Estimating COVID Case Fatality Rate in Bulgaria for 2020–2021

Latchezar Tomov[1(✉)], Hristiana Batselova[2], and Tsvetelina V. Velikova[3]

[1] Department of Informatics, New Bulgarian University, 1618 Sofia, Bulgaria
lptomov@nbu.bg

[2] Department of Epidemiology and Hygiene, University Hospital "Saint George", Medical University, 6000 Plovdiv, Bulgaria
dr_batselova@abv.bg

[3] Department of Clinical Immunology, University Hospital Lozenetz, Sofia University St. Kliment Ohridski, 1407 Sofia, Bulgaria
Tsvelikova@medfac.mu-sofia.bg

Abstract. We estimate the case fatality rate from COVID-19 with our method by age groups for three waves - September 2020 to January 2021 (wild type), February 2021 to May 2021 (alpha), and July 2021 to January 2022 (delta). We use linear regression with optimal lag with 21 days moving averaging to correct for reporting delays. We take the coefficient from the regression as the case fatality ratio. We unite the lower age groups into one to achieve a good correlation. We have new cases by age group and deaths by age group and sex. Our results indicate that the delta variant is more severe than alpha, and this is enough to outweigh any improvements in treatment since the first major wave, 14.08.2020–01.01.2021.

Keywords: COVID-19 · Case fatality rate · Delta wave · Statistics · Linear regression

1 Introduction

1.1 The COVID-19 Global Pandemic

At the end of 2019, the first reports of clusters of pneumonia with unknown etiology appeared in Wuhan City, Hubei Province of China. Many of the patients were linked with a wet seafood market where other wildlife species were also sold. An unknown virus was isolated from all the patients, and molecular analysis has proven that the pathogen was a new coronavirus (CoV). The scientist first named it 2019-nCoV, and the name of the disease was given by WHO- COVID-19. The virus was widespread and, from an outbreak in Wuhan, turned into a public health emergency of international concern [1].

For decades Coronaviruses have been known to be a high pandemic risk. SARS-CoV-2 is the ninth coronavirus known to be a pathogen to humans and the seventh identified in the last 20 years [2]. All human coronaviruses are proven to be with zoonotic origins. The emergence of SARS-CoV-2 had many similarities to SARS-CoV that appeared

T. Zlateva and R. Goleva (Eds.): CSECS 2022, LNICST 450, pp. 102–115, 2022.
https://doi.org/10.1007/978-3-031-17292-2_9

among humans in Foshan, Guangdong province, China, in November 2002 and again in Guangzhou, Guangdong province, in 2003 [3].

All of these SARS-CoV emergence cases among humans were linked to markets where live animals were sold, particularly civets and raccoon dogs [4], and were also sold live in Wuhan markets in 2019 [5] and are known to be a host of SARS-CoV-2 [6, 7].

COVID-19 is an infectious respiratory illness caused by the severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) [8]. Since May 2022 WHO reported 512,607,587 confirmed cases of COVID-19, including 6,243,038 deaths [9]. The median incubation period of COVID-19 ranges from 5 to 6 days [10].

COVID-19 can affect any age group, and all humans are susceptible. However, the risk for severe disease is higher in the elderly and patients with comorbidities such as cardiovascular disease, diabetes mellitus, chronic respiratory disease, hypertension, and cancer can experience worse outcomes [11]. Therefore, the consequences of the infection depend on the interplay between the viral and immune mechanisms [10].

1.2 The COVID-19 Pandemic in Bulgaria

The first person to test positive for SARS-2-COV was on 08.03.2020 [12]. The Bulgarian government acted swiftly before the clusters transitioned into diffusion, thus stopping the first wave [13]. The first wave started in the middle of August - there was a transient spike in June and July due to the ending of lockdown and temporarily removing all measures on the 15$^{\text{th}}$ of June [14], coupled with holiday migrations from the big cities towards the countryside and the Black sea coastline. The seasonal autumn wave accelerated when schools opened with very few measures and was reversed after the gradual transition towards online education in schools [15]. The second wave started on 01.01.2021 and was driven by a new variant, called “alpha”, which soon became dominant. The third wave was driven by the Delta variant and infected and killed the most people – almost 40% of all deaths from these three waves are from the third one – the three large waves took approximately 8000,10000 and 12000 each with 200 000,220 000 and over 300 000 infected. A crude estimate of overall mortality risk shows 4%, 4.5% and 4%, which could lead to an incorrect assessment of risks for the waves. That’s why we need a better estimation of the Case Fatality Rate (CFR, %), separate for all waves and age groups. Furthermore, different age groups have different average lags between infection and death, which introduces variance in an estimation method, that only incorporates single lag.

2 Methods

We have data from the open portal for new cases by age groups since 06.06.2020 [16] and data for new deaths by age groups since 09.03.2020 [16, 17]. The data between 09.03.2020 and 5.6.2020 is gathered and prepared manually from the coronavirus website by Ralitza Ilieva-Markova, our assistant.

Several different methods for estimating case fatality rate (CFR,%) exist. Possibly the simplest is to get the ratio between deaths and the new cases with some lag and obtain a function with varying values at varying moments of time. Its mean value could be used for the case fatality ratio [18]. Another approach is to use a linear or polynomial regression [19]. This crude case fatality ratio has some limitations - it does not disseminate by risk factors such as age, sex, and comorbidities. For such purposes, a much more sophisticated method is employed in various risk calculators in the UK, for example [20]. Time series analysis is also a viable option that can be scaled for more detailed or limited data [21]. These models are useful for predicting actual mortality during a particular wave but do not give a single value for CFR by age group. A single value is more valuable in communication with authorities when discussing policies to prevent or mitigate a pandemic and its subsequent medium and long term effects on the public health an the public health system.

Furthermore, the adjustment of lags of predictors is not an automated procedure. We need an estimate for CFR by age for the different waves, driven by different conditions such as different infected populations, different variants, etc. This way, we can compare the variants and their "burden" in the first approximation. Our data for Bulgaria is very limited - no detailed info for comorbidities of the deceased, no sex for the infected, comorbidities of infected, no seroprevalence study, etc. We do not need a complicated method for CFR analysis, so we use simple linear regression. What is new is our algorithm, because we identified waves by objective markers such as the effective reproductive number and/or changepoint analysis. We also used the sample cross-correlation function to extract the optimal lag (with the highest correlation between new cases and deaths) for the deceased. We used the coefficient of the linear regression for our case fatality ratio. We applied a longer moving average with a 21-day-period both to the new cases by age group and the deceased by age group to correct for noise from random delays and periodicity in reporting on a weekly basis. This is the largest possible n-weekly period that does not alter the dynamics of the processes. We use the Robert Koch formula for the effective reproductive number R_t [22] - we don't need a more complicated method for our goals. We use the non-commercial software GNU Octave for our purpose.

The defining criteria for waves are formed from a minimum value of calculated R_t in proximity or coinciding with changepoints. Analysis of regime change from [13]-Fig. 1. This is not yet fully automated criteria since there are multiple local extrema of calculated R_t (Fig. 2) in the period between two waves and, in some cases, more than one changepoint. However, this could be automated as *"the last local minima in R_t in the proximity of changepoint before a prolonged period of R_t > 1"* (the length of the period being at least twice longer than the longest period before two successive local extrema in the "in-between" zone, also the "proximity" needs to be quantified). The three waves according to these criteria are Wave 1: 14.08.2020– 01.01.2021, Wave 2: 02.01.2021–30.06.2021, and Wave 3: 01.07.2021–27.12.2022.

3 Results and Discussion

In the period 2020–2021, there were three major waves of this pandemic, attacking different subpopulations with different age and comorbidities. Although we do not have access to the data for comorbidities, they strongly correlate with age. We believe that estimating the case fatality rate has some meaning, especially for pandemic control policies and healthcare management, due to the exponential nature of the hospitalization risk and the different CFR in different waves. There are more suitable models to predict deaths from new cases, and we explored some of them in previous publications [21]. However, having a single number per age group is useful in forming policies and communicating risks towards the general population and the responsible government officials in charge of these policies.

The results from our almost completely automated procedure are shown for the three waves in Table 1, 2 and 3, respectively, and in Figs. 3, 4 and 5 (Wave 1), Figs. 6, 7 and 8 (Wave 2), Figs 9, 10 and 11 (Wave 3). In Tables 1, 2 and 3, we give the estimated CFR for 10 different models - the linear regression coefficient for each age group, their p values, their t-stats and the RMSE of the linear regressions, showing how close are the actual death counts to the model fit. We give two of the models in Fig. 13–14 as examples – the fit for age group 70–79 for the first wave an 90+ for the third wave. The comparisons of the three waves (Fig. 12) show several interesting results:

- The risk of death exponentially increases with age up to 70 years of age, after which there is saturation. This could have several different explanations, one of which is the lower probability of surviving 80+ years of age with many comorbidities and the lack of control for such in our research due to lack of data.
- The delays between new cases and deaths from wave 2, driven mostly by variant alpha, are decreased compared to the first wave, driven by the wild-type virus. This indicates increased severity and corresponds to anecdotal evidence from doctors in Bulgaria's COVID wards and research [23].
- Adjusted R^2 for Wave 2 is smaller than in the other waves, reflecting more dispersion around the maximum correlation – the maximum correlation with the optimal lag is not as high due to substantial contributions from other lags – or the deaths are more spread out in time than in the other waves. A single number here captures the mortality risk a bit worse than in the other waves.
- Despite the increased severity of alpha, the estimated case fatality rate is substantially lower than in Wave 1. This could be explained by the substantial increase in testing (this is case fatality rate, not infection fatality rate) and/or by improvements in treatment, but also partially with the smaller correlations in model estimation (hidden mortality risks)
- Delay for children is substantially higher, and mortality risks are considerably lower. However, a good model fit was successful only for the first wave. This could also be due to a change in reporting of deaths (listing the comorbidities as a primary cause of death).
- Delta showed CFR, similar to the first wave despite increased testing and improved treatment protocol, but early signs of an overburdened hospital system [24]. There were also substantial reinfections that work towards decreased CFR [25]. The Delta wave also happened in a partially vaccinated population [26]. This indicates significantly increased severity of the Delta variant, which as a hypothesis needs additional research to support it.

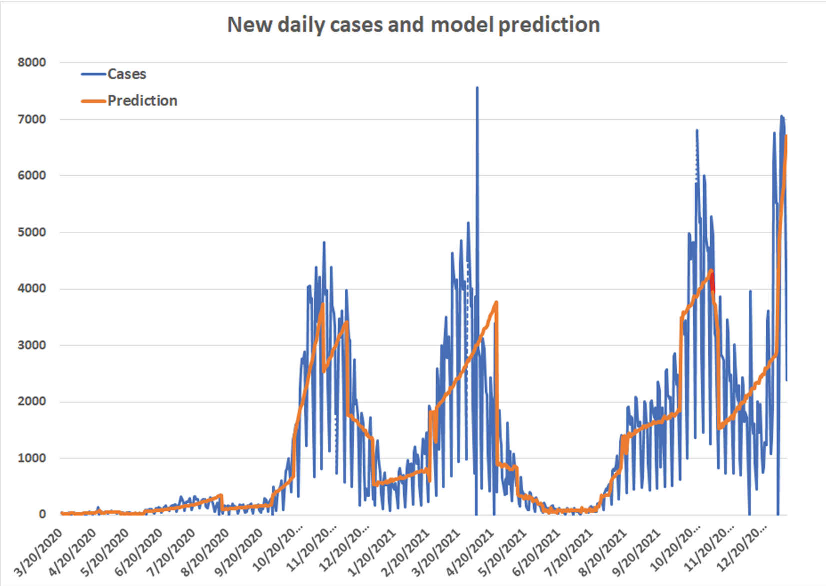

Fig. 1. Model prediction of COVID-19 daily cases with automatic detection of regime changes and stochastic modeling of trajectories [13]. The points of discontinuity refer to regime changes.

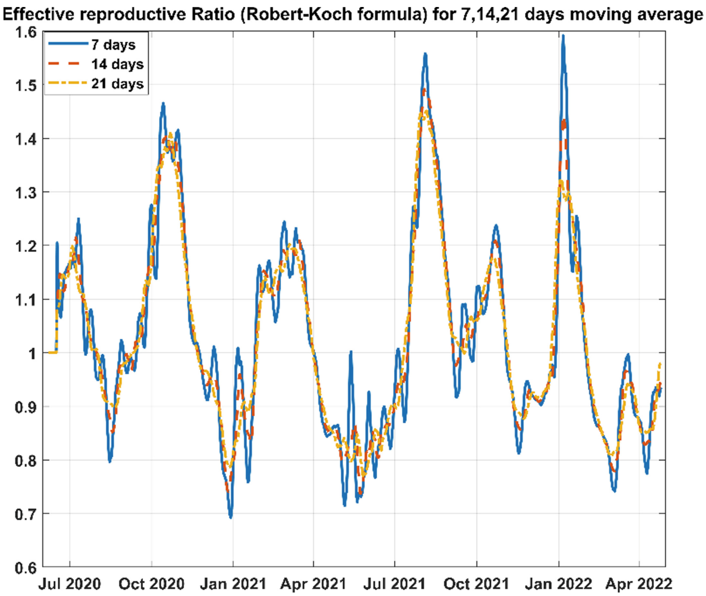

Fig. 2. Effective reproduction ratio, estimated by the formula of Robert Koch institute with 4 days of the incubation period for moving average of daily cases with periods of 7,14 and 21 days.

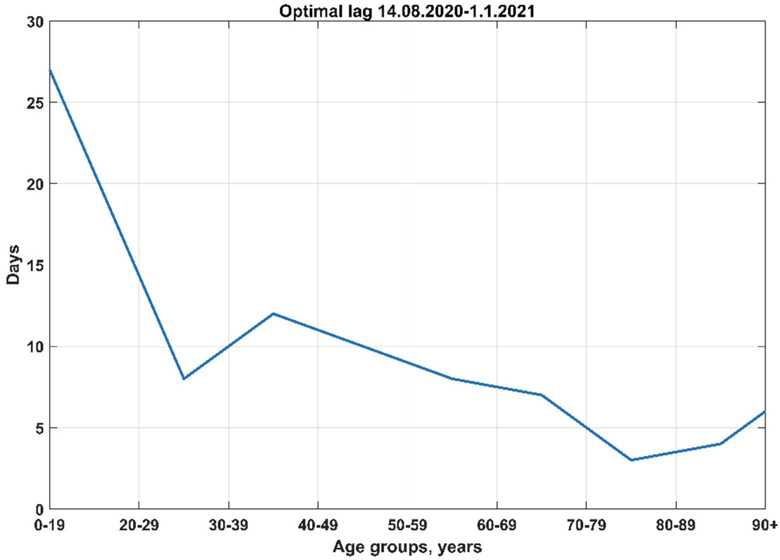

Fig. 3. Optimal lag of sample cross-correlation between new cases and deaths per age group for Wave 1.

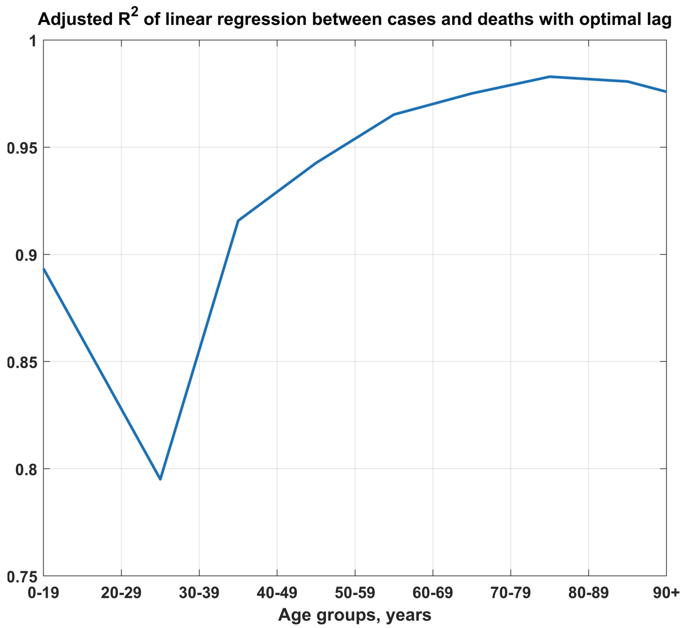

Fig. 4. Adjusted R^2 of the linear regression between new cases and deaths with optimal lag per age group for Wave 1.

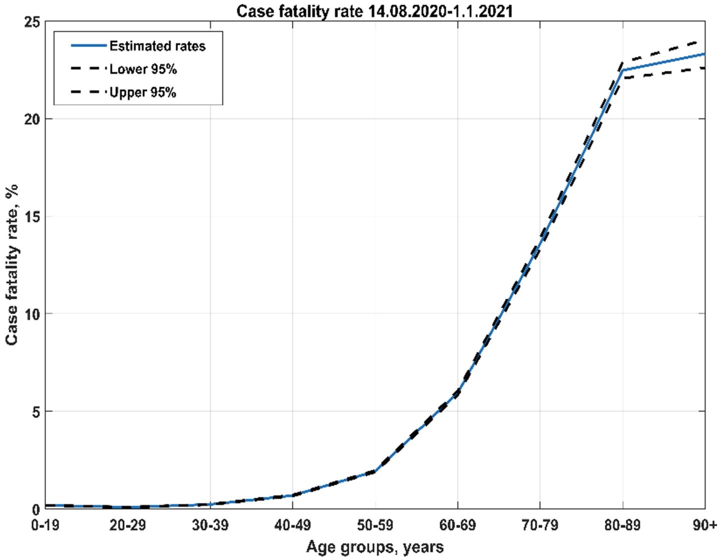

Fig. 5. Estimated case fatality rate and 95% confidence intervals as the coefficient from linear regression with optimal lag per age group for Wave 1.

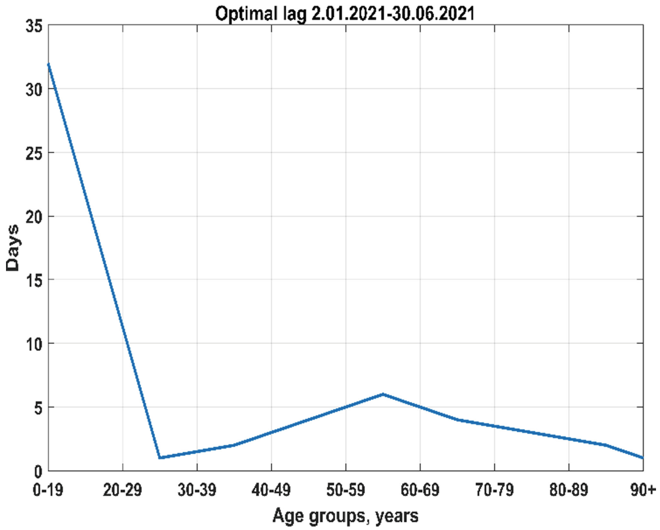

Fig. 6. Optimal lag of sample cross-correlation between new cases and deaths per age group for Wave 2.

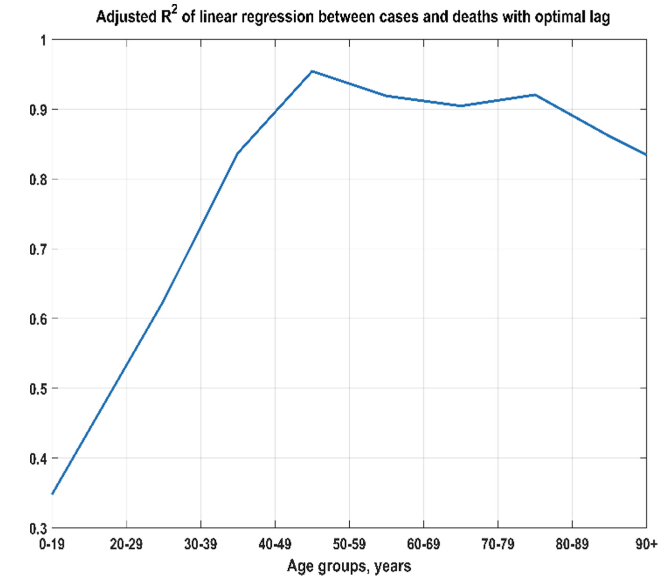

Fig. 7. Adjusted R^2 of the linear regression between new cases and deaths with optimal lag per age group for Wave 2.

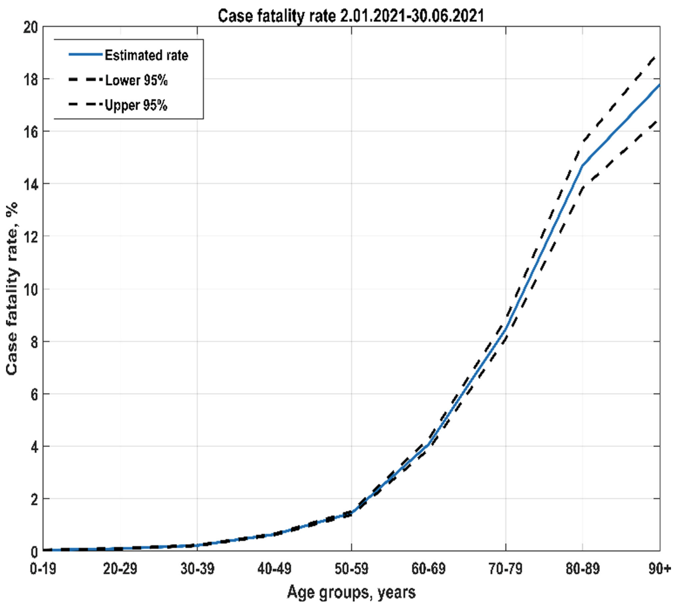

Fig. 8. Estimated case fatality rate and 95% confidence intervals as the coefficient from linear regression with optimal lag per age group for Wave 2.

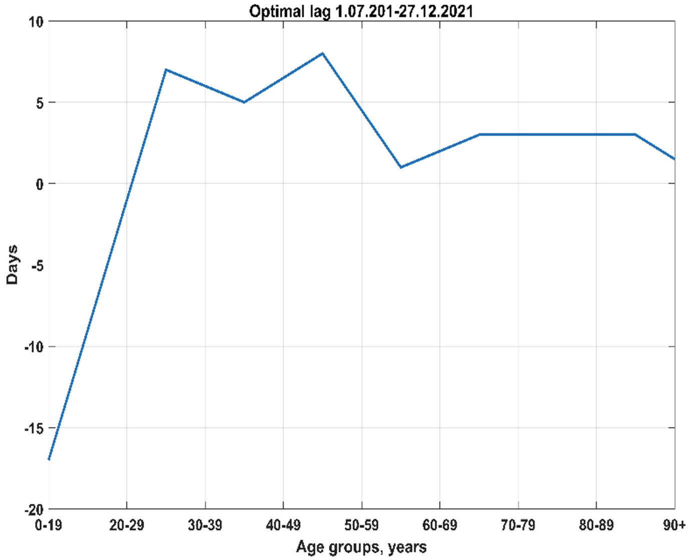

Fig. 9. Optimal lag of sample cross-correlation between new cases and deaths per age group for Wave 3.

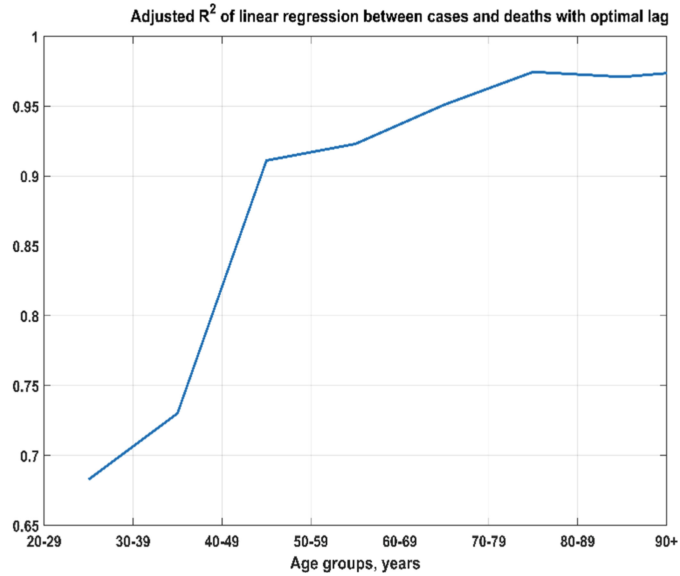

Fig. 10. Adjusted R^2 of the linear regression between new cases and deaths with optimal lag per age group for Wave 3.

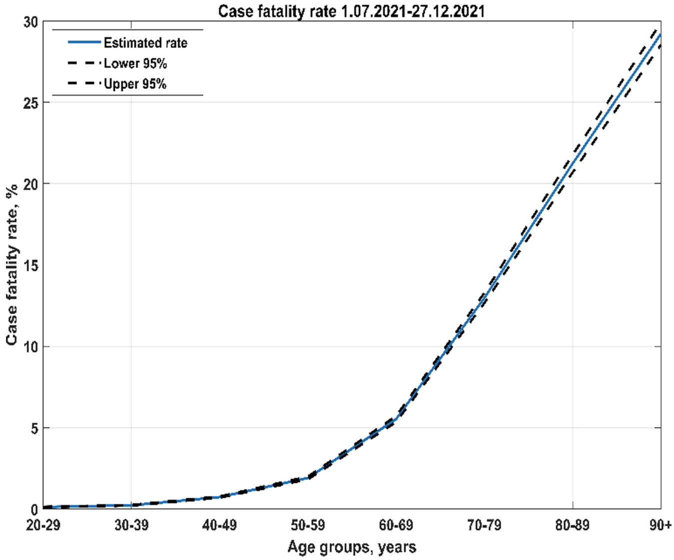

Fig. 11. Estimated case fatality rate and 95% confidence intervals as the coefficient from linear regression with optimal lag per age group for Wave 2.

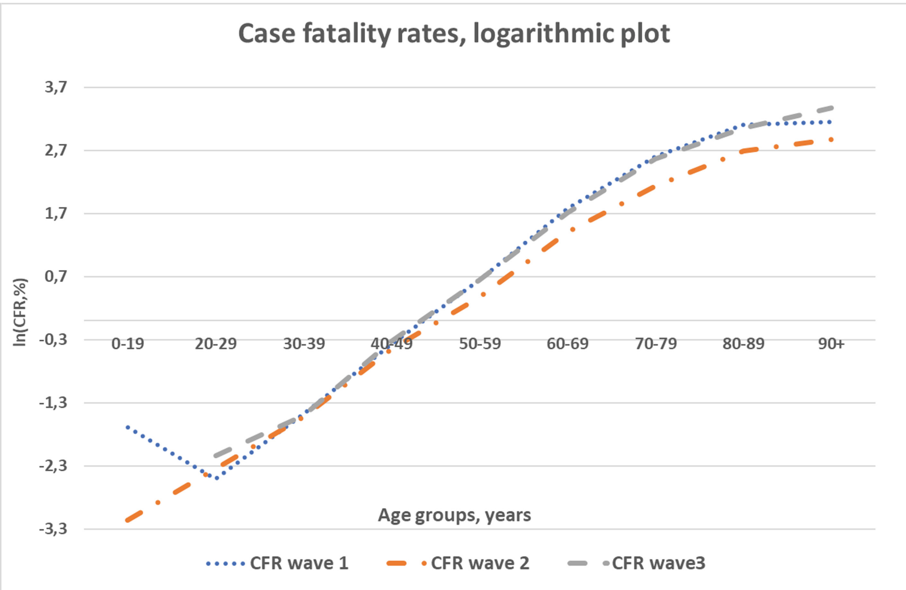

Fig. 12. Comparison of the estimated case fatality rates for the three waves (logarithmic plot).

Table 1. Estimated CFR and model parameters for Wave 1

Age group	CFR,%	T-stat	P value	RMSE
0–19	0.185343422487180	38.9110979328123	1.30817035064048e−76	0.021732
20–29	0.0814640663243919	18.2326275468917	1.09988316471984e−38	0.055223
30–39	0.230159017738559	38.7935582526591	1.92310126757591e−76	0.144791
40–49	0.689188248236168	53.6055189594344	9.84784198059810e−95	0.379568
50–59	1.93038503082146	92.9271625360074	5.23345933982721e−127	0.633749
60–69	5.93965884283796	105.184425073305	2.20675477239857e−134	1.452484
70–79	13.5895553151869	92.9362143877190	5.16418027007381e−127	2.218877
80–89	22.4676792760740	108.672603974902	2.49925803109170e−136	1.061962
90+	23.3160356243301	64.6250278180166	1.38759960707132e−105	0.167789

CFR: …Case Fatality Rate……….; RMSE: ……Root Mean Square Error…….

Table 2. Estimated CFR and model parameters for Wave 2.

Age group	CFR,%	T-stat	P value	RMSE
0–19	0.0422356296567256	9.79693481057488	2.22851007371688e−18	0.021509
20–29	0.0956918315236564	17.2611764374006	8.70168515847137e−40	0.063986
30–39	0.217996100910862	30.2424115433259	7.80149815693797e−72	0.151505
40–49	0.638694058840290	60.9730277484419	9.46038326682465e−121	0.277128
50–59	1.44984967745342	44.9122024954124	1.09579425737544e−98	0.899099
60–69	4.07056706902815	41.0765543261353	2.01408414848497e−92	2.681178
70–79	8.45935083431752	45.4107576525972	1.81346937173692e−99	3.646795
80–89	14.7076759170350	33.2386618263124	5.15824052965261e−78	2.81023
90+	17.7741156731515	27.3619701862002	1.72506047259888e−65	0.32967

CFR: …Case Fatality Rate……….; RMSE: ……Root Mean Square Error…….

Table 3. Estimated CFR and model parameters for Wave 3.

Age group	CFR,%	T-stat	P value	RMSE
0–19	-	-	-	-
20–29	0.118384398162955	19.6531966249049	1.83592637752605e−46	0.094262
30–39	0.223735563220952	22.0294355443674	9.96578783625195e−53	0.281782
40–49	0.729399256401650	42.8391931708295	1.09018475772862e−95	0.556642
50–59	1.90537446158302	46.3166942774476	3.14992631355092e−101	1.164027
60–69	5.52926275243861	58.9483408966994	9.98593754990433e−119	2.371468
70–79	12.9733371591063	82.6712913696373	6.11083999904439e−144	3.00583
80–89	21.2299468812402	77.3111919608126	6.73558995179243e−139	1.906377
90+	29.2022949933292	85.5649458137322	1.55419014993249e−146	0.210967

CFR: …Case Fatality Rate……….; RMSE: ……Root Mean Square Error…….

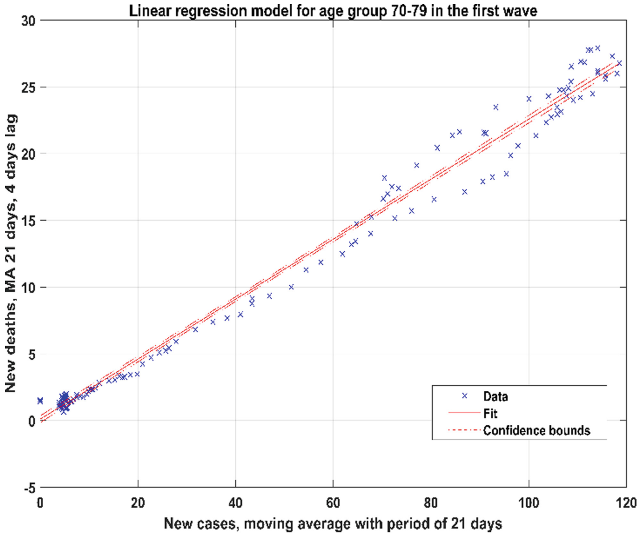

Fig. 13. Example of linear regression model, generated by our algorithm – linear regression for age group 70–79 years old with lag of 4 days.

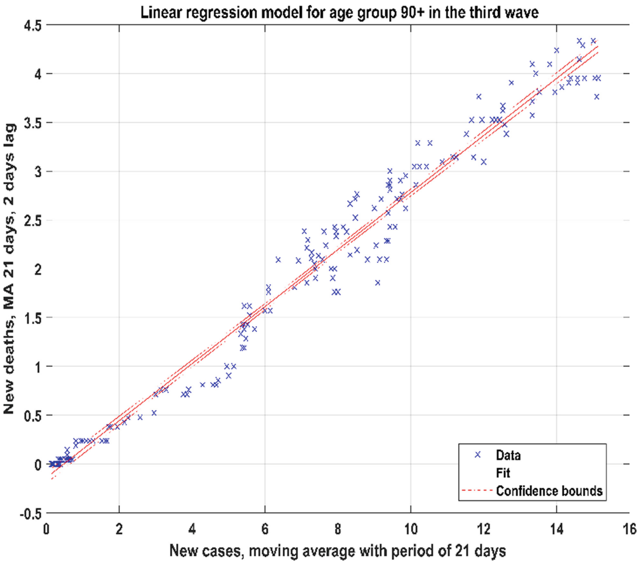

Fig. 14. Example of linear regression model, generated by our algorithm – linear regression for age group 90+ years old.

4 Conclusion

The goal of this research is to compare case fatality rates of different covid-19 waves by age groups in order to obtain first approximation of the burden of different waves in terms of lethality. There are other models, that the authors use to predict deaths by age groups, and to predict excess mortality, but they are not easy to communicate to authorities – there is need of single number for CFR instead of complicated model, to compare visually the different waves and to assist in development of new policies. For this reason, a broader spectrum of models, methods and algorithms needs to be employed, to deliver the messages with the adequate level of complexity to policymakers. This method is not new, its implementation is novel – using optimum lags for different age groups from cross-correlation analysis, and application of larger periods for smoothing of data due to large delays of reporting of deaths an weekly periodicity. More complex models (part of ongoing research) are used to detect the start and the end of different waves.

Our results indicate that the delta variant is more severe than alpha, which is enough to outweigh any improvements in treatment since the first significant wave. The crude methods of estimation give higher burden of the alpha wave, because they do not adjust by age. The larger absolute number of deaths during the alpha wave was from the significant increase of the proportion of people over 60 years old. Furthermore, low vaccination coverage and lack of effective measurements led to more significant waves. Additionally, we confirmed that the CFR is valuable parameter to evaluate, follow up and compare waves by age groups.

References

1. Sun, J., et al.: COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. Trends Mol. Med. **26**(5), 483–495 (2020). ISSN 1471-4914
2. Lednicky, A., et al.: Emergence of porcine delta-coronavirus pathogenic infections among children in Haiti through independent zoonoses and convergent evolution. medRxiv (2021). https://doi.org/10.1101/2021.03.19.21253391
3. Xu, R.H., et al.: Epidemiologic clues to SARS origin in ChinaEmerg. Infect. Dis. **10**, 1030–1037 (2004)
4. Guan, Y., et al.: Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science **302**, 276–278 (2003)
5. Close, R.H., et al.: Epidemiologic clues to SARS origin in China. Emerg. Infect. Dis. **10**, 1030–1037 (2004)
6. Freuling, C.M., et al.: Susceptibility of raccoon dogs for experimental SARS-CoV-2 infection Emerg. Infect. Dis. **26**, 2982–2985 (2020)
7. Edward, C., et al.: The origins of SARS-CoV-2: a critical review. Cell **184**(19), 4848–4856 (2021)
8. Dhar Chowdhury, S., Oommen, A.M.: Epidemiology of COVID-19. J. Dig. Endosc. **11**(1), 3–7 (2020). https://doi.org/10.1055/s-0040-1712187
9. https://covid19.who.int. WHO Coronavirus (COVID-19) Dashboard. Accessed 06 May 2022
10. Linton, N.M., Kobayashi, T., Yang, Y.: Incubation period and other epidemiological characteristics of 2019 novel Coronavirus infections with right truncation: a statistical analysis of publicly available case data. J Clin Med. **9**(02), E538–E538 (2020)
11. Wu, Z., McGoogan, J.M.: Characteristics of and important lessons from the Coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020;(e-pub ahead of print). doi https://doi.org/10.1001/jama.2020.2648

12. https://bnr.bg/post/101237668/potvardeni-sa-dva-sluchaa-na-koronavirus-v-balggaria. Accessed 06 May 2022
13. Tchorbadjieff, A., Tomov, L.P., Velev, V., Dezhov, G. Manev, V., Mayster, P.: On regime changes of Covid-19 outbreak. J. Appl. Stat. (2020, under review)
14. https://news.lex.bg/%D0%BF%D1%80%D0%B5%D0%BC%D0%B8%D0%B5%D1%80%D1%8A%D1%82-%D1%81%D0%BB%D0%B5%D0%B4-15-%D1%8E%D0%BD%D0%B8-%D0%BE%D1%82%D0%BF%D0%B0%D0%B4%D0%B0%D1%82-%D0%B2%D1%81%D0%B8%D1%87%D0%BA%D0%B8-%D0%BC%D0%B5/. Accessed 06 May 2022
15. Tomov, L., Lazova, S., Velikova, T.S., Batselova, Hr.: Epidemiology of COVID-19 spread in Bulgaria. In: National Cardiologic Conference on Covid-19, Bulgaria, 19–20th November (2021). https://doi.org/10.13140/RG.2.2.12061.18403
16. https://data.egov.bg/data/resourceView/8f62cfcf-a979-46d4-8317-4e1ab9cbd6a8. Accessed 06 May 2022
17. http://coronavirus.bg. Accessed 06 May 2022
18. Baud, D., Qi, X., Nielsen-Saines, K., Musso, D., Pomar, L., Favre, G.: Real estimates of mortality following COVID-19 infection. Lancet Infect Dis. **20**(7), 773 (2020) https://doi.org/10.1016/S1473-3099(20)30195-X. PMID: 32171390; PMCID: PMC7118515
19. Suleiman, A.A., Suleiman, A., Abdullahi, U.A., Suleiman, S.A.: Estimation of the case fatality rate of COVID-19 epidemiological data in Nigeria using statistical regression analysis. Biosaf. Health, 3(1), 4–7 (2021), Öztoprak et al., COVID-19: Turkey's epidemiology; March 2020.https://doi.org/10.14744/ejmo.2020.60998
20. Hippisley-Cox, J., Coupland, C.A., Mehta, N., Keogh, R.H., Diaz-Ordaz, K., Khunti, K., et al.: Risk prediction of covid-19 related death and hospital admission in adults after covid-19 vaccination: national prospective cohort study. BMJ, **374**, n2244 (2021) https://doi.org/10.1136/bmj.n2244
21. Tomov, L., Tchorbadjieff, A., Angelov, S.: Age-specific mortality risk from Covid-19 in Bulgaria. Comput. Sci. Educ. Comput. Sci. (2021, accepted)
22. Georg, Q.: The reproduction number and its measurement. A critique of the Robert Koch Institute. (2020). https://doi.org/10.13140/RG.2.2.29012.83849
23. Grint, D.J., et al.: Severity of SARS-CoV-2 alpha variant (B.1.1.7) in England. Clin Infect Dis. (2021). Ciab754
24. Tomov, L.P., Batselova, H.M., Velikova, T.V.: COVID-19 Delta wave caused early overburden of hospital capacity in the Bulgarian healthcare system in 2021. Healthcare **10**(4), 600 (2022). https://doi.org/10.3390/healthcare10040600
25. Marinov, G.K., Mladenov, M., Rangachev, A., Alexiev, I.: SARS-CoV-2 reinfections during the first three major COVID-19 waves in Bulgaria. https://doi.org/10.1101/2022.03.11.22271527
26. https://vaccinetracker.ecdc.europa.eu/public/extensions/covid-19/vaccine-tracker.html#uptake-tab. Accessed 06 May 2022