Heliyon 10 (2024) e31410

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Compartmental modeling for pandemic data analysis: The gap between statistics and models

Leonidas Sakalauskas [a,d,*], Vytautas Dulskis [b], Rimas Jonas Jankunas [c,d]

[a] Klaipeda University, H. Manto st. 84, Klaipeda, LT-92294, Lithuania
[b] Vilnius University Institute of Data Science and Digital Technologies, Akademijos st. 4, Vilnius, LT-08412, Lithuania
[c] Medical Academy, Lithuanian University of Health Sciences, Mickeviciaus st. 9, Kaunas, LT-44307, Lithuania
[d] Health Law Institute, Ulonu st. 5-303, Vilnius, LT-08240, Lithuania

ARTICLE INFO

Keywords:
COVID-19
COVID-19 passports
Compartmental modeling
Maximum likelihood estimation
COVID-19 deaths

ABSTRACT

A scrutiny analysis of the COVID-19 data is required to get insights into effective strategies for pandemic control. However, there is a gap between official data and methods used to assess the effectiveness of the potential measures, which was partly addressed in an editorial-letter-type discussion on the impact of the COVID-19 passport in Lithuania. The therein-applied descriptive statistics method provides only limited evidence, while detailed analysis requires more sensitive and reliable methods. In this regard, this paper advocates a maximum likelihood compartmental modeling approach, which provides the flexibility to raise various hypotheses about infection, recovery, and mortality dynamics and to find the most likely answers given the data. Our paper is based on COVID-19 deaths, which are more reliable and essential than infection cases. It should also be noted that officially collected data are unsuitable for in-depth analyses, including compartmental modeling, as they do not capture important information. Overall, this paper does not aim to solve the underlying problems completely but rather stimulate a discussion.

1. Introduction

The COVID-19 pandemic remains a hot area of research even after its relative decline [1–4]. Analysis of officially collected infection, mortality, and related statistics may lead to findings that may prove crucial in predicting the course of other pandemics and developing strategies to contain them [5,6]. However, a particular gap can be seen between officially collected COVID-19 pandemic statistics and models that allow in-depth analysis of the efficacy of its management measures such as quarantines [7], masks [8], COVID-19 passports [9], or mass vaccination [10], thus raising concerns about whether high-quality inferences can be made from available data [11]. The paper aims to highlight this gap, propose an adequate approach for working with the available data so that at least some of its limiting impacts can be dealt with, and encourage pandemic data stakeholders to rethink data collection and employment.

The motivation for examining such a seemingly unproblematic topic stems from the discussion raised in editorial letters on the impact of the COVID-19 passport on the spread of the virus in Lithuania [12–14]. In the letter that initiated the discussion, this impact was assessed using the principles and methods of descriptive statistics as well as some common sense assumptions. However,

* Corresponding author at: Vilnius University Institute of Data Science and Digital Technologies, Akademijos st. 4, Vilnius, LT-08412, Lithuania.
E-mail address: leonidas.sakalauskas@mif.vu.lt (L. Sakalauskas).

https://doi.org/10.1016/j.heliyon.2024.e31410
Received 31 January 2024; Received in revised form 10 May 2024; Accepted 15 May 2024

Available online 22 May 2024

in response to it, the discussion counterparty made notice of the potential fragility of such an approach, hinting that the obtained conclusions may be misleading due to issues with both the method and the data.

In fact, when examining the characteristics of pandemic-related processes, researchers often resort to descriptive statistics that enable summarizing or describing the characteristics of a data set through the calculation of measures of central tendency, variability, and frequency distribution [15]. However, such methods tend to achieve only some illustrative goals with limited scientific conclusion potential rather than a comprehensive and reliable analysis of the phenomena of interest since any aggregation of data, in the broadest sense, results in a loss of information, reducing the sensitivity of the modeling to often essential details, which in turn undermines the quality of results [16].

On the contrary, compartmental models are often used in epidemiology to model at a lower level, which is enabled by dividing objects of scientific interest into smaller conceptual units until the underlying mechanisms become apparent (feeding the model with delicate granularity data constitutes yet another condition for low-level modeling) [17]. In essence, compartmental models are a very general modeling technique. In this modeling framework, the population is assigned to compartments with labels such as, for example, S, I, R, and D (which stand for Susceptible (i.e., those who can potentially be infected in the future), Infected (i.e., those who are currently infected), Recovered (i.e., those who have already recovered), and Deceased (i.e., those who have already deceased), respectively) [18]. People are allowed to progress from one compartment to another according to some rules imposed by model structure, which is the realization of modeling assumptions [19]. The models are most often run with ordinary differential equations (which are deterministic) but can also be used with a stochastic (random) framework, which is more realistic but much more complicated to analyze [20]. Models of this type try to predict how a disease spreads, the number of people infected, or the duration of a pandemic and to estimate various epidemiological parameters such as the reproduction number. Such models can also attempt to explain how different public health interventions imposed by authorities may affect the outcome of the pandemic [21,22].

In principle, epidemiological compartmental models are dynamic models with a structure and free parameters usually expressed by an overall non-linear set of equations [23]. To use such models to analyze real-world processes, one needs to estimate the model's unknown parameters from available actual data. The accuracy of the parameter estimates broadly depends on the estimation method used and the adequacy, quality, and volume of data available [24,25].

In terms of method, parameter estimation should ideally be carried out using classical methods with optimal estimator properties, such as the maximum likelihood estimation [26], which allows the most asymptotically robust conclusions possible. It is true that the maximum likelihood approach often proves to be hardly tractable (especially given high model complexity and a large amount of data), leading researchers to resort to various heuristics that are usually easier to compute but harm the quality of the estimates [27]. Thus, it must be explicitly stressed that efficient parameter estimation inevitably requires a certain level of effort to establish suitable mathematical apparatus [28].

In terms of data, parameter estimation ideally requires data describing all the compartments that comprise the compartmental model. However, the officially collected and available COVID-19 data are somewhat limited in this regard, thus burdening the proper and widespread applicability of compartmental models for modeling the pandemic. It is because there are simple no (assumption-free) data for specific compartments of modeling interest, or the seemingly appropriate data turns out to suffer from various issues affecting its accuracy towards intended modeling purposes [11,25,29]. As a matter of fact, some of the data limitations might be offset by using more complex models (i.e., more equations and parameters) that account for them by design. However, the increased model complexity might hinder its solvability and, in turn, application, thus leading to the similar result as that caused by the inappropriate data or the lack thereof.

Remark 1. When we refer to available COVID-19 data throughout the paper, we mainly refer to data available under the COVID-19 topic page of OurWorldInData.org [30].

Reviewing the available data through the prism of the aforementioned (daily-level, which corresponds to the finest available data granularity) SIRD model, which consists of just four compartments, can already help demonstrate the issues present with them. Here, the suitability-wise closest data for the Infected assessment are *confirmed cases (new per day)* [31] (see Fig. 1). However, it must be stressed that these data are characterized by several problems: 1) they are inseparable from the testing volumes (which are expressed by the *tests (new per day)* [32] data), which vary significantly from day to day (see Fig. 2); 2) only part of the population is tested, so the number of *confirmed cases (new per day)* [31] is not representative of the population as a whole; 3) the tests may give false results (a person who tested positively may be not infected, or vice versa); 4) the time lag between the publication of test results and the execution of the tests is, in principle, variable and unknown (Fig. 3 shows the *share of positive tests* [33] metric with *confirmed cases (new per day)* [31] shifted by varying number of days with respect to *tests (new per day)* [32]; it can be seen that the data relatively loses periodicity when this shift is one day, but even in this case it remains significant).

These problems significantly complicate the identification of the actual daily flow going into the Infected compartment (in order to avoid some of the problems (i.e., 1 and 4), often smoothed data are used (see Fig. 4), but despite the problems mentioned above with the aggregation of data, it also complicates the application of compartmental models from a technical point of view, as this effectively introduces data dependencies [34]). Moreover, entry to the Recovered compartment poses a similar challenge, as there is simply no general practice of testing for recovery. As a consequence, one can consider just an estimate that stems from assuming that the current number of infected persons is the total number of cases confirmed during the last n days (the State Data Agency of Lithuania uses $n = 20$ [35]).

However, the situation is more favorable with entry to the Deceased compartment, as it is relatively well tracked by the *confirmed deaths (new per day)* [36] data (see Fig. 5). This is because these data are relatively easily captured in full on the actual date of death.

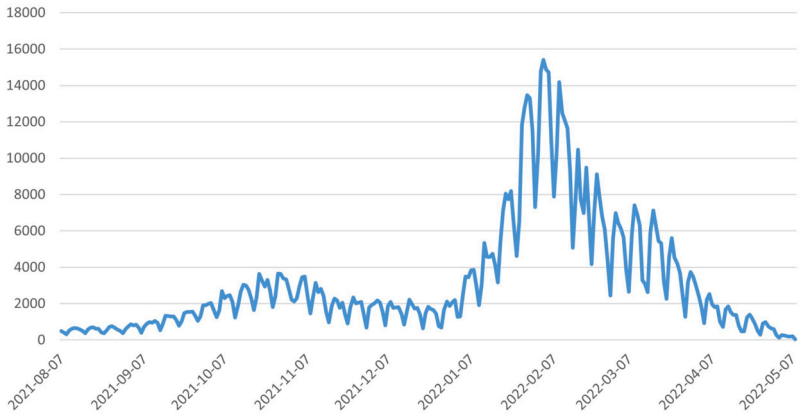

Fig. 1. Confirmed cases (new per day) in Lithuania.

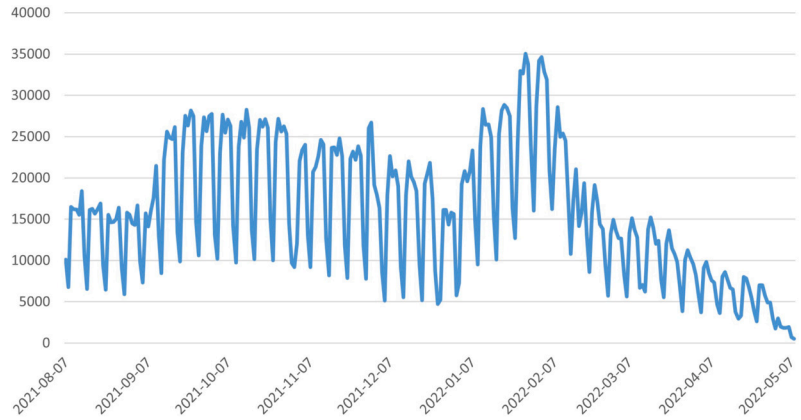

Fig. 2. Tests (new per day) in Lithuania.

The only real drawback of them is the possible misattribution of the cause of death (analogous to problem point 3 of the *confirmed cases (new per day)* [31] data).

In the absence of good-quality data for each of the modeled compartments, it might make sense to perform compartmental modeling by expressing the model in terms of compartments with the most satisfactory data. In this paper, we argue that, in the case of the SIRD model, such a compartment is Deceased, and we advocate the maximum likelihood method for estimating the parameters of the SIRD model based solely on COVID-19 deaths, the limitations of which are accounted for by introducing observational noise into the model. Having turned this approach into a particular modeling solution, we then apply it to actual COVID-19 death data from Lithuania, analyzing the impact of the COVID-19 passport in the delta and omicron wave cases. Finally, we discuss possible improvements in both the data and the model.

Remark 2. We explicitly acknowledge weaknesses of modeling based on COVID-19 deaths, as there are no standard foolproof criteria for their attribution, and very few autopsies were performed on people presumably dead from COVID-19. Nevertheless, these data are more reliable than COVID-19 cases used in other publications [12,37–39] because cases depend on the volume and strategy of testing as well as test used.

2. Materials and methods

The paper considers a discrete-time deterministic SIRD model. In this model, the movement between compartments is defined by the following equations (the meaning of the compartments is given in Section 1):

$$S_{i+1} = S_i - aI_iS_i \tag{1}$$

$$I_{i+1} = I_i\left(1 + aS_i - b - c\right) \tag{2}$$

$$R_{i+1} = R_i + bI_i \tag{3}$$

$$D_{i+1} = D_i + cI_i \tag{4}$$

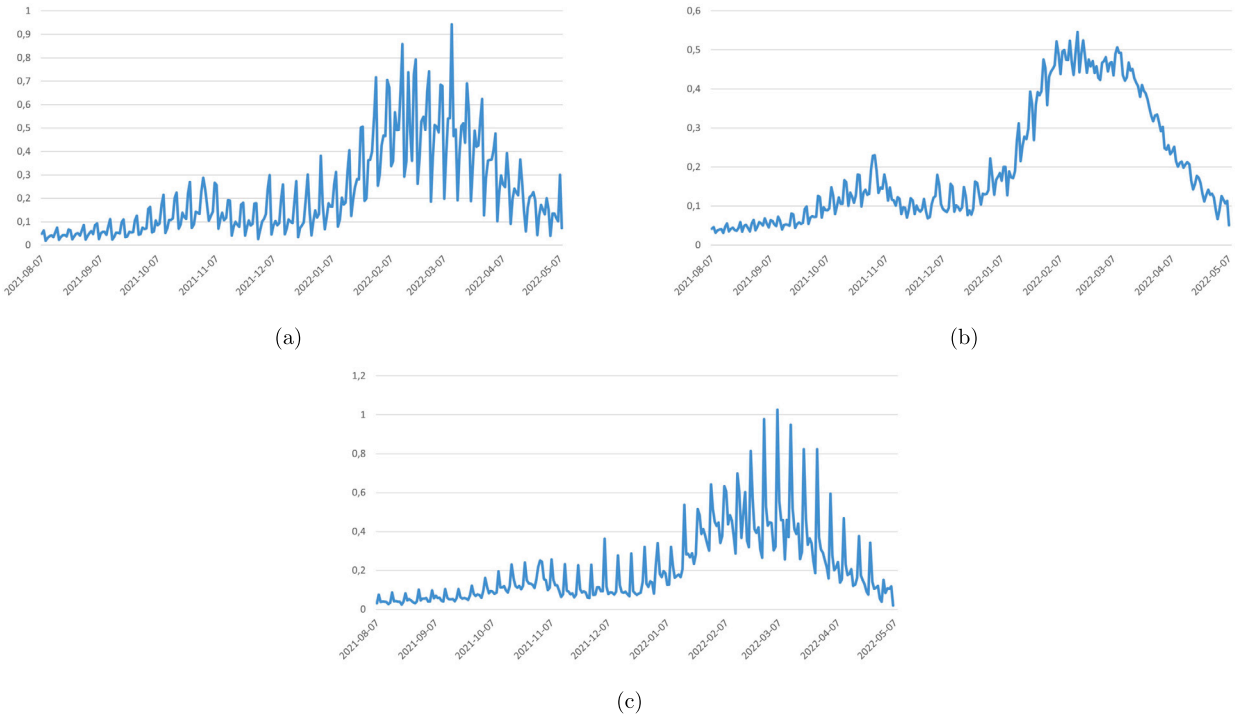

Fig. 3. Share of positive tests in Lithuania: (a) Confirmed cases (new per day) shifted 0 days into the future with respect to tests (new per day); (b) Confirmed cases (new per day) shifted 1 day into the future with respect to tests (new per day); (c) Confirmed cases (new per day) shifted 2 days into the future with respect to tests (new per day).

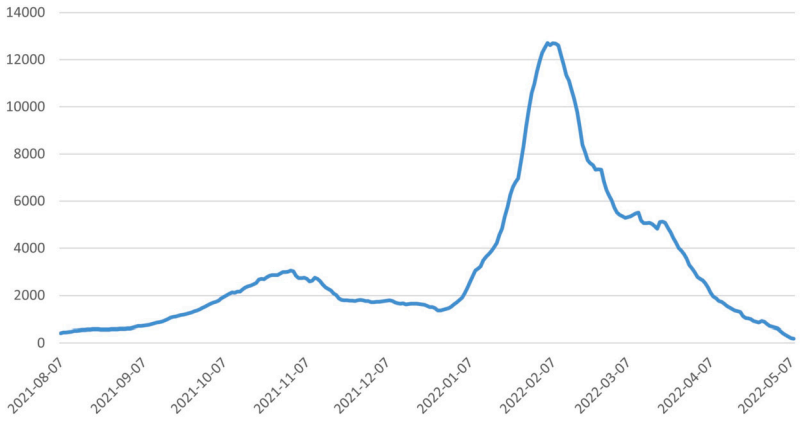

Fig. 4. Confirmed cases (7-day rolling average) in Lithuania.

Here $i = 0, \ldots, T-1$; $S_0 = 1 - I_0$, $I_0 = \mu$, $R_0 = 0$, $D_0 = 0$. The coefficients a, b, and c are the transition probabilities from S to I, I to R, and I to D, respectively; μ is the fraction of the infected population at the starting point in time, which has to be greater than zero for the modeled infection dynamics to propel.

To use this model, one needs to estimate its parameters (i.e., a, b, c, and μ) from the given data for compartments S, I, R, and D (ideally). For this purpose, the maximum likelihood method is one of the most appropriate or preferred methods, as it is characterized by (asymptotically) optimal estimation properties. It is important to note here that the chosen equations of the model can be expressed in terms of the rest, thus bringing the model to a suitable form for working only with the selected compartments (data).

In the paper, we use COVID-19 deaths of the finest available granularity, that is, daily (i.e., the discrete-time index i corresponds to days), to estimate the parameters by the maximum likelihood method. The model-related quantity that corresponds to the *confirmed deaths (new per day)* [36] data is defined as follows (specifically, it is expressed in units per million population and divided by a million):

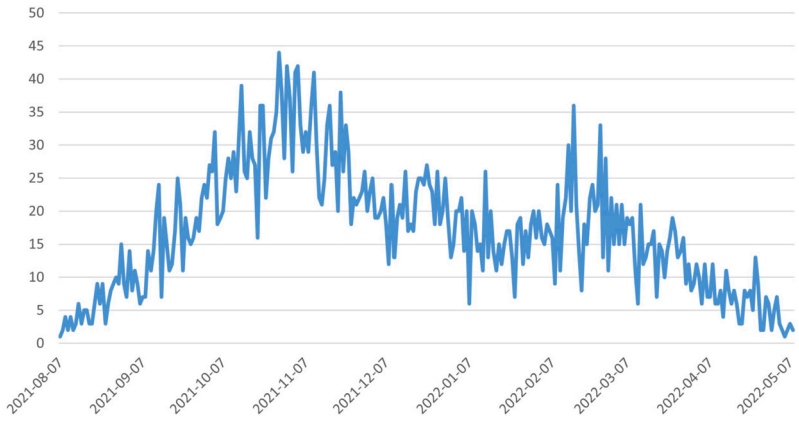

Fig. 5. Confirmed deaths (new per day) in Lithuania.

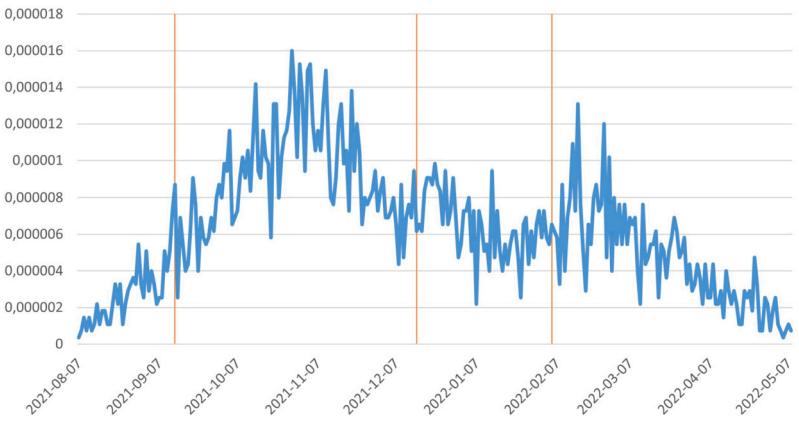

Fig. 6. Confirmed deaths (new per day) per million in Lithuania (divided by a million).

$$\Delta D_i = D_{i+1} - D_i \tag{5}$$

ΔD of Lithuania for the period from 7 August 2021 to 8 May 2022 is shown in Fig. 6.

The period depicted in Fig. 6 broadly covers the delta and omicron waves (the middle vertical dash marks the date when the first omicron cases were detected, i.e., 15 December 2021), and the start (the left vertical dash, which marks the date 13 September 2021) and end (the right vertical dash, which marks the date 5 February 2022) of the use of the COVID-19 passport in Lithuania. For the whole period shown, ΔD is not equal to 0, and for the first days not shown on both extreme sides (i.e., 6 August 2021 and 9 May 2022), ΔD is equal to 0. This period is investigated in Section 3 by the proposed method to study the impact of the COVID-19 passport on the spread of the COVID-19 pandemic in Lithuania.

In terms of the quantity defined by Equation (5), the model defined by Equations (1)–(4) is expressed as follows:

$$\Delta D_0 = \mu c \tag{6}$$

$$\Delta D_1 = \Delta D_0 \left(1 + a\left(1 - \frac{\Delta D_0}{c}\right) - b - c\right) \tag{7}$$

$$\Delta D_{i+1} = \Delta D_i \left(1 + a\left(1 - \frac{\Delta D_i}{c} - \left(\frac{b}{c} - 1\right)\sum_{j=0}^{i-1} \Delta D_j\right) - b - c\right) \tag{8}$$

Here $i = 1, \ldots, T-2$.

In correspondence to the limitation laid out in Section 1 concerning ΔD, we complement it with observational noise of multiplicative Gaussian nature (the normal distribution is widely applied for error modeling [40,41]):

$$Z_i = \Delta D_i \left(1 + \sigma \epsilon_i\right) \tag{9}$$

Here $i = 0, \ldots, T-1$; ϵ_i are observational uncertainties modeled as independent realizations of the standard normal random variable (i.e., having zero mean and unit variance); and σ is the corresponding standard deviation (i.e., the severity of observational noise).

Remark 3. It is evident from design that the stand-alone model defined by Equations (6)–(8) lacks capabilities to fit the data as in Fig. 6 because the model produces smooth curves, while the data is somewhat choppy. However, the addition of Equation (9) is precisely the addition of choppiness to the model, thus aligning it with the requirements posed by the data. It should be noted that even though we call this addition observational noise (as per the state-space modeling design [42]), it can be interpreted as more general noise that encompasses not only the observational noise but also the noise that stems from factors of unknown or not modeled origin.

The model defined by Equations (6)–(9) can be considered the ultimate mathematical object of the proposed method. The following (and final) step consists of the corresponding parameter estimation, for which we employ selected aspects of the estimator recursioning technique [28] and a standard optimization algorithm, thus effectively carrying out the maximum likelihood estimation. We consider the exact mathematical details to be of secondary importance for the message the paper is meant to deliver; hence, we present them only as supplementary material in the form of an application, which can also be used to reproduce results.

The proposed method enables SIRD model applications under the actual data availability situation. It showcases the general way of thought for treating compartmental modeling in the face of data insufficiency problems. However, such an approach generally does not come without its disadvantages either, as the expression of a compartmental model through a subset of compartments introduces a certain level of parameter over-identifiability (e.g., in the SIRD model case, the observed number of deaths can be equally well explained by lower infection rate a but higher mortality rate c or vice versa). Having said that, whether this affects the particular model defined by Equations (6)–(9) remains a question of its likelihood function analysis, which is outside the scope of this paper.

It should be stressed that the analytical analysis of the model defined by Equations (1)–(4) shows that the number of infections increases exponentially at the beginning of an epidemic. Initially, the number of infected people may represent only a tiny proportion of the population. However, the model's assumption that the number of infections is proportional to the number of the susceptible, which is continuously decreasing, implies that the number of the infected (and hence the number of the deceased that are of interest to us here) naturally reaches a peak, followed by a decline. Thus, in order to deal jointly (i.e., with a single likelihood function) with periods covering more than one wave (i.e., with more than one local maximum), it is necessary to employ the time-varying model parameters (fortunately, their implementation is relatively straightforward). Moreover, such parameters are also needed for modeling various phenomena (e.g., enforcing a COVID-19 passport) within a single wave.

3. Results

In this section, we analyze the following phenomena in the case of Lithuania:

1. The impact of the introduction of the COVID-19 passport on the spread of the delta wave;
2. The impact of the termination of the COVID-19 passport on the spread of the omicron wave.

We model the two phenomena independently of each other, using distinct data sets and likelihood functions (alternatively, one can use the joint data set and likelihood function, but the difference is mainly in the technicalities and not the outcomes).

Remark 4. It should be noted that the analysis of phenomena such as those listed above does not suffer from the over-identifiability limitation, if any, described in Section 2. It is because it focuses only on parameter a (i.e., infection intensity) change throughout the modeling horizon, with the rest of the parameters being constant (hence the reason for over-identifiability acting no role here).

To analyze the first phenomenon, we take the period from the start of our data (i.e., 7 August 2021) to 27 December 2021. Here, we assume that the omicron wave started on 28 December 2021. Although the first cases formally appeared on 15 December 2021 (and omicron had likely already been present earlier), it started taking over the delta wave only later. Given that the 7-day-rolling average of confirmed cases was going down to a reading below 500 at that time, indicating that delta was retreating and omicron was still not significant, our assumed omicron start date is the date of renewed upward momentum, which we define as the rise above the 500 threshold. In principle, the moment of omicron appearance can be modeled as a parameter that also requires estimation from data [43], but for simplicity, we treat this moment as particularly defined.

We divide this selected period into the period without the COVID-19 passport (i.e., until 12 September 2021) and the period with the COVID-19 passport (i.e., from 13 September 2021). Since the COVID-19 passport potentially affects only the spread of the virus (modeled by parameter a) and not the recovery and mortality (modeled by parameters b and c, respectively), in this case, the model has only variable parameter a. We apply a rectangular trend to its variation, that is, $a = a_1$ until 12 September 2021 and $a = a_2$ from 13 September 2021 (all the other model parameters remain constant with time), thus assuming that the impact of the COVID-19 passport is instantaneous. This overall modeling setting corresponds to the assumption that the characteristics of the delta wave itself, as well as any significant external factors (except for the COVID-19 passport), do not alter over the analyzed period.

Remark 5. The fact that we use death data to analyze a pandemic control measure mainly related to infection data introduces a specific limitation. For example, it might be the case that we see a significant surge of deaths right after the moment of parameter a change, thus indicating that a_2 may be higher than a_1. However, it is highly likely that the infections corresponding to these increased deaths actually occurred before the moment of parameter a change, making this situation somewhat undeservedly biased toward a_2 being higher than a_1.

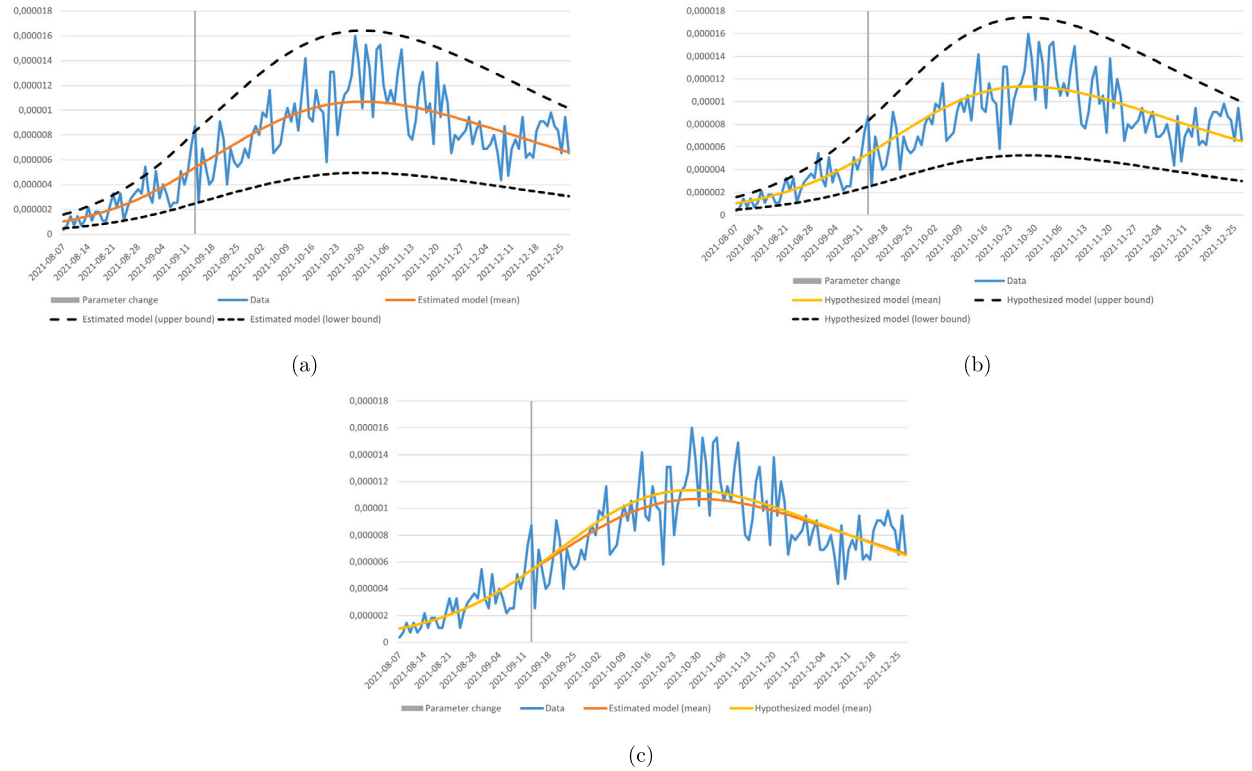

Fig. 7. Modeling of the impact of the introduction of the COVID-19 passport on the spread of the delta wave in Lithuania: (a) Estimated model against data; (b) Hypothesized model against data; (c) Estimated vs. hypothesized model mean.

The corresponding results are provided in Fig. 7. Fig. 7a depicts the estimated model fitted to the data (the model consists of a deterministic part (i.e., mean) and a two-standard-deviation-bounding interval that contains ≈ 95% of values that are due to it). The model is considered a good fit for the data because the mean runs through the middle of the data, and the bounding interval contains most of the data points. Fig. 7b depicts the same as Fig. 7a but for the hypothesized model, which refers to a model applied over the whole period with parameter a_1 of the estimated model. The fact that the hypothesized model is also a good fit for the data suggests that the difference between the estimated a_1 and a_2 is not significant, which means that a_2 is not different from a_1 apart from random chance (Fig. 7c depicts the means of both models, from which one can see that the estimated a_2 is lower than the estimated a_1). Consequently, there is not enough evidence to conclude that the COVID-19 passport impacted the spread of COVID-19 during the delta wave in Lithuania.

Remark 6. The obtained conclusion contrasts that of [12], in which the authors conclude that the COVID-19 passport did curb the spread of COVID-19 during the delta wave in Lithuania. This difference can essentially be attributed to the different accounts of the natural course of the pandemic in the models used. In [12], it is assumed that actions such as the employment of the COVID-19 passport and (or) people's self-awareness were needed for the infection dynamics to ultimately reach a peak and break through to the downside (the authors provide a scenario with none of these actions in place, leading to the exponential growth of infections to infinity), which is a particular overstatement, as these dynamics can first and foremost be explained by the fact that once the critical mass of the population has been through infection, the cases will inevitably start to decrease. Meanwhile, the compartmental model applied in our paper is capable of adhering to this fact by design (see how the hypothesized model in Fig. 7b, which stands for an analogous scenario as that described exponential one of [12], captures the natural dynamics of the pandemic), thus allowing for a fairer comparison of different scenarios (see Fig. 7c).

For the analysis of the second phenomenon, we proceed the same with the part of the data from 28 December 2021 to 8 May 2022, which corresponds to the omicron wave. In this case, the change point of parameter a is 5 February 2022, from which the COVID-19 passport was no longer applicable.

The corresponding results are provided in Fig. 8. Fig. 8a depicts the estimated model fitted to the data, showing a good fit. Fig. 8b depicts the hypothesized model, whose fit is inappropriate. The fact that the hypothesized model fails to fit the data suggests that the difference between the estimated a_1 and a_2 is significant, which means that a_2 is different from a_1 even outside of random chance (Fig. 8c depicts the means of both models, from which one can see that the estimated a_2 is higher than the estimated a_1). However, the obtained result is somewhat ambiguous, as the closer inspection of the data (see Remark 7) hints at a high probability for this

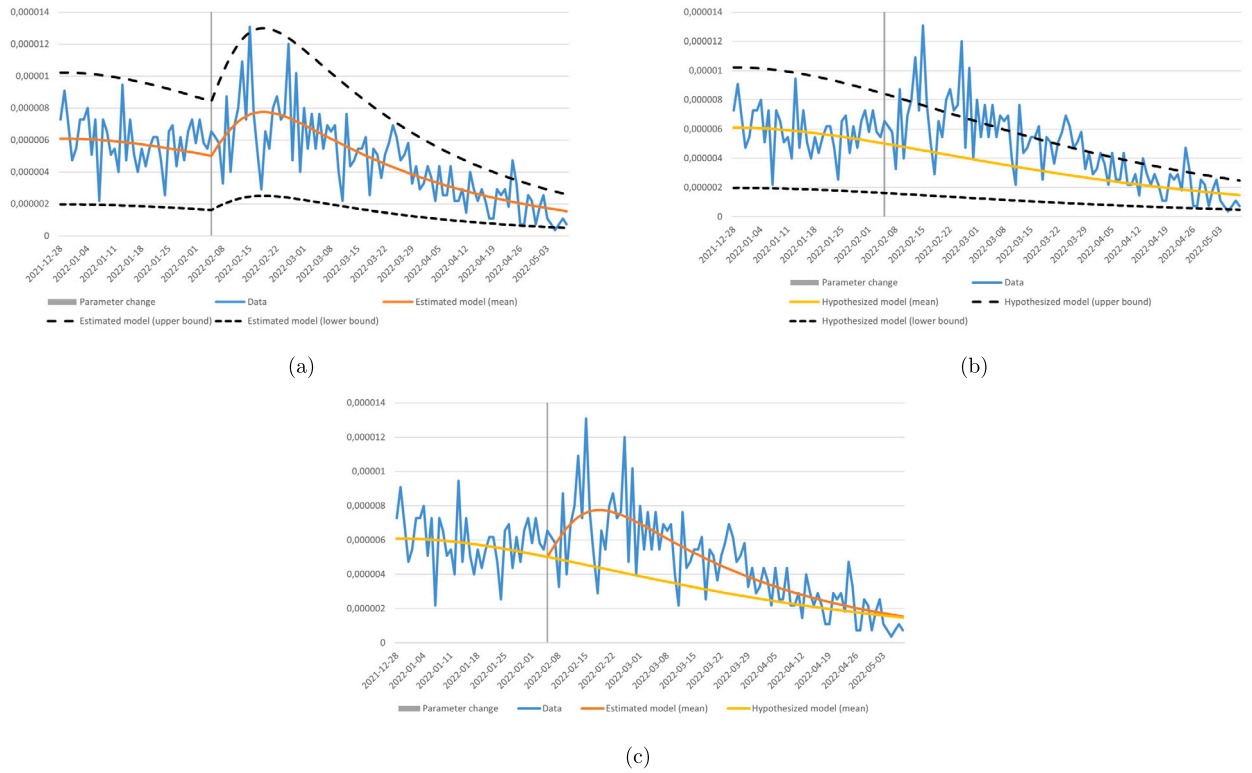

Fig. 8. Modeling of the impact of the termination of the COVID-19 passport on the spread of the omicron wave in Lithuania: (a) Estimated model against data; (b) Hypothesized model against data; (c) Estimated vs. hypothesized model mean.

case analysis to suffer from the limitation laid out in Remark 5, thus requiring further investigation under more comprehensive data before any valid conclusions can be made.

Remark 7. In the omicron wave, the cases started to drop dramatically immediately after the revocation of the COVID-19 passport. However, COVID-19 deaths continued to rise for the next several weeks, as anticipated, considering the lag between infection and death. Those scientists who neglect the natural course of pandemics and consider only infection cases may conclude that the revocation of the COVID-19 passport in Lithuania stopped the pandemic, and those who consider only COVID-19 deaths may conclude that it resulted in the opposite, while both are actually wrong. Curves of both COVID-19 cases and COVID-19 deaths in Lithuania followed the same pattern as they did in countries without COVID-19 passports, such as Norway, Sweden, and Poland.

4. Discussion

The paper highlights the advantages of compartmental modeling for modeling epidemiological phenomena and stresses that officially collected COVID-19 pandemic statistics contain limitations that threaten their full-fledged applications. Minding this, it seeks to encourage stakeholders to question their data and think about room for improvement by offering its share of ideas.

In particular, the paper advocates the maximum likelihood approach to estimating the parameters of a daily-level discrete-time deterministic SIRD model using only COVID-19 deaths, as these data are argued to contain relatively few drawbacks that can be fairly dealt with the acknowledgment of observational noise in the model. The proposed method allows for a reasonable assessment of the core dynamic characteristics of the pandemic (i.e., infection, recovery, and mortality intensities), thus enabling analyses on the effectiveness of various pandemic control measures, among others. It should be noted, however, that while maximum likelihood estimation can hardly be displaced as a method of parameter estimation (because of its fundamentally known properties, which are asymptotically optimal), the same does not hold considering the particular model used in the paper, as it is only one somewhat limited abstraction out of all the possible ones, many of which contain much more detail (in fact, the choice of the most appropriate model for a given research question is a delicate matter that requires profound subject-related considerations so that the correct details can be discerned and employed).

Moreover, the proposed method is concerned with treating compartmental modeling through the perspective of a single compartment, which, at the same time, acts as a workaround in the face of the problematic COVID-19 data setting and is yet vulnerable to limitations that stem from such a practice. To reduce the adverse effects, one must direct efforts toward improvements in data, as no method can ultimately escape their flaws. In this regard, one can consider linking the confirmed case data with that of tests, both of which are currently somewhat mingled. It can be achieved by noting the test date of a confirmed case on an individual level. Such

a procedure basically requires only a change in reporting the currently collected data. It can already improve the opportunities for compartmental modeling of COVID-19, as it is attributable to better discernment of the actual flow going into the Infected compartment. Moreover, a similar linkage can be established between confirmed deaths and confirmed cases, thus allowing for treatment of the limitation noted in Remark 5.

Assuming that the current official data situation is hardly actually to change, the future work should involve the extension of the proposed method to models that offer increased modeling flexibility (e.g., models with stochastic latent part, as this would allow for discernment of the different types of noise present with data, or models with finer-grained compartments and relations between them). Such extensions would be characterized by more complex mathematics that could be handled in the framework of estimator recursioning technique [28]. Moreover, another promising future opportunity pertains to applying the proposed method to modeling based on overall excess mortality (instead of COVID-19 deaths) because factual deaths comprise an objective criterion and expected deaths are a rather well-implied one [44].

Ethics statement

All data used in the study was obtained from public databases; hence, ethics approval and informed consent were not required.

CRediT authorship contribution statement

Leonidas Sakalauskas: Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Vytautas Dulskis:** Writing – original draft, Visualization, Software, Formal analysis, Data curation. **Rimas Jonas Jankunas:** Writing – review & editing, Writing – original draft, Validation, Investigation, Data curation.

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.

Data availability statement

Data is included in supplementary material and referenced in the article.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e31410.

References

[1] W. Gong, S. Parkkila, X. Wu, A. Aspatwar, Sars-cov-2 variants and covid-19 vaccines: current challenges and future strategies, Int. Rev. Immunol. 42 (6) (2023) 393–414.
[2] S. Luthra, S. Agrawal, A. Kumar, M. Sharma, S. Joshi, J. Kumar, Psychological well-being of young adults during covid-19 pandemic: lesson learned and future research agenda, Heliyon 9 (5) (2023) e15841, https://doi.org/10.1016/j.heliyon.2023.e15841.
[3] M. Aleem, M. Sufyan, I. Ameer, M. Mustak, Remote work and the covid-19 pandemic: an artificial intelligence-based topic modeling and a future agenda, J. Bus. Res. 154 (2023) 113303.
[4] L. Rinaldi, Accounting and the covid-19 Pandemic Two Years on: Insights, Gaps, and an Agenda for Future Research, Accounting Forum, vol. 47, Taylor & Francis, 2023, pp. 333–364.
[5] D. Panarello, G. Tassinari, One year of covid-19 in Italy: are containment policies enough to shape the pandemic pattern?, Socio-Econ. Plan. Sci. 79 (2022) 101120.
[6] C. Fan, R. Lee, Y. Yang, A. Mostafavi, Fine-grained data reveal segregated mobility networks and opportunities for local containment of covid-19, Sci. Rep. 11 (1) (2021) 16895.
[7] K. Auranen, M. Shubin, E. Erra, S. Isosomppi, J. Kontto, T. Leino, T. Lukkarinen, Efficacy and effectiveness of case isolation and quarantine during a growing phase of the covid-19 epidemic in Finland, Sci. Rep. 13 (1) (2023) 298.
[8] S. SeyedAlinaghi, A. Karimi, A.M. Afsahi, P. Mirzapour, S. Varshochi, H. Mojdeganlou, P. Mojdeganlou, A. Razi, S. Alilou, M. Dashti, et al., The effectiveness of face masks in preventing covid-19 transmission: a systematic review, Infect. Disord.-Drug Targets 23 (8) (2023) 19–29.
[9] M.P. Walkowiak, J.B. Walkowiak, D. Walkowiak, Covid-19 passport as a factor determining the success of national vaccination campaigns: does it work? The case of Lithuania vs. Poland, Vaccines 9 (12) (2021) 1498.
[10] Z.L. Jiesisibieke, W.-Y. Liu, Y.-P. Yang, C.-W. Chien, T.-H. Tung, Effectiveness and safety of covid-19 vaccinations: an umbrella meta-analysis, Int. J. Public Health 68 (2023) 1605526.
[11] C. Kuhbandner, S. Homburg, H. Walach, S. Hockertz, Was Germany's lockdown in spring 2020 necessary? How bad data quality can turn a simulation into a delusion that shapes the future, Futures 135 (2022) 102879.
[12] M. Stankūnas, A. Džiugys, G. Skarbalius, E. Misiulis, R. Navakas, Evaluating the potential impact of COVID-19 passports in Lithuania, J. Infect. 85 (3) (2022) 334–363.
[13] R. Jankunas, L. Sakalauskas, K. Zamaryte-Sakaviciene, D. Stakisaitis, M. Helmersen, Commentary on the impact of the COVID-19 passports in Lithuania, J. Infect. 86 (3) (2022) e78–e79.
[14] M. Stankūnas, A. Džiugys, G. Skarbalius, E. Misiulis, R. Navakas, Authors' reply to a commentary on the potential impact of COVID-19 passports to epidemiological situation, J. Infect. 87 (3) (2023) e51–e53.
[15] M.J. Fisher, A.P. Marshall, Understanding descriptive statistics, Aust. Crit. Care 22 (2) (2009) 93–97.

[16] D.J. Salkeld, M.F. Antolin, Ecological fallacy and aggregated data: a case study of fried chicken restaurants, obesity and lyme disease, EcoHealth 17 (2020) 4–12.
[17] A.F. Siegenfeld, P.K. Kollepara, Y. Bar-Yam, et al., Modeling complex systems: a case study of compartmental models in epidemiology, Complexity (2022) 2022.
[18] J. Fernández-Villaverde, C.I. Jones, Estimating and simulating a sird model of covid-19 for many countries, states, and cities, J. Econ. Dyn. Control 140 (2022) 104318.
[19] Ö. Özmen, J.J. Nutaro, L.L. Pullum, A. Ramanathan, Analyzing the impact of modeling choices and assumptions in compartmental epidemiological models, Simulation 92 (5) (2016) 459–472.
[20] C. Champagne, B. Cazelles, Comparison of stochastic and deterministic frameworks in Dengue modelling, Math. Biosci. 310 (2019) 1–12.
[21] G.C. Calafiore, C. Novara, C. Possieri, A time-varying sird model for the covid-19 contagion in Italy, Annu. Rev. Control 50 (2020) 361–372.
[22] Y. Zhu, F. Liu, Y. Bai, Z. Zhao, C. Ma, A. Wu, L. Ning, X. Nie, Effectiveness analysis of multiple epidemic prevention measures in the context of covid-19 using the svird model and ensemble Kalman filter, Heliyon 9 (3) (2023) e14231, https://doi.org/10.1016/j.heliyon.2023.e14231.
[23] D.Y. Trejos, J.C. Valverde, E. Venturino, Dynamics of infectious diseases: a review of the main biological aspects and their mathematical translation, Appl. Math. Nonlinear Sci. 7 (1) (2022) 1–26.
[24] P. Crepey, H. Noël, S. Alizon, Challenges for mathematical epidemiological modelling, Anaesth. Crit. Care Pain Med. 41 (2) (2022) 101053.
[25] B. Jahn, S. Friedrich, J. Behnke, J. Engel, U. Garczarek, R. Münnich, M. Pauly, A. Wilhelm, O. Wolkenhauer, M. Zwick, et al., On the role of data, statistics and decisions in a pandemic, AStA Adv. Stat. Anal. 106 (3) (2022) 349–382.
[26] I.J. Myung, Tutorial on maximum likelihood estimation, J. Math. Psychol. 47 (1) (2003) 90–100.
[27] L.S.T. Ho, F.W. Crawford, M.A. Suchard, Direct likelihood-based inference for discretely observed stochastic compartmental models of infectious disease, Ann. Appl. Stat. 12 (3) (2018) 1993–2021, https://doi.org/10.1214/18-AOAS1141.
[28] L. Sakalauskas, V. Dulskis, D. Plikynas, A technique for efficient estimation of dynamic structural equation models: a case study, Struct. Equ. Model. (2024) 1–16, https://doi.org/10.1080/10705511.2023.2282378.
[29] J.M. Mendes, P.S. Coelho, Generalised seir modelling of the covid-19 pandemic course: data quality issues and structural analysis, 2021.
[30] E. Mathieu, H. Ritchie, L. Rodés-Guirao, C. Appel, C. Giattino, J. Hasell, B. Macdonald, S. Dattani, D. Beltekian, E. Ortiz-Ospina, M. Roser, Coronavirus pandemic (covid-19), Our World in Data, https://ourworldindata.org/coronavirus, 2020.
[31] WHO COVID-19 Dashboard - processed by Our World in Data, “New cases”, [dataset], WHO COVID-19 Dashboard [original data].
[32] Official data collated by Our World in Data – processed by Our World in Data, “New tests”, [dataset], Official data collated by Our World in Data [original data].
[33] Official data collated by Our World in Data – processed by Our World in Data, “Positive test rate”, [dataset], Official data collated by Our World in Data [original data].
[34] C. Godske, On the time dependence of smoothed variables, Tellus 18 (4) (1966) 714–721.
[35] Covid dashboards from statistics Lithuania, https://osp.stat.gov.lt/covid-dashboards. (Accessed 29 December 2023).
[36] WHO COVID-19 Dashboard – processed by Our World in Data, “New deaths”, [dataset], WHO COVID-19 Dashboard [original data].
[37] E. Kuhl, Data-driven modeling of covid-19—lessons learned, Extreme Mech. Lett. 40 (2020) 100921.
[38] H. Zhao, N.N. Merchant, A. McNulty, T.A. Radcliff, M.J. Cote, R.S. Fischer, H. Sang, M.G. Ory, Covid-19: short term prediction model using daily incidence data, PLoS ONE 16 (4) (2021) e0250110.
[39] Q. Griette, J. Demongeot, P. Magal, What can we learn from covid-19 data by using epidemic models with unidentified infectious cases? medRxiv, 2021.
[40] C. Zimmer, R. Yaesoubi, T. Cohen, A likelihood approach for real-time calibration of stochastic compartmental epidemic models, PLoS Comput. Biol. 13 (1) (2017) e1005257.
[41] C. Zimmer, S.I. Leuba, T. Cohen, R. Yaesoubi, Accurate quantification of uncertainty in epidemic parameter estimates and predictions using stochastic compartmental models, Stat. Methods Med. Res. 28 (12) (2019) 3591–3608.
[42] M. Aoki, State Space Modeling of Time Series, Springer Science & Business Media, 2013.
[43] R. Bali Swain, X. Lin, F.Y. Wallentin, Covid-19 pandemic waves: identification and interpretation of global data, Heliyon (2024) e25090, https://doi.org/10.1016/j.heliyon.2024.e25090.
[44] T. Beaney, J.M. Clarke, V. Jain, A.K. Golestaneh, G. Lyons, D. Salman, A. Majeed, Excess mortality: the gold standard in measuring the impact of covid-19 worldwide?, J. R. Soc. Med. 113 (9) (2020) 329–334.