., 2023).

In Table 4, percentages of fully vaccinated individuals at the beginning and at the end of the study period are given for 12 different states in the US, representing different geographic locations, population sizes, and intervention policies. To pre-estimate the vaccination rate, p, for each state we divide the increase in vaccination percentage by the length of the study window, which is 140 days (from July 9 to November 25, 2021).

Unlike the case of synthetic data, reconstructions from real incidence are subject to modeling errors and errors in pre-estimated disease parameters in addition to errors resulting from measurements and numerical algorithms. This makes the process considerably less stable, and the right choice of the regularization sequence, $\{\tau_k\}$, becomes even more significant (and, at times, more difficult). Besides, the presence of other variants, rather than Delta, increases the likelihood of iterative scheme (3.12) slipping into a local minimum and fitting the model to one sub-wave instead of a complete data set. For some states, to ensure the convexity of the cost functional, one has to start with extremely heavy penalty (that is, a very large value

Table 4
Proportions of vaccinated individuals (Centers for Disease Control and Preventiona) and pre-estimated COVID-19 Delta variant vaccination rates, July 9 to November 25, 2021, for different states in the US (United States Census Bureau; Centers for Disease Control and Preventiona; Centers for Disease Control and Prevention, 2022, p. 2022).

States	**Population, N, as of 7/9/2021**	Percentage of vaccinated people as of 7/9/2021	Percentage of vaccinated people as of 11/25/2021	**Vaccination rate, p, 7/9–11/25 2021**
Alabama	5,031,362	44.3	54	0.000914286
California	39,501,653	50.8	68.8	0.001285714
Delaware	992,114	50.9	60.9	0.000714286
Florida	21,589,602	46.7	61	0.001021429
Georgia	10,729,828	37.1	49.4	0.000878571
Kentucky	4,507,445	44.2	51.9	0.00055
Maryland	6,173,205	57	67.2	0.000728571
New Hampshire	1,378,587	57.1	64.6	0.000535714
North Carolina	10,449,445	42.4	54	0.000828571
Pennsylvania	12,994,440	50.5	58	0.000535714
Texas	29,232,474	42	54.5	0.000892857
Wisconsin	5,896,271	50.5	59.3	0.000628571

of the initial regularization parameter, τ_0) and then to aggressively reduce the impact of the regularization term by driving τ_k to zero in order to secure convergence to global minimum.

Table 5 shows regularization sequences and stopping times for the same 12 states as in Table 4. To double-check the reconstructed reporting rates, $\Psi = 1/\psi$, apart from incidence data, the experiments have also been carried out with cumulative data.

As pointed out in (Saucedo, Martcheva, & Annor, 2019), incidence data are corrupt with independent and identically distributed (i.i.d.) errors potentially leading to more accurate parameter estimations as compared to estimations from cumulative data. However, due to reporting irregularities and cases being under-reported (or not reported) over weekends and holidays in some of the states, incidence data sets are often "spread out" making it hard to clinch an adequate data fit. For this reason, repeating parameter estimation with cumulative data allows to ensure that the estimation is stable, and the error in our reconstructed parameters, Ψ and $\mathcal{R}_e(t)$, is consistent with the level of noise in reported cases.

As opposed to incidence, cumulative data suffer from the dominance of earlier cases, which results in a growing noise propagation towards the end of the study window. But on the bright side, cumulative sets are smooth, and any misfit can be easily detected and mitigated with a better choice of regularization parameters.

One can see from Table 5 that, due to error buildup in cumulative data, for almost all cumulative sets the initial values of regularization parameters, τ_0, are several orders of magnitude higher as compared to incidence. This, in turn, requires a higher rate of decay for τ_k in some of the cases. Nevertheless, the approximate reporting rates, Ψ, obtained from incidence and cumulative data are very similar, and most values of "cumulative Ψ" are slightly lower than the corresponding values of "incidence Ψ" (see Table 6). The bundles of reconstructed curves for effective reproduction numbers, $\mathcal{R}_e(t)$, and their underlying transmission rates, $\beta(t)$, are virtually identical.

6. Conclusions and future plans

Prior research underscores the importance of structural and practical identifiability of system parameters reconstructed from compartmental deterministic and stochastic models for infectious disease transmission (Eisenberg, Robertson, & Tien, 2013; Nemeth, Tuncer, & Martcheva, 2023; Roosa & Chowell, 2019; Tuncer et al., 2016, 2022). The lack of identifiability, when multiple parameters fit the same data with a comparable level of accuracy, may be attributed to model structure (structural identifiability (Nemeth et al., 2023; Tuncer et al., 2022)) or to incomplete and/or inconsistent data used for validation of the model (practical identifiability (Roosa & Chowell, 2019)). This is closely related to the notion of ill-posedness of parameter estimation inverse problems, and advanced regularization techniques, originated from the fields of inverse scattering and biomedical imaging (Kaltenbacher et al., 2008), have emerged as a powerful tool aimed at stable estimation of disease parameters that are crucial for future projections, prevention, and control.

Regularization is introduced into compartmental models as *a priori* information in the form of some pre-estimated parameters (to address the issue of structural identifiability) and into the cost functionals, designed to fit nonlinear observation operators to epidemic data, as problem-specific penalties that secure practical identifiability.

Unlike other system parameters, i.e., incubation and recovery rates, the case reporting rate, Ψ, and the time-dependent effective reproduction number, $\mathcal{R}_e(t)$, are directly influenced by a large number of factors making it impossible to pre-estimate these parameters in any meaningful way. In this study, we proposed a novel iteratively-regularized optimization algorithm (3.12), combined with $S_uS_vI_uI_vRD$ compartmental model (2.1)–(2.6) (Luo et al., 2023), for stable reconstruction of Ψ and $\mathcal{R}_e(t)$ from reported data on vaccination percentages, incidence cases, and daily deaths. The innovative regularization procedure exploits (and takes full advantage of) a unique structure of the Jacobian and Hessian approximation for the nonlinear observation operator. The proposed inversion method is thoroughly tested with synthetic and real SARS-CoV-2 Delta variant data for different regions in the United States of America from July 9, 2021, to November 25, 2021.

Table 5
Regularization sequences and stopping times in algorithm (3.12) for different states in the US.

States	**Regularization sequence, τ_k, and stopping time, k, for incidence data**	**Regularization sequence, τ_k, and stopping time, k, for cumulative data**
Alabama	$\tau_k = 10^{-8}/(k+1)^{0.1}$, $k = 15$	$\tau_k = 10^{-1}/(k+1)^{3.5}$, $k = 15$
California	$\tau_k = 5 \cdot 10^{26}/(k+1)^{32}$, $k = 15$	$\tau_k = 5 \cdot 10^{24}/(k+1)^{32}$, $k = 15$
Delaware	$\tau_k = 10^{-7}/(k+1)^{0.1}$, $k = 15$	$\tau_k = 10^{-1}/(k+1)^{3.5}$, $k = 15$
Florida	$\tau_k = 10^{8}/(k+1)^{9}$, $k = 50$	$\tau_k = 10^{8}/(k+1)^{9}$, $k = 50$
Georgia	$\tau_k = 10^{9}/(k+1)^{10}$, $k = 50$	$\tau_k = 10^{12}/(k+1)^{9}$, $k = 60$
Kentucky	$\tau_k = 2 \cdot 10^{-8}/(k+1)^{0.5}$, $k = 10$	$\tau_k = 2 \cdot 10^{-6}/(k+1)^{0.5}$, $k = 10$
Maryland	$\tau_k = 5 \cdot 10^{-9}/(k+1)^{0.1}$, $k = 15$	$\tau_k = 5 \cdot 10^{-6}/(k+1)^{0.1}$, $k = 15$
New Hampshire	$\tau_k = 10^{-6}/(k+1)^{1.6}$, $k = 15$	$\tau_k = 10^{-5}/(k+1)^{1.6}$, $k = 15$
North Carolina	$\tau_k = 3 \cdot 10^{-8}/(k+1)^{0.5}$, $k = 15$	$\tau_k = 3 \cdot 10^{-5}/(k+1)^{0.5}$, $k = 15$
Pennsylvania	$\tau_k = 10^{-8}/(k+1)^{0.5}$, $k = 15$	$\tau_k = 5 \cdot 10^{-6}/(k+1)^{0.5}$, $k = 30$
Texas	$\tau_k = 10^{-8}/(k+1)^{0.1}$, $k = 15$	$\tau_k = 5 \cdot 10^{-2}/(k+1)^{6}$, $k = 5$
Wisconsin	$\tau_k = 2 \cdot 10^{-5}/(k+1)^{0.1}$, $k = 15$	$\tau_k = 10^{-2}/(k+1)^{0.1}$, $k = 25$

Table 6
COVID-19 Delta variant case reporting rates reconstructed from incidence and cumulative data (Centers for Disease Control and Preventionb), July 9 to November 25, 2021, for different states in the US.

States	**Reconstructed reporting rate, $\Psi = 1/\psi$, from incidence data**	**Reconstructed reporting rate, $\Psi = 1/\psi$, from cumulative data**
Alabama	0.154 (95%CI:[0.149,0.159])	0.141 (95%CI:[0.14,0.141])
California	0.214 (95%CI:[0.209,0.219])	0.206 (95%CI:[0.206,0.207])
Delaware	0.241 (95%CI:[0.220,0.271])	0.220 (95%CI:[0.216,0.224])
Florida	0.132 (95%CI:[0.130,0.134])	0.125 (95%CI:[0.124,0.125])
Georgia	0.188 (95%CI:[0.183,0.190])	0.166 (95%CI:[0.166,0.167])
Kentucky	0.181 (95%CI:[0.176,0.187])	0.166 (95%CI:[0.165,0.167])
Maryland	0.182 (95%CI:[0.172,0.192])	0.173 (95%CI:[0.171,0.175])
New Hampshire	0.364 (95%CI:[0.336,0.397])	0.335 (95%CI:[0.329,0.341])
North Carolina	0.229 (95%CI:[0.223,0.234])	0.216 (95%CI:[0.215,0.217])
Pennsylvania	0.161 (95%CI:[0.154,0.166])	0.180 (95%CI:[0.179,0.181])
Texas	0.177 (95%CI:[0.174,0.179])	0.152 (95%CI:[0.151,0.156])
Wisconsin	0.359(95%CI:[0.342,0.376])	0.325 (95%CI:[0.323,0.327])

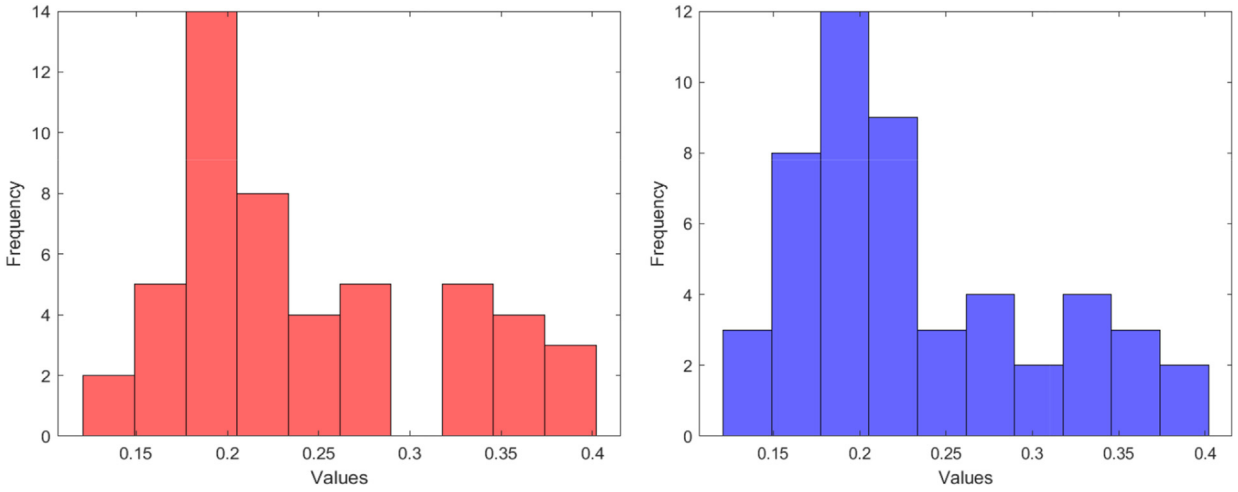

Fig. 8. Delta variant case reporting rates reconstructed from incidence (left) and cumulative (right) data (Centers for Disease Control and Preventionb), July 9 to November 25, 2021, for 50 states in the US.

Our study shows that case reporting rate during the Delta wave of COVID-19 pandemic in the US is between 12% and 37%, with most states being in the range from 15% to 25% (see Table 6 and Fig. 8). This confirms earlier accounts on considerable under-reporting of COVID-19 cases due to the impact of silent transmissions, resulting from a combination of presymptomatic and asymptomatic infections, and the limitations of testing (Luo et al., 2023; Smirnova et al., 2022; Tuncer et al., 2022). An important goal for our future research is to design a regularized numerical procedure with observation operators connected to a stochastic compartmental model. The model will be governed by Wiener processes, which take into account the uncertainty of disease transmission, incubation period, and variability of detection.

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

CRediT authorship contribution statement

Alexandra Smirnova: Conceptualization, Formal analysis, Investigation, Project administration, Software, Supervision, Writing – original draft, Writing – review & editing. **Mona Baroonian:** Investigation, Software.

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.

References

Agossou, O., Nicodème Atchadé, M., & Moussa Djibril, A. (2021). Modeling the effects of preventive measures and vaccination on the COVID-19 spread in Benin Republic with optimal control. *Results in Physics, 31*, Article 104969.

Aguilar, J. B., Samuel Faust, J., Westafer, L. M., & Gutierrez, J. B. (2020). *A model describing COVID-19 community transmission taking into account asymptomatic carriers and risk mitigation*. medRxiv.
Angeli, M., Neofotistos, G., Mattheakis, M., & Kaxiras, E. (2022). Modeling the effect of the vaccination campaign on the COVID-19 pandemic. *Chaos, Solitons & Fractals, 154*, Article 111621.
Bakushinsky, A. B., & Kokurin, M. Y. (2004). *Iterative methods for ill-posed operator equations with smooth operators*. Dordrecht, Great Britain: Springer.
Betti, M., Luigi Bragazzi, N., Heffernan, J. M., Kong, J., & Raad, A. (2021). Integrated vaccination and non-pharmaceutical interventions based strategies in Ontario, Canada, as a case study: A mathematical modelling study. *Journal of The Royal Society Interface, 18*(180).
Bjørnstad, O. N., Finkenstädt, B. F., & Grenfell, B. T. (2002). Dynamics of measles epidemics: Estimating scaling of transmission rates using a time series sir model. *Ecological Monographs, 72*(2), 169–184.
Calafiore, G. C., Novara, C., & Possieri, C. (2020). A time-varying SIRD model for the COVID-19 contagion in Italy. *Annual Reviews in Control, 50*, 361–372.
CDC. (2022). *Trends in number of COVID-19 cases and deaths in the US reported to CDC*. State/Territory.
Centers for Disease Control and Preventiona. (2023). Covid-19 vaccinations in the United States jurisdiction. https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-Jurisdi/unsk-b7fc. (Accessed 6 June 2023).
Centers for Disease Control and Preventionb. United States covid-19 cases and deaths by state over time (archived). https://data.cdc.gov/Case-Surveillance/United-States-COVID-19-Cases-and-Deaths-by-State-o/9mfq-cb36. (Accessed 6 June 2023).
Centers for Disease Control and Prevention. (2022). *Trends in number of covid-19 vaccinations in the us*.
Chowell, G., Viboud, C., Simonsen, L., & Moghadas, S. (2016). Characterizing the reproduction number of epidemics with early sub-exponential growth dynamics. *Journal of The Royal Society Interface, 13*, Article 20160659.
Diekmann, O., Heesterbeek, J. A. P., & Metz, J. A. J. (1990). On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology, 28*(4), 365–382.
Doria-Rose, N., Suthar, M. S., Makowski, M., O'Connell, S., McDermott, A. B., Flach, B., et al. (2021). Antibody persistence through 6 months after the second dose of mRNA-1273 vaccine for Covid-19. *New England Journal of Medicine, 384*(23), 2259–2261.
van den Driessche, P., & Watmough, J. (2002). Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences, 180*(1–2), 29–48.
Eisenberg, M. C., Robertson, S. L., & Tien, J. H. (2013). Identifiability and estimation of multiple transmission pathways in cholera and waterborne disease. *Journal of Theoretical Biology, 324*, 84–102.
Fenichel, E. P., Castillo-Chavez, C., Ceddia, M. G., Chowell, G., Gonzalez Parra, P. A., Hickling, G. J., et al. (2011). Adaptive human behavior in epidemiological models. *Proceedings of the National Academy of Sciences, 108*(15), 6306–6311.
Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., & Colaneri, M. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine, 26*, 855–860.
Giordano, G., Franco, B., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A., et al. (2020). Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine, 26*(6), 855–860.
Gustavo Barbosa Libotte, Fran Sérgio Lobato, Gustavo Mendes Platt, & Silva Neto, A. J. (2020). Determination of an optimal control strategy for vaccine administration in COVID-19 pandemic treatment. *Computer Methods and Programs in Biomedicine, 196*, Article 105664.
He, S., Peng, Y., & Sun, K. (2020). SEIR modeling of the COVID-19 and its dynamics. *Nonlinear Dynamics, 101*, 1667–1680.
Johns Hopkins University. Covid-19 deaths dataset for the United States. https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_deaths_US.csv. (Accessed 20 June 2023).
Johnson, A. G. (2022). COVID-19 incidence and death rates among unvaccinated and fully vaccinated adults with and without booster doses during periods of delta and omicron variant emergence—25 US jurisdictions, april 4–december 25, 2021. *MMWR. Morbidity and Mortality Weekly Report, 71*.
Kaltenbacher, B., Neubauer, A., & Scherzer, O. (2008). *Iterative regularization methods for nonlinear ill-posed problems. Radon Series on Computational and Applied Mathematics*. Berlin: Walter de Gruyter.
Kosmidis, K., & Macheras, P. (2020). A fractal kinetics SI model can explain the dynamics of COVID-19 epidemics. *PLoS One, 15*(8), Article e0237304.
Kudryashov, N. A., Chmykhov, M. A., & Vigdorowitsch, M. (2021). Analytical features of the SIR model and their applications to COVID-19. *Applied Mathematical Modelling, 90*, 466–473.
Leontitsis, A., Abiola, S., Alsheikh-Ali, A., Al Nasser, Y., Tom, L., & Alshamsi, A. (2021). Seahir: A specialized compartmental model for covid-19. *International Journal of Environmental Research and Public Health, 18*(5), 2667.
Lewis, N., Chambers, L. C., Chu, H. T., Taylor, F., De Vito, R., Gargano, L. M., et al. (2022). Effectiveness associated with vaccination after COVID-19 recovery in preventing reinfection. *JAMA Network Open, 5*(7), Article e2223917–e2223917.
Li, L., Yang, Z., Dang, Z., Meng, C., Huang, J., Meng, H., et al. (2020). Propagation analysis and prediction of the COVID-19. *Infectious Disease Modelling, 5*, 282–292.
Lopez Bernal, J., Andrews, N., Gower, C., Gallagher, E., Simmons, R., Simon, T., et al. (2021). Effectiveness of covid-19 vaccines against the B. 1.617. 2 (delta) variant. *New England Journal of Medicine, 385*(7), 585–594.
Luo, R., Herrera-Reyes, A. D., Kim, Y., Rogowski, S., White, D., & Smirnova, A. (2023). Estimation of time-dependent transmission rate for COVID-19 SVIRD model using predictor-corrector algorithm. *Mathematical Modeling for Women's Health - Collaborative Workshop for Women in Mathematical Biology*. Revision submitted.
Ma, Q., Liu, J., Qiao, L., Kang, L., Liu, R., Jing, W., et al. (2021). Global percentage of asymptomatic SARS-CoV-2 infections among the tested population and individuals with confirmed COVID-19 diagnosis: A systematic review and meta-analysis. *JAMA Network Open, 4*(12), Article e2137257–e2137257.
Michael Otunuga, O. (2021). Time-dependent probability distribution for number of infection in a stochastic SIS model: Case study COVID-19. *Chaos, Solitons & Fractals, 147*, Article 110983.
Mizumoto, K., Kagaya, K., Alexander, Z., & Chowell, G. (2020). Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan. *Euro Surveillance, 25*(10), 2020.
Moghadas, S. M., Fitzpatrick, M. C., Sah, P., Pandey, A., Shoukat, A., Singer, B. H., et al. (2020). The implications of silent transmission for the control of COVID-19 outbreaks. *Proceedings of the National Academy of Sciences, 117*(30), 17513–17515.
Nemeth, L., Tuncer, N., & Martcheva, M. (2023). Structural and practical identifiability analysis of a multiscale immuno-epidemiological model. In *Computational and mathematical population dynamics* (pp. 169–201). World Scientific.
Oran, D. P., & Eric, J. T. (2021). The proportion of SARS-CoV-2 infections that are asymptomatic: A systematic review. *Annals of Internal Medicine, 174*(5), 655–662.
Patalon, T., Gazit, S., Pitzer, V. E., Prunas, O., Warren, J. L., & Weinberger, D. M. (2022). Odds of testing positive for SARS-CoV-2 following receipt of 3 vs 2 doses of the BNT162b2 mRNA vaccine. *JAMA Internal Medicine, 182*(2), 179–184.
Roosa, K., & Chowell, G. (2019). Assessing parameter identifiability in compartmental dynamic models using a computational approach: Application to infectious disease transmission models. *Theoretical Biology and Medical Modelling, 16*(1), 1.
Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J. M., et al. (2020). Short-term forecasts of the COVID-19 epidemic in guangdong and zhejiang, China: February 13–23. *Journal of Clinical Medicine, 9*, 596.
Saucedo, O., Martcheva, M., & Annor, A. (2019). Computing human to human Avian influenza ${\mathcal{R}}_{0}$ via transmission chains and parameter estimation. *Mathematical Biosciences and Engineering, 16*(5), 3465–3487.
Shang, W., Kang, L., Cao, G., Wang, Y., Gao, P., Liu, J., et al. (2022). Percentage of asymptomatic infections among SARS-CoV-2 omicron variant-positive individuals: A systematic review and meta-analysis. *Vaccines, 10*(7), 1049.
Sharma, S., Volpert, V., & Banerjee, M. (2020). Extended SEIQR type model for COVID-19 epidemic and data analysis. *Mathematical Biosciences and Engineering, 17*(6), 7562–7604.

Smirnova, A., & Bakushinsky, A. (2020). On iteratively regularized predictor–corrector algorithm for parameter identification. *Inverse Problems, 36*(12), Article 125015.
Smirnova, A., Pidgeon, B., & Luo, R. (2022). On stable parameter estimation and short-term forecasting with quantified uncertainty with application to COVID-19 transmission. *Journal of Inverse and Ill-Posed Problems, 30*(6), 823–844.
Smirnova, A., Sterrett, N., Mujica, O. J., Munayco, C., Suárez, L., Viboud, C., et al. (2020). Spatial dynamics and the basic reproduction number of the great cholera epidemic in Peru. *PLoS Neglected Tropical Diseases, 14*(7), Article e0008045.
Subramanian, R., He, Q., & Pascual, M. (2021). Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity. *Proceedings of the National Academy of Sciences, 118*(9), Article e2019716118.
Tang, B., Wang, X., Qian, L., Bragazzi, N. L., Tang, S., Xiao, Y., et al. (2020). Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *Journal of Clinical Medicine, 9*(2).
Thompson, R. N. (2020). Epidemiological models are important tools for guiding COVID-19 interventions. *BMC Medicine, 18*(152).
Tuncer, N., Gulbudak, H., Cannataro, V. L., & Martcheva, M. (2016). Structural and practical identifiability issues of immuno-epidemiological vector–host models with application to rift valley fever. *Bulletin of Mathematical Biology, 9*(78), 1796–1827.
Tuncer, N., Timsinaa, A., Nunob, M., Chowell, G., & Martcheva, M. (2022). Parameter identifiability and optimal control of an SARS-CoV-2 model early in the pandemic. *Journal of Biological Dynamics, 16*(1), 412–438.
United States Census Bureau. State population totals and components of change: 2020-2022. https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-total.html. (Accessed 6 June 2023).