



OPEN

## Excess mortality in Belarus during the COVID-19 pandemic as the case study of a country with limited non-pharmaceutical interventions and limited reporting

Alexander Kirpich<sup>1✉</sup>, Aleksandr Shishkin<sup>1</sup>, Thomas A. Weppelmann<sup>2</sup>, Alexander Perez Tchernov<sup>3</sup>, Pavel Skums<sup>4</sup> & Yuriy Gankin<sup>5</sup>

Public health intervention to contain the ongoing COVID-19 pandemic significantly differed by country since the SARS-CoV-2 spread varied regionally in time and in scale. Since vaccinations were not available until the end of 2020 non-pharmaceutical interventions (NPIs) remained the only strategies to mitigate the pandemic spread at that time. Belarus in Europe is one of a few countries with a high Human Development Index where no lockdowns have ever been implemented and only limited NPIs have taken place for a period of time. Therefore, the Belarusian case was evaluated and compared in terms of the mortality burden. Since the COVID-19 mortality was low, the excess *overall* mortality was studied for Belarus. Since no overall mortality data have been reported past June 2020 the analysis was complemented by the study of Google Trends funeral-related search queries up until August 2021. Depending on the model, the Belarusian mortality for June of 2020 was 29 to 39% higher than otherwise expected with the corresponding estimated excess death was from 2953 to 3690 while the reported COVID-19 mortality for June 2020 was only 157 cases. The Belarusian excess mortality for June 2020 was higher than for all neighboring countries with an excess of 5% for Poland, 5% for Ukraine, 8% for Russia, 11% for Lithuania and 11% for Latvia. The relationship between Google Trends and mortality time series was studied using Granger's test and the results were statistically significant. The results for Google Trends searches did vary by key phrase with the largest excess of 138% for April 2020 and 148% for September 2020 was observed for a key phrase "coffin", while the largest excess of 218% for January 2021 was observed for "funeral services". In summary, there are indications of the excess overall mortality in Belarus, which is larger than the reported COVID-19-related mortality.

The novel coronavirus SARS-CoV-2 was originally detected in December 2019 in Wuhan, China<sup>1–5</sup> and within just three months<sup>6</sup> rapidly spread around the world. The corresponding public health interventions to curb the pandemic, however, significantly differed between regions<sup>7</sup>. This happened since the spread of SARS-CoV-2 varied in time and scale as well as the overall preparedness of healthcare systems and hospitals that could be utilized simultaneously<sup>8</sup>. The large-scale vaccination interventions were not available until the end of 2020 due to time required for vaccine development and licensing<sup>9–17</sup>. As a result mandatory non-pharmaceutical interventions (NPIs) remained the only strategies to control the pandemic. The NPIs included multiple measures ranging from basic precautions and some social distancing to tracing of all infected individuals together with their contacts and complete lockdowns of non-essential public services and business activities.

Belarus in Eastern Europe is one of a few countries with a high Human Development Index (vhHDI)<sup>18</sup> where no lockdowns have ever been implemented and only limited NPIs took place<sup>7</sup>. The available NPIs included some contact tracing, 14-days self-isolation for close contacts of laboratory-confirmed cases and for those returning from abroad, delay in class starting times for schools, remote learning for some universities, and increased

<sup>1</sup>Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, GA, USA. <sup>2</sup>Department of Internal Medicine, University of South Florida, Tampa, FL, USA. <sup>3</sup>Faculty of Mechanics and Mathematics, Belarusian State University, Minsk, Belarus. <sup>4</sup>Department of Computer Science, Georgia State University, Atlanta, GA, USA. <sup>5</sup>Quantori, Cambridge, MA, USA. ✉email: akirpich@gsu.edu

frequency of public transportation to avoid crowding<sup>19</sup>. Since the social distancing measures were not strictly implemented or enforced, it remained up to individuals whether and how to change their behavior patterns<sup>20</sup>. In particular, the mask regimen was recommended but not evaluated for compliance<sup>21</sup>. Businesses and public services which required personal interactions, such as restaurants, gyms, barber shops, taxis, and theaters always remained fully operational. The mask regimen was not mandatory until November of 2020 and no country exit and entry limitations were implemented by Belarusian authorities until December of 2020. Even when the vaccination campaign eventually started in the very end of 2020<sup>22</sup>, the vaccination rates remained one of the lowest in Europe for a period of time<sup>23,24</sup>. As a consequence, Belarusians remained relatively free in their choices and in personal risks that they were willing to undergo up until the end of 2020. Even then, the newly mandated NPIs were mild in comparison to neighboring Poland, Russia, Ukraine, Lithuania and Latvia.

Given such circumstances, the thorough study of the Belarusian approach towards battling the COVID-19 pandemic is crucial, since the Belarusian cases illustrates: (1) the implications of limited NPIs in the country with high HDI and the universal healthcare system and (2) the limited reporting and limited publicly available surveillance data. In the meantime, despite aforesaid reporting such study is still possible since Belarus has a surveillance system comparable to the other countries with high HDI.

At first glance the lab confirmed COVID-19 mortality serves as a direct and unbiased measure of the epidemic burden since (almost) all such deaths are recorded properly. This measure, however, has flaws. In reality, most of the asymptomatic individuals and some individuals with mild (and even severe) symptoms remain untested. As a consequence, death of *untested* individuals is not reflected in COVID-19-induced mortality statistics. This also happens due to limited testing capacities which vary by region and over time. On the other hand, some *tested positive* deceased individuals may have had causes of death other than COVID-19. Those may include various pre-existing health complicating conditions such as deadly chronic diseases or recent severe traumas. For such individuals the exact cause of death is ambiguous but their deaths can be included in COVID-19 mortality.

To address this uncertainty in COVID-19 mortality the study of the *overall and excess mortalities* during the pandemic period comes to places. More specifically, the estimated change in the overall population mortality from *all causes* during the pandemic period provides a well defined measurement estimate of the epidemic toll for a given region which can be compared across regions<sup>25,26</sup>.

Such statistical estimates are robust since they (1) do not depend on the preformed amount of laboratory tests and (2) do not rely on proper identification of causes of death and related ambiguity. Belarus as a country does not participate in various regional monitoring initiatives such as a European mortality monitoring activity (EuroMOMO)<sup>27</sup>. Therefore, this work fills this knowledge gap for Belarus by studying the excess *overall* mortality and related characteristics such as the related Google Trends search queries<sup>28</sup> during the epidemic period. This also complements the limited available knowledge about the Belarusian COVID-19 burden<sup>25,29</sup>. The limitations of such an excess mortality approach are that the exact (clinical) causes of each death are not reported for Belarus. Therefore, the excess mortality summaries by clinical cause cannot be directly attributed to COVID-19 as well as compared with such summaries for the neighboring countries which do report such data. They can only be interpreted as the overall “mortality burden” during the pandemic with potentially multiple underlying causes.

This work contains: (1) the analysis of officially reported overall death counts and excess mortality, (2) the demographic characteristic summaries and their incorporation into models, (3) the comparisons of mortalities via standardized scores with neighboring counties, (4) the studies of the auxiliary Google Trends search data and their relationship with mortality data during the pandemic period.

## Data sources

**Demographics and overall mortality.** No individual level data have been used in this work. The analysis was performed on publicly available aggregated population summaries which did not include any individual level personal information. The annual population structure and annual mortality were downloaded from the National Statistical Committee of the Republic of Belarus (NSCRB)<sup>30,31</sup>. For the past ten years the NSCRB published population data *annually* as “Demographic Yearbook of the Republic of Belarus” and (as of November 2021) the latest available yearbook was published for 2019<sup>32</sup>. Therefore, for year 2020 the Belarusian census summaries from October 2019 were used instead.

The Belarusian *monthly* mortality from January 2011 to June 2020 were downloaded from the United Nations Demographic Statistics Database<sup>33</sup>. At authors knowledge as of November 2021 this is the most detailed and the most up to date overall mortality dataset<sup>25</sup>. In particular, the most recent edition “Belarus in Numbers 2021” by the NSCRB was released on April 29, 2021<sup>34</sup> and *did not* include the annual mortality counts while they were included in pre-COVID-19 “Belarus in Numbers 2020”<sup>35</sup>.

**COVID-19 cases and mortality.** The COVID-19 epidemiological data were reported by the Ministry of Public Health of the Republic of Belarus<sup>36</sup>. Since the beginning of the pandemic (March 2020) the reports were in the format of online press releases<sup>36–38</sup> and daily numbers of confirmed cases, deaths and conducted tests. The daily updates were reported via the Ministry of Public Health official Telegram Messenger channel<sup>38</sup>. Since Telegram channel data were in unstructured text format the news agency *onliner.by* generated interactive visual summaries for incidence, recovery, and death counts<sup>39</sup>.

**Google trends.** Google Trends *monthly* searchers<sup>28</sup> from January 2015 until August 2021 were considered as auxiliary data to complement the existing knowledge. The following search query time series were considered: (1) “гроб” (“grob”) i.e. coffin, (2) “поминки” (“pominki”) i.e. memorial service and (3) “ритуальные услуги” (“ritualnie usluzhi”) i.e. funeral services.

## Methods

**Excess mortality quantification.** The study is focused on a retrospective analysis and comparison of the (standardized) Belarusian excess mortality with the neighboring countries. Broadly, the *excess mortality* during a period of time (e.g. month) is defined as the actual death counts minus the death counts that would have “normally expected”<sup>26</sup>, where “normally expected” can be interpreted pretty broadly. In this work the formal definition of “excess death” during a certain period (i.e. month) is the difference between the actual death counts minus the “expected counts” which are predicted from statistical modeling. Such modeling can incorporate auxiliary data such as population age characteristics over time. The approach is to formulate and to fit models to historic mortality data over  $n$  time periods (i.e. months) that directly precede the epidemic start date. Then the same models are used to produce the “expected” predicted mortality during the epidemic period for comparisons with the reported mortality data. The differences between reported and predicted produce the formal quantification of “excess mortality”.

**Standardized mortality scores.** While the modeling efforts do provide a formal quantification of the “excess” mortality numbers they are lacking a meaningful way to compare those counts between countries or regions. The later comparison may be very much desired to compare the effectiveness of the applied NPIs.

The formal comparisons between countries (or regions) can be done via the *standardized scores*<sup>40</sup>. In particular, the  $P$ -scores allow to compare mortality counts for different time periods and populations of different regions with different sizes<sup>41</sup>. The scores are defined either non-parametrically or parametrically<sup>26</sup> and both are evaluated in this work. The non-parametric version for  $I$  studied periods has the form:

$$P(t_i) = \frac{x(t_i) - \bar{x}(t_i)}{\bar{x}(t_i)} \text{ for } i = 1, 2, \dots, I, \quad (1)$$

where  $\bar{x}(t_i)$  is the average mortality of all  $t_i$  periods from the past  $n$  years that precede the studied epidemic period.

The parametric version for  $I$  study periods has the form:

$$\mathcal{P}(t_i) = \frac{x(t_i) - \hat{\mu}(t_i)}{\hat{\mu}(t_i)} \text{ for } i = 1, 2, \dots, I \quad (2)$$

where  $\hat{\mu}(t_i)$  is the model-predicted population mortality for  $t_i$  studied period based on the model fit using  $n$  years of data that precede the studied epidemic period. The differences between non-parametric (1) and parametric (2) and expected to be minor<sup>26</sup>. The scores  $P(t_i)$  and  $\mathcal{P}(t_i)$  are typically converted to percent for easier interpretation. For example, the converted score value of 120.47 indicates that mortality is exceeded by 120.47%. Since the mortality scores are on the same scale regardless of the studied population size, they can be easily interpreted and compared *between regions*.

The non-parametric (1) and parametric (2) scores are not limited to mortality data and corresponding time series. The same scores can be computed and studied for any time series such as Google Trends search queries which is also done in this work.

**Predictive models for parametric estimates.** The model which is chosen to produce the expected estimates  $\hat{\mu}(t_i)$  for  $i = 1, 2, \dots, I$  for formula (2) has to account for (competing) processes which are simultaneously happening within the studied population and has to incorporate the ability to disentangle the effects of each process on the mortality. In particular, the model should at least incorporate annual seasonality of mortality<sup>42</sup>, potential time lags as well as changes in the studied population structure and size<sup>43</sup>. For COVID-19 the demographic change is crucial, since the Belarusian population is ageing and the number of high risk individuals older than 65 is increasing<sup>44</sup>. Such demographic features have not been incorporated in excess mortality studies earlier<sup>25</sup>, while such changes can potentially affect the mortality dynamics. Two time series models have been used: the Prophet forecasting model<sup>45</sup> and an autoregressive integrated moving average (ARIMA) model<sup>46</sup>.

The Prophet model<sup>45</sup> is implemented in R package `prophet` and consists of three main components: trend, seasonality and holidays. The model is formulated as:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (3)$$

for a given time interval  $t$ . The formula (3) consists of dependent (response) model variable  $y(t)$  which represents mortality counts or search trend values, independent (exploratory) variables  $g(t)$ ,  $s(t)$  and  $h(t)$  which represent model trends, and the model error  $\epsilon_t$ . Variable  $g(t)$  represents the non-periodic modeling trend component,  $s(t)$ —periodic modeling component, and  $h(t)$ —extreme events (or holidays) which can occur with an irregular schedule. The model error  $\epsilon_t$  represents the processes which are unexplained by the model. All the model variables and the error depend on time  $t$ . The model (3) can also incorporate a covariate i.e. an independent predictor for  $y(t)$ <sup>47</sup>. Therefore, the number of individuals in the population of Belarus who were 65 years or older was considered as such covariate in the model fitting process.

The autoregressive integrated moving average (ARIMA) model is implemented in R package `forecast`<sup>46</sup>. The ARIMA model is determined by the set of three parameters  $(p, d, q)$ . The parameter  $p$  determines the lag of the model i.e. the number of previous observations  $Y(t-1), Y(t-2), \dots, Y(t-p)$  which are used to fit the current observation  $Y(t)$ . The parameter  $d$  stands for the “degree” of the model, i.e. for how many times the previous response has to be subtracted to ensure the required stationarity model assumption<sup>48</sup>. For example, for  $d = 0$  the newly defined response variable is  $y(t) = Y(t)$ , for  $d = 1$  the modeled variable is  $y(t) = Y(t) - Y(t-1)$ , and for  $d = 2$  the modeled variable is  $y(t) = [Y(t) - Y(t-1)] - [Y(t-1) - Y(t-2)]$ . The third parameter

$q$  stands for the number of moving average parameters within the model. The formal ARIMA modeling equation has the form:

$$y(t) = \mu + \sum_{i=1}^p \varphi_i y(t-i) + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t. \quad (4)$$

The formula (4) contains the response model variable  $y(t)$  which represents mortality counts (or search trend values) and is parameterized by the overall mean parameter  $\mu$ , the autoregressive parameters  $\varphi_i$  where  $i = 1, 2, \dots, p$ , the moving average parameters  $\theta_j$  and the model errors  $\epsilon_{t-j}$  where  $j = 0, 1, 2, \dots, q$ .

**Score computations.** The non-parametric  $P$ -scores for mortality data for January–June of 2020 were computed based on the monthly averages  $\bar{x}(t_i)$  for the first six months of the year i.e. for  $t_1, t_2, \dots, t_6$ . The averages were based on *five* (2015–2019) and *nine* (2011–2019) years of data. The analogous computations of  $P$ -scores were preformed for three Google Trends based on *five* (2015–2019) years of data.

The parametric  $\mathcal{P}$ -scores for mortality data were computed based on the predicted averages  $\hat{\mu}(t_i)$  that came from the Prophet forecasting model and the ARIMA model. The both models were fit to two time series with *five* (January 2015–February 2020) and *nine* (January 2011–February 2020) years of monthly mortality. For all fits the *four* following epidemic months (March 2020–June 2020) were predicted by those models. The Prophet and ARIMA models were fitted *both* without a covariate and with a covariate which was the number of individuals of age 65 or older. For the Prophet model the default package settings were used while for the ARIMA model `auto.arima` function was used for *nine* years to determine  $(p, d, q)$  parameters. The corresponding ARIMA settings were used for *five* years fits since those `auto.arima` options were deemed to be the most reliable.

For Prophet model the 95% upper bound of model fits (and forecasts for the last four months) were used to produce the expected mortality  $\hat{\mu}(t_i)$  and the corresponding  $\mathcal{P}(t_i)$ -scores (2). Those estimates are conservative since they use the upper bound of forecasted expected mortality thus producing the lower bound estimate of  $\mathcal{P}(t_i)$ -scores. The `forecast` package only provides ARIMA upper uncertainty bounds for the model forecasts and not for the model fits. Therefore, instead of 95% upper bound those model-produced values  $\hat{\mu}(t_i)$  both for the fitted and for the four forecasted months were used to produce the  $\mathcal{P}(t_i)$ -scores.

The analogous scores (2) based on *five* (January 2015–February 2020) years of monthly trends data and the Prophet model were used to produce the expected averages  $\hat{\mu}(t_i)$  and the corresponding  $\mathcal{P}(t_i)$ -scores of three Google Trends. Only the Prophet model was fitted for Google Trends since the results for the Prophet and ARIMA models for mortality were very similar.

**Mortality and google trends data linkage.** To study the associations and (potential) relationships between mortality and Google Trends the corresponding time series were compared pairwise. The comparisons were performed for the time period when they all were available, i.e. for monthly records from January 2015 to June 2020. For easier visual comparison, the series were standardized to be on the same scale. More specifically, the values of each time series were divided by the corresponding median of that series values and the resulting values were multiplied by 100. In addition to that all four standardized series were smoothed using `loess` function in R with `span = 0.25` parameter to remove potential noise. The resulting standardized and smoothed standardized values were compared using the Granger dependency test for time series *with dependent records*<sup>49</sup>. The graphical summaries were also produced.

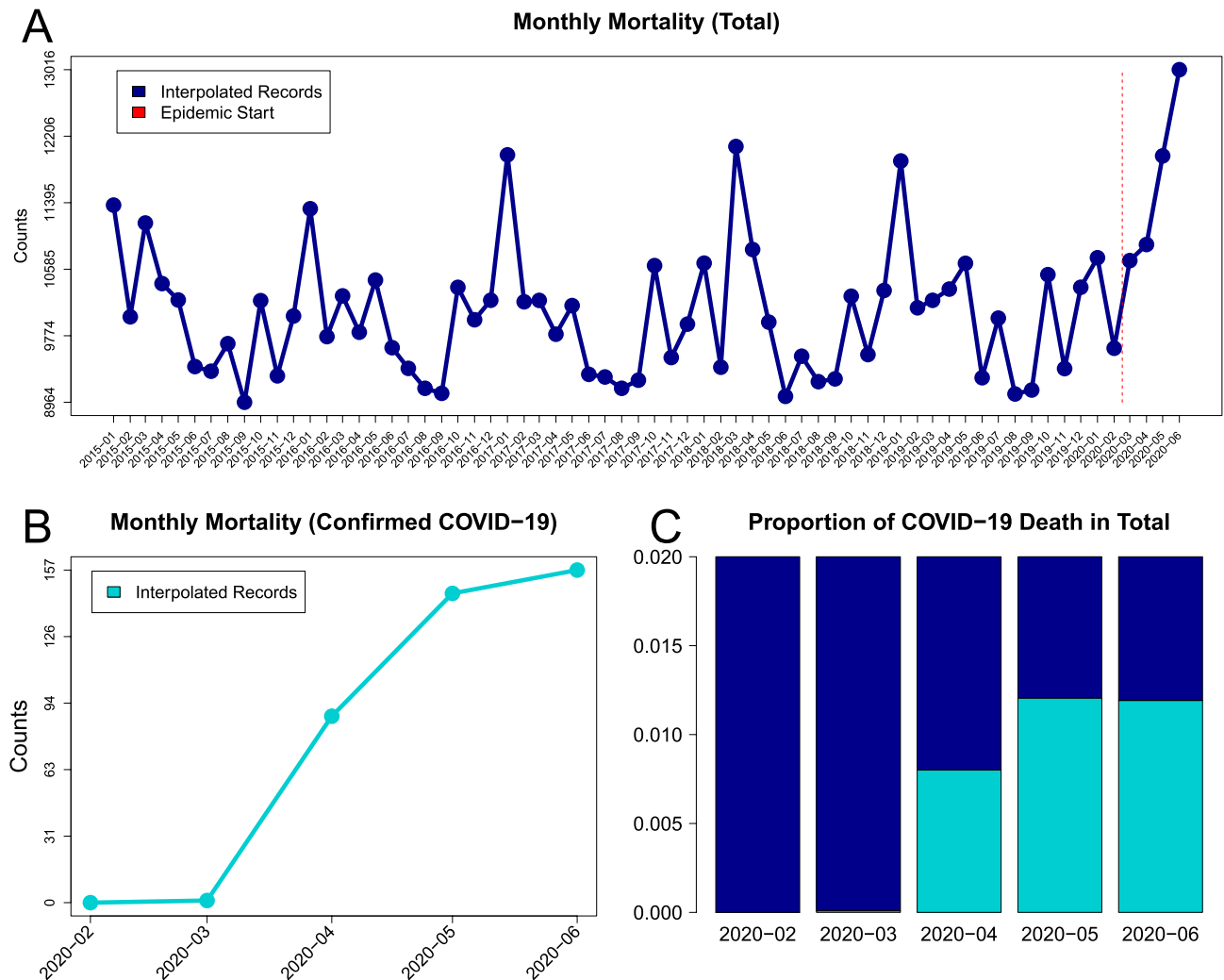
## Results

**Raw data summaries.** The visual summaries of overall and COVID-19-related mortalities are presented in Fig. 1. The overall mortality increased since April 2020 (Fig. 1 Panel A) while the number (Fig. 1 Panel B) and the proportion of COVID-19-related deaths among the total mortality remained notoriously low for the first four available months of the pandemic (Fig. 1 Panel C). In particular, the reported COVID-19-related mortality constitutes less than 1.5% of the overall mortality during the reported April–June of 2020 (Fig. 1 Panel C).

The Google Trends data visual summaries for “grob” which means coffin, “pominki” which means memorial service and “ritualnie uslugi” which means funeral services key phrases are provided in Fig. 2. Such data are extremely variable and the trend depends heavily on the chosen key phrase. Still, the patterns do change visually after the beginning of the pandemic in March 2021. The coffin (“grob”) variable has the most pronounced trend.

**Non-parametric  $P$ -scores.** The non-parametric  $P$ -scores for mortality data based on *five* and *nine* years histories are presented in Fig. 3. The  $P$ -score values for early 2020 are *negative* indicating low recorded mortality. The scores *increased* from April 2020 and for the latest reported month June they were 39.56% for *five* and 35.92% for *nine* years histories. The summaries for *five* years averages (Fig. 3 panels A and C) and for *nine* years averages (Fig. 3 panels B and D) have very similar results while there is a clear trend of substantial incidence increase during the pandemic. The summaries from non-parametric mortality  $P$ -scores from Fig. 3 agree with high excess mortality identified earlier<sup>25</sup>.

The non-parametric  $P$ -scores for the Google Trends search data based on *five* years histories are presented in Fig. 4. The Google Trends search results do vary substantially from one key phrase to the other. The key phrase “grob” i.e. coffin had  $P$ -score values of 138.1% for April of 2020 and of 148.3% for September of 2020. The peak values for coffin were observed in the middle of 2020 and later they started to decrease. The key phrase “pominki” i.e. memorial service had a large variability with no clear trend. The largest  $P$ -score value was 256.48% for October of 2020 and the smallest value was  $-100\%$  for August of 2020. Finally, the key phrase “ritualnie uslugi” i.e. funeral



**Figure 1.** The visual summaries of the raw data for overall and COVID-19-related mortality: **(A)** Total monthly mortality counts joined by a linear interpolator. The start of the epidemic is marked with a vertical red bar. **(B)** COVID-19-related monthly mortality counts joined by a linear interpolator. **(C)** Monthly proportions of COVID-19-related mortality among the total mortality.

services had the pronounced excess of 218.47% for January of 2021. In summary: (1) the excess of non-parametric  $P$ -scores was observed during the pandemic period and (2) the excess was not uniform across different Google Trends, which emphasizes potential self-reporting search biases.

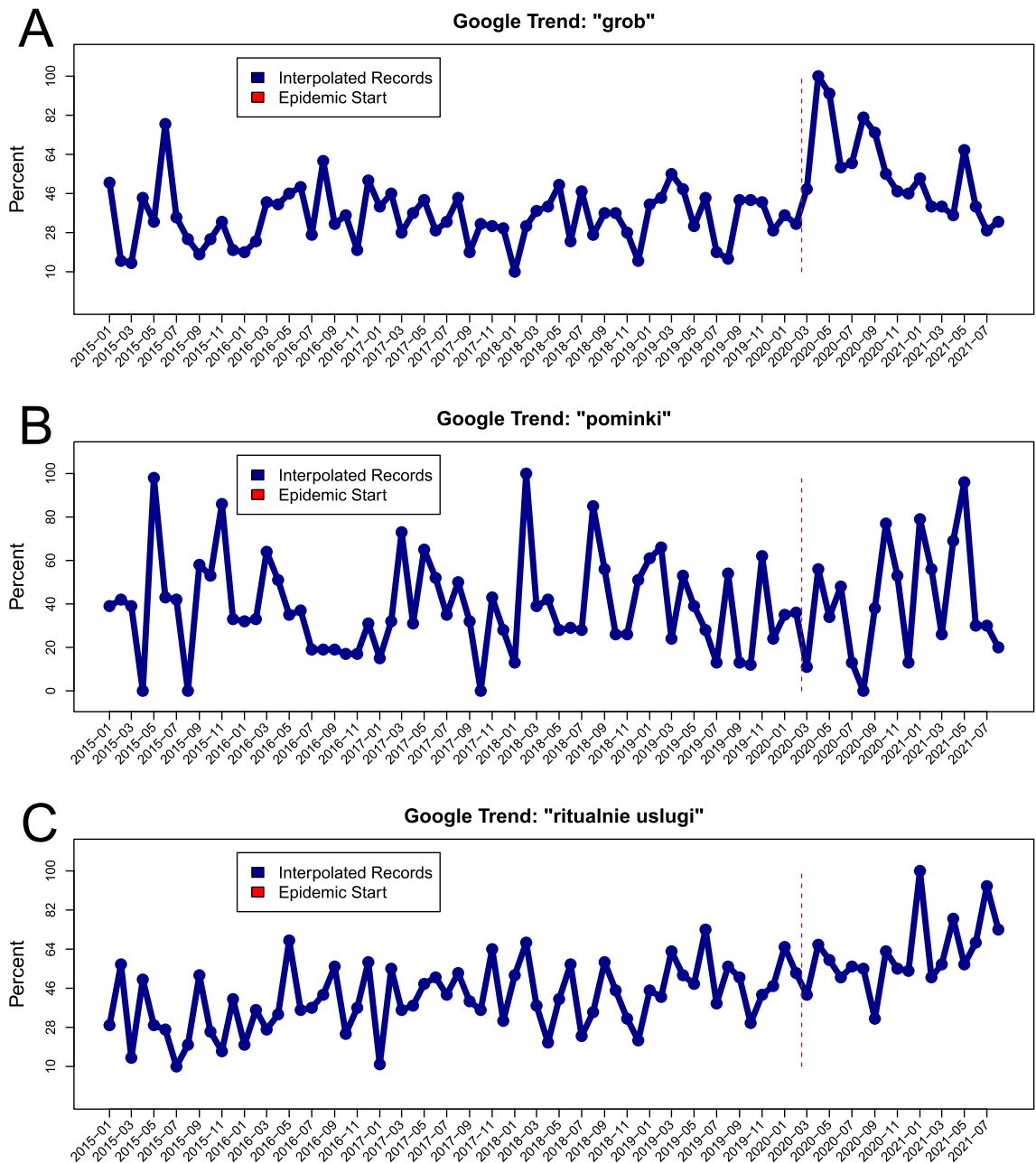
The presented non-parametric  $P$ -scores are *robust* for misspecifications since they do not depend on any model assumptions but they also cannot include covariates.

**Parametric  $\mathcal{P}$ -scores.** The parametric  $\mathcal{P}$ -scores for mortality data based on the Prophet forecasting model fitted to *five* and *nine* years histories *without* demographic characteristics are presented in Fig. 5. The  $\mathcal{P}$ -score values for early 2020 are *negative* indicating low recorded mortality. The scores *increased* from April 2020 and for the latest reported month June they were 29.81% for *five* and 30.89% for *nine* years histories. The summaries for *five* years histories (Fig. 5 panels A and C) and for *nine* years histories (Fig. 5 panels B and D) have very similar results while there is a clear trend and substantial mortality increase during the pandemic.

The corresponding version of the Prophet model *with* demographic characteristics (i.e. the number of individuals of age 65 and older) produced very similar results (33.29% and 33.13%) which have been summarized in Supplement Fig. S1. In addition to that, the comparisons of the Prophet model fitting for *nine* years *with* and *without* demographic characteristics are presented in Supplement Fig. S2. There is no indication from such comparisons that the ageing population had an effect on mortality during the past pre-pandemic years while the pandemic period was substantially different from pre-pandemic period.

The parametric  $\mathcal{P}$ -scores for mortality data based on the ARIMA model had very similar trends and results to the ones from the Prophet model. This was true *both* for the models *with* demographic characteristics (i.e. the number of individuals of age 65 and older) and for those *without* them. The scores *increased* from April 2020 and for all ARIMA models. For the models *without* demographic characteristics for the latest reported month June the

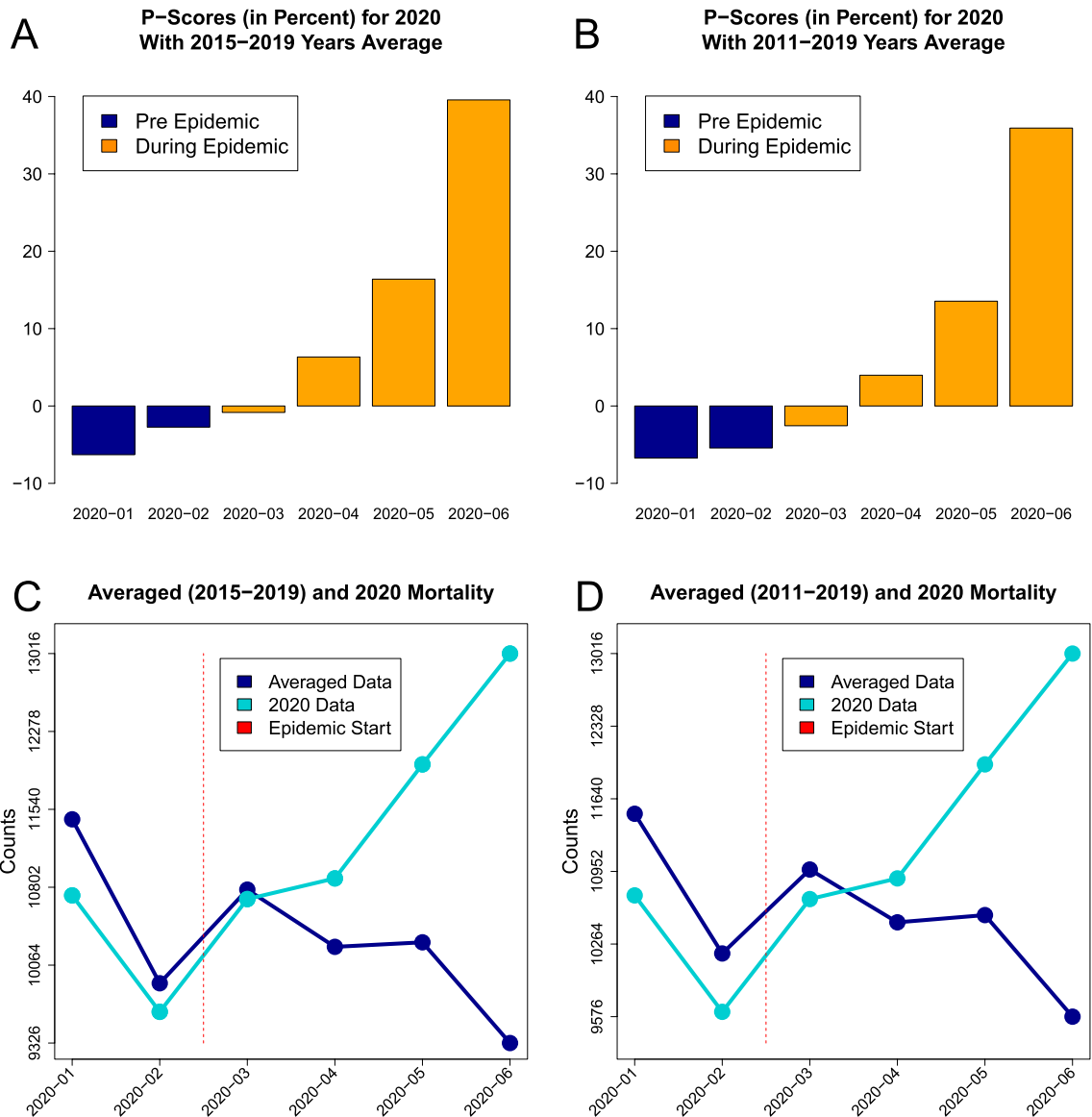




**Figure 2.** The visual summaries of the monthly raw Google trends data from January 2015 until August 2021 for (A) "grob" i.e. coffin, (B) "pominki" i.e. memorial service and (C) "ritualnie usługi" i.e. funeral services. The vertical red bar emphasizes the epidemic start, i.e. time of the first confirmed cases.

scores were 31.07% for *five* and 33.43% for *nine* years histories. For the models *with* demographic characteristics (i.e. the number of individuals of age 65 and older) for the latest reported month June the scores were 29.46% for *five* and 35.59% for *nine* years histories. The corresponding visual summaries are provided in Supplement Fig. S3 and Fig. S4. In addition to that, the comparisons of the ARIMA model fitting for *nine* years *with* and *without* demographic characteristics are presented in Supplement Fig. S5. In the same way as for the Prophet model there is no indication from such comparisons that the ageing population had an effect on mortality over the past pre-pandemic years while the pandemic period was substantially different from pre-pandemic period.

The parametric  $\mathcal{P}$ -scores for the Google Trends search data based on the Prophet forecasting model are presented in Fig. 6. In the same way as for non-parametric scores the results for  $\mathcal{P}$ -scores do vary substantially from one key phrase to the other. The key phrase "grob" i.e. coffin had the pronounced excess of 97.1% for April of 2020 and 86.6% for September of 2020. The peak values for coffin were observed in the middle of 2020. The key phrase "pominki" i.e. memorial service did not have a clear trend over the study period and large variability was observed. The largest score was the excess of 66.91% for October of 2020 and the smallest score was the shortage of -100% for August of 2020. Finally, the key phrase "ritualnie usługi" i.e. funeral services had the pronounced excess of 95.40% for January of 2021. In summary: (1) the excess of parametric  $\mathcal{P}$ -scores was observed during

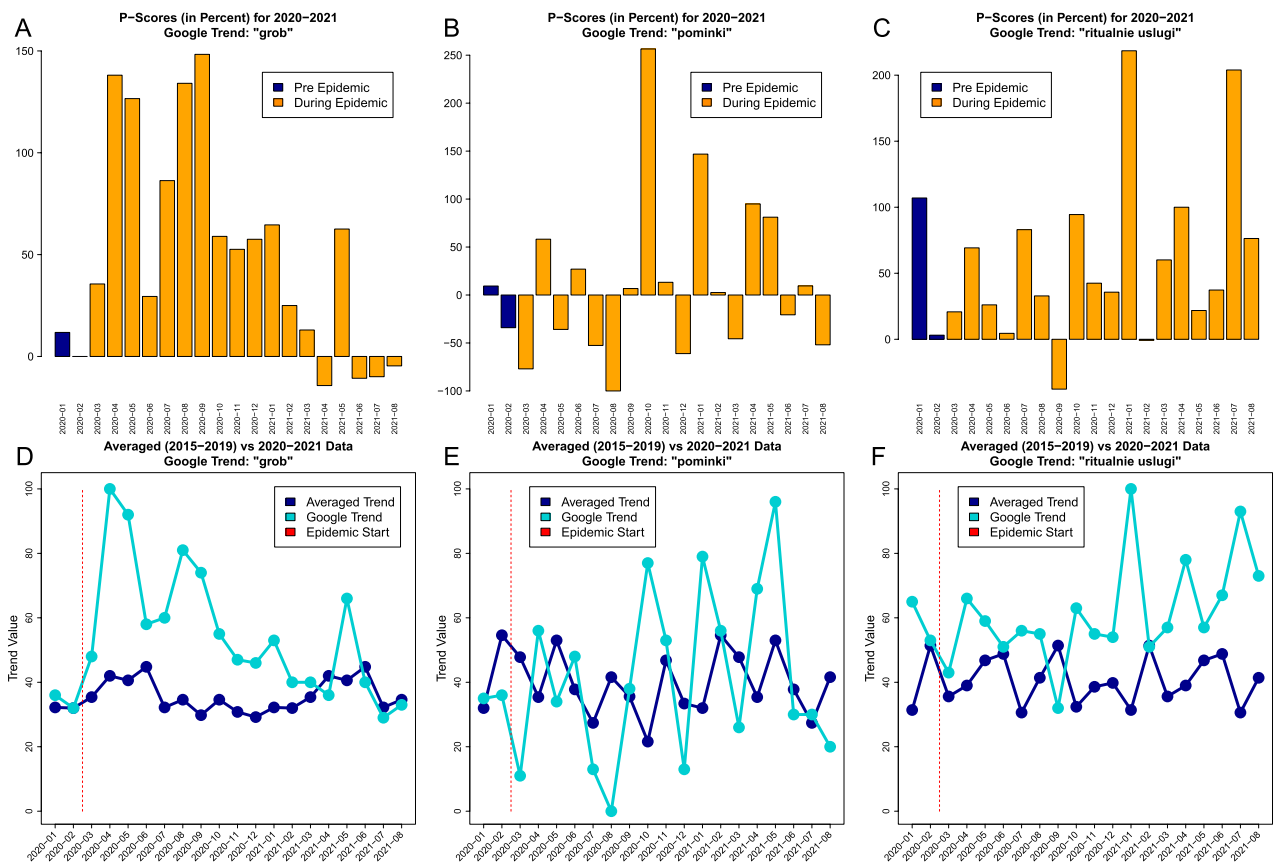


**Figure 3.** The visual summaries of non-parametric mortality  $P$ -scores presented for *five* (panel A) and *nine* (panel B) years where orange bars represent the epidemic period. The corresponding averaged data (dark blue) based on the previous *five* (panel C) and *nine* (panel D) years are presented along with the reported data for 2020 (cyan). Vertical red bars indicate the epidemic start period.

the pandemic period, (2) the excess was not uniform across different Google Trends which emphasizes potential self-reporting search biases, and (3) the parametric scores were more “smooth” and “less extreme” than the non-parametric scorers causing smaller variability and less extreme values.

**Compassion of excess mortality scores across methods and regions.** The formal comparison between non-parametric (1) and parametric (2) scores and the corresponding “excess mortality” based on *five* years histories are presented in Tables 1 and 2. The summaries for *nine* years histories are very similar and are available in Supplement Tables S1 and S2. Regardless of the method both  $P$  and  $\mathcal{P}$  scores increased since March 2020 and were in the range from  $\sim 10\%$  to  $\sim 20\%$  for May 2020 and from  $\sim 29\%$  to  $\sim 40\%$  for June 2020. The corresponding excess mortality estimates were from 1116 to 2035 deaths for May 2020 and from 2953 to 3690 deaths for June 2020. For all models the estimated “excess death” far exceeded the reported COVID-19-related mortality for the same time period (Fig. 1).

In the meantime, the excess mortality is expected everywhere during the pandemic. Therefore, since  $P$  and  $\mathcal{P}$ -scores have similar values the formal comparisons of Belarusian  $P$ -scores with neighboring and more distant countries can provide a formal comparison of the pandemic burden. In particular, based on the interactive maps of ourworldindata.org<sup>50</sup> all scores computed for Belarus for June 2020 were higher than for *all* neighboring counties with 5% for Poland, 5% for Ukraine, 8% for Russia, 11% for Lithuania, 11% for Latvia. Our findings also agree with exceptionally high mortality in Belarus which has been documented earlier<sup>25</sup> but provide extra



**Figure 4.** The visual summaries of non-parametric Google Trends searches  $P$ -scores are presented for coffin (panel A), memorial service (panel B) and funeral services (panel C) where orange bars are used for the epidemic period. The corresponding averaged data based on five years (dark blue) are plotted together with the queries data (cyan) for coffin (panel D), memorial service (panel E) and funeral services (panel F). Vertical red bars indicate the epidemic start period.

insides in terms of multiple estimation approaches (non-parametric, Prophet and ARIMA) with the inclusion of demographic characteristics (i.e. number of individuals older than 65) in the study.

As of November 2021 Belarus is the only country among its neighbors where no mortality has been reported past June 2020. Moreover, based on the available data, the mortality demonstrated the exponential growth from March 2020 until the reporting ended. It is also worth mentioning that June was not the peak month for mortality in the region. For example, neighboring Poland has demonstrated the peak in excess mortality in November 2020 with the score value of 117%. The other available studies of excess mortality also indicate the highest underreporting COVID-19 rate for Belarus in comparison to all neighbors<sup>51</sup> which agrees with the findings of this manuscript.

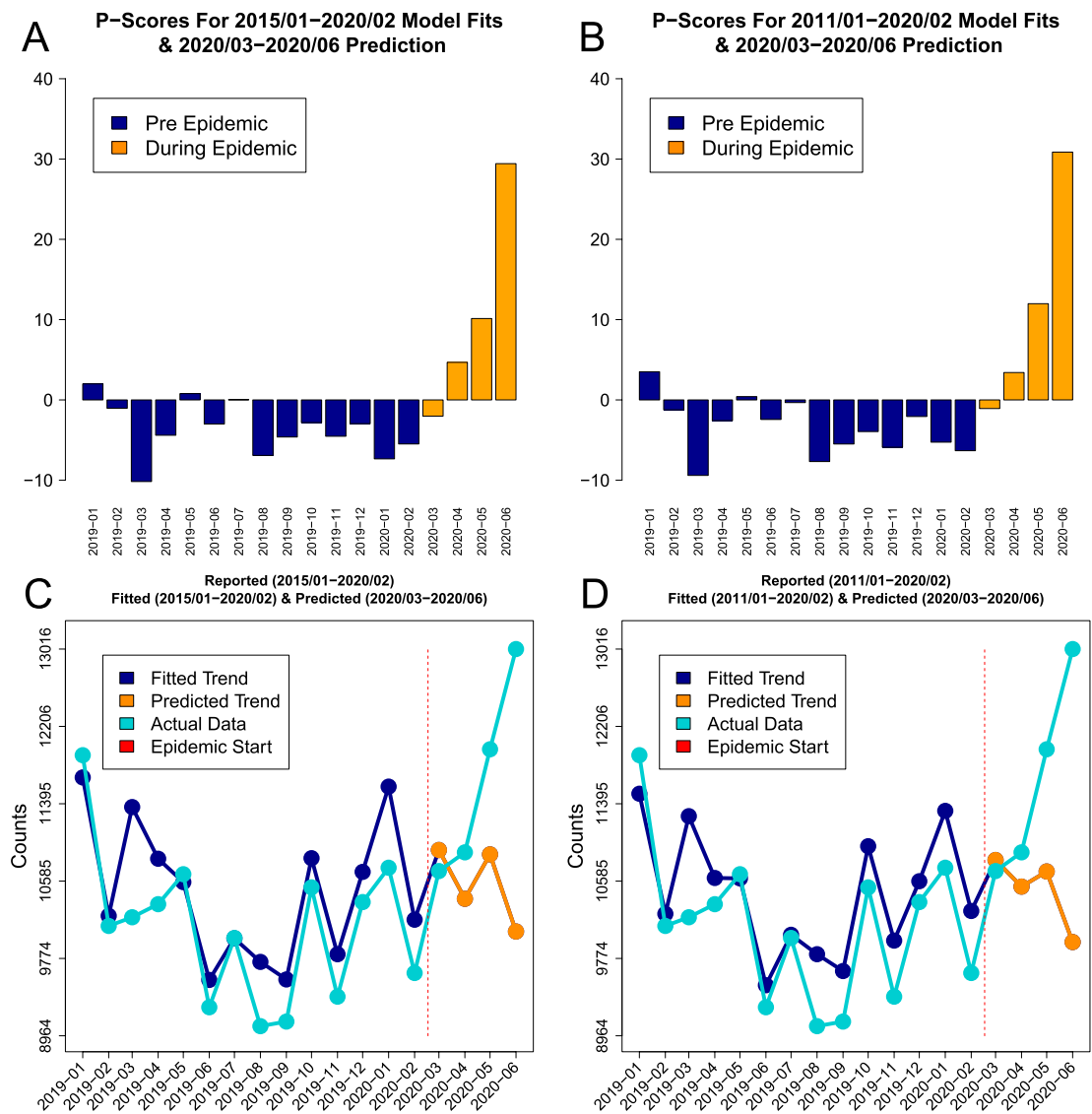
For a complete picture, the analysis results for Belarus were compared with the more distant Western European neighbors: United Kingdom, Spain, Belgium, Italy, Netherlands, and France<sup>52</sup>. For those countries the corresponding  $P$ -scores for peak weeks in April 2020 were 105–108% for the United Kingdom, 117–156% for Spain, 90–105% for Belgium, 55–81% for Italy, 52–75% for Netherlands, and 57–65% for France depending on the week<sup>53</sup>. In the meantime, the direct comparison of Western European pandemic “peak” in April with the values in June for Eastern Europe is not correct: the epidemic “peak” in Western Europe was exactly in April, which was followed by a steady decline while in Eastern Europe and in Belarus in particular the epidemic was just starting in April and peaked much later than June. The manuscript findings also agree with the previous study which concluded that early lockdowns reduced mortality<sup>54</sup>.

**Mortality and google trends data relationships.** The pairwise summaries of the standardized and smoothed standardized time series are presented in Fig. 7. The smoothed trends appear visually related, which is especially pronounced for “grob” i.e. coffin and for “ritualnie usługi” i.e. funeral services variables.

The more formal comparisons of Granger’s test  $p$ -values for scaled ( $p_{GR}$ ) and smoothed scaled ( $p_{GR(Sm)}$ ) trends are provided in Table 3.

The results of Granger’s test indicate moderate significance for scaled and strong significance for smoothed scaled data.





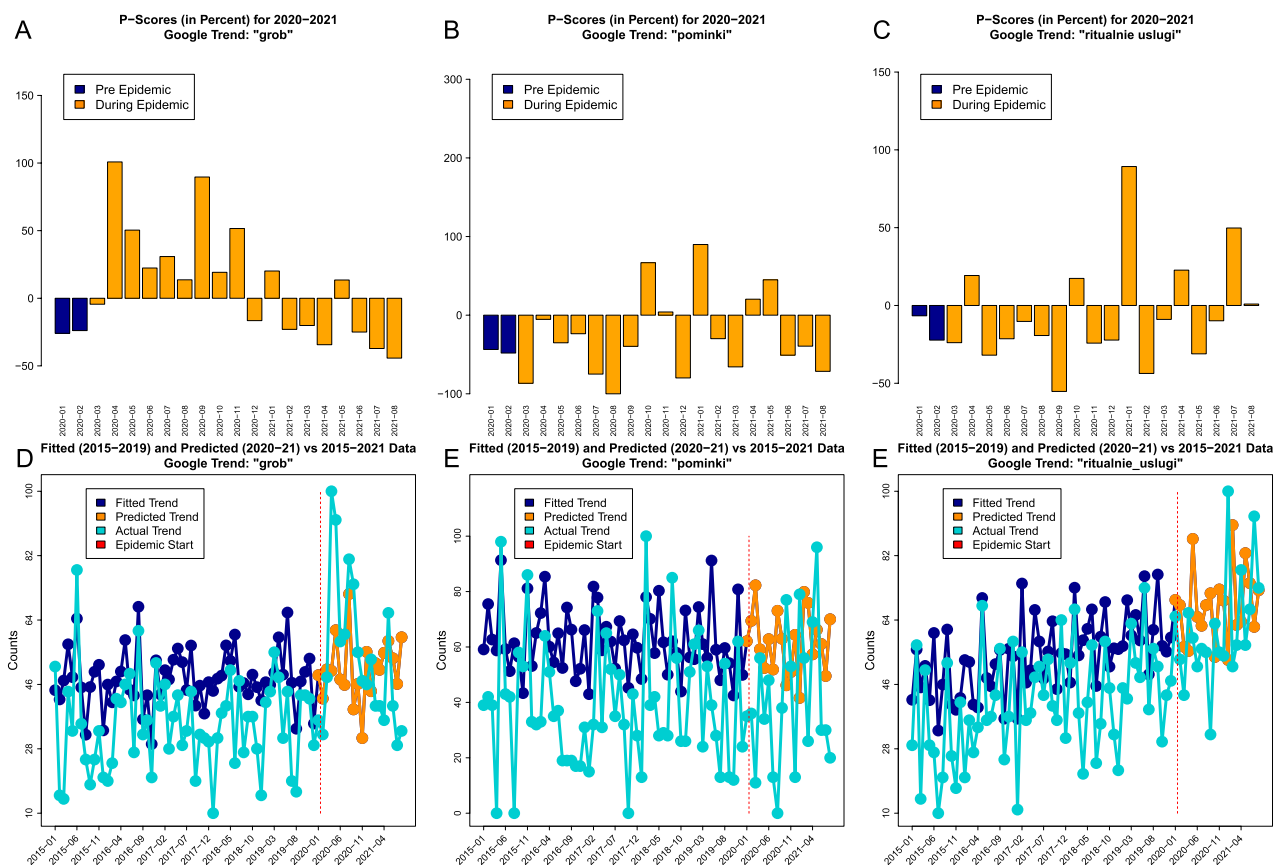
**Figure 5.** The visual summaries of parametric mortality  $\mathcal{P}$ -scores based on the Prophet model *without* demographic characteristics are presented for *five* (panel A) and *nine* (panel B). The orange bars represent the epidemic period. The corresponding Prophet model predictions (dark blue) based on the previous *five* (panel C) and *nine* (panel D) years are presented along the reported data (cyan). Vertical red bars indicate the epidemic start period.

## Discussion

The findings of this study suggest that there are indications that an approach with limited NPIs could have resulted in the overall mortality increase. While such an increase was observed among all neighboring countries where more stringent NPIs were implemented, the Belarusian mortality was the highest for the last reported month of June 2020. The exact causes of such increase in the overall mortality are unknown, but the timeline of the increase corresponded to the pandemic time period in Belarus. Either way, there are indications that the overall mortality in Belarus during the pandemic period which was reported was substantially higher than expected otherwise and also higher than for all neighboring countries.

The limited NPIs were the primary factors that distinguished Belarus from its direct neighbors. The direct comparisons of the county's NPIs and the mortality rates, however, did not necessarily imply causal relationship. There could be many other factors, which affected the increase in mortality, such as multiple differences in healthcare systems and population structure between the countries. In particular, the effects of the population age structure changes were investigated. Based on the available demographics data there were (1) no substantial changes in the number of individuals in the age group of 65+ during the studied period and (2) the inclusion of such data into the models resulted in similar mortality predictions and scores as from models without such data.

The inclusion of Google Trends search data and the corresponding analysis, served as an auxiliary indicator that a higher increase in mortality could have happened until the end of 2020. No further studies, however, can be performed on exact mortality counts until more data will become available.



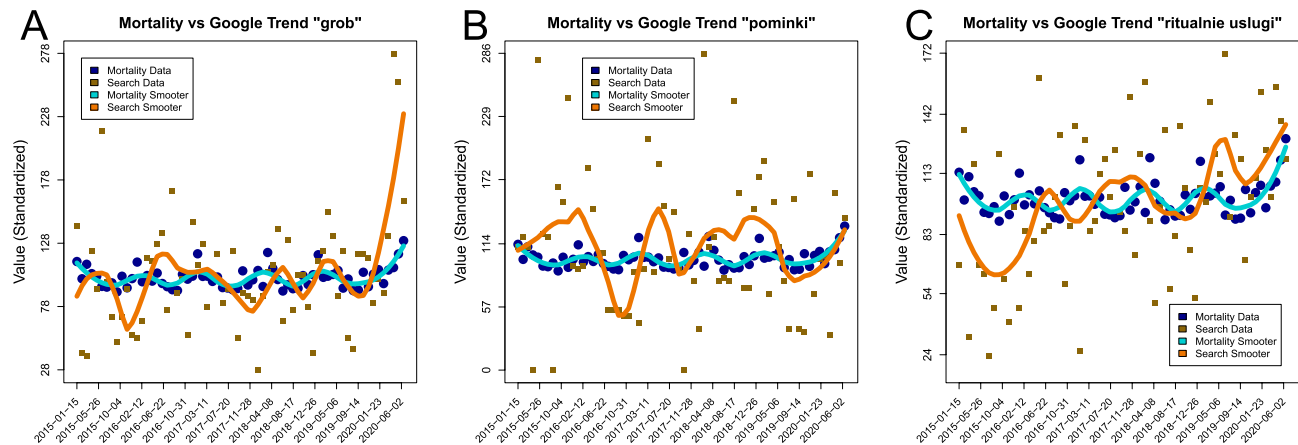
**Figure 6.** The visual summaries of parametric Google Trends searches  $\mathcal{P}$ -scores are presented for the coffin (panel A), memorial service (panel B) and funeral services (panel C) where orange bars are used for the epidemic period. The corresponding model predictions based on *five* years (dark blue) are plotted together with the queries data (cyan) for coffin (panel D), memorial service (panel E) and funeral services (panel E). Vertical red bars indicate the epidemic start period.

Method	2020-01	2020-02	2020-03	2020-04	2020-05	2020-06
Non-parametric	− 6.30	− 2.74	− 0.83	6.33	16.40	39.56
Prophet	− 7.23	− 5.08	− 1.82	4.40	10.29	29.35
Prophet & 65+	− 5.47	− 2.94	0.65	7.86	13.37	33.29
ARIMA	7.11	− 4.49	8.16	9.67	20.49	31.07
ARIMA & 65+	4.45	− 3.51	6.88	7.84	19.33	29.46

**Table 1.** The values of  $P$  and  $\mathcal{P}$  for the first six months of 2020 based on *five* previous years of historic data for different modeling approaches.

Method	2020-01	2020-02	2020-03	2020-04	2020-05	2020-06
Non-parametric	− 721	− 271	− 90	648	1686	3690
Prophet	− 836	− 515	− 198	459	1116	2953
Prophet & 65+	− 621	− 291	69	793	1411	3251
ARIMA	712	− 452	806	960	2035	3086
ARIMA & 65+	457	− 350	688	791	1938	2962

**Table 2.** The estimated excess mortality counts for the first six months of 2020 based on *five* previous years of historic data for different modeling approaches. Negative values indicate less than predicted counts.



**Figure 7.** The visualization of pairwise comparisons between the standardized mortality records and the standardized Google Trends for (A) "grob" i.e. coffin, (B) "pominki" i.e. memorial service and (C) "ritualnie usługi" i.e. funeral services key phrases. The standardized mortality values are presented as blue dots, while the standardized search data are presented as brown squares. The corresponding smoothers are presented as cyan and dark orange lines.

Name	<i>p</i> <sub>GR</sub>	<i>p</i> <sub>GR(Sm)</sub>
"grob" (coffin)	0.14	< 0.01
"pominki" (memorial service)	0.08	< 0.01
"ritualnie usługi" (funeral services)	0.07	0.08

**Table 3.** The summary of *p*-values for standardized (*GR*) and smoothed standardized (*GR(Sm)*) data from Granger's test.

The retrospective analysis of mortality for a single region and the comparisons of such analyses between regions remain challenging due to inherited limitations. At a first glance, the analysis of excess mortality seems to be a perfect measure of the epidemic burden. Moreover, the use of standardized scores for comparison between regions or countries is handy. In reality, the mortality data accumulation and reporting are far from perfect and have flaws<sup>55</sup>. For Belarus those limitations are exacerbated by limited reporting which does not cover any time past June 2020.

The first challenge is the frequency and completeness of reporting and the corresponding reporting lag. In particular, while the lab confirmed COVID-19 mortality during the pandemic is reported very frequently (e.g. daily), the overall mortality is reported with some delay and for longer time periods i.e. it can be reported weekly, monthly, quarterly or (even) annually. Moreover, the overall mortality tends to be incomplete for developing countries and even for developed countries the proper and complete documentation of *all cases* can take up to one year and beyond<sup>40</sup>. This happens since it may not be immediately possible to identify causes of death and, therefore, it may take some time to produce the appropriate death certificates and to incorporate them into the statistics<sup>40</sup>. As a consequence, even though, those cases do not comprise the majority of statistics, they can introduce reporting delays.

The second challenge is the constantly changing population structure which is not always incorporated into the analysis. In particular, the non-parametric *P*-scores which are commonly used in research cannot incorporate the number of people older than 65 at each given point of time, which is critical for ageing populations like Belarus. Such analysis can only be incorporated in parametric *P*-scores (2) and via modeling of means  $\mu(t)$  which can include the modeling of demographic characteristics over time. This latter parametric approach also comes with limitations, since the population age structure and corresponding population sizes and total counts may not be available with the studied time intervals precision. Therefore, averages over time for longer periods have to be used instead. This was the case for the presented Belarusian analysis where the population size and age structure are only reported annually.

The third challenge is that various demographic and social processes are happening differently within the society during the pandemic. In particular, the overall mortality is affected by multiple factors which can *both* increase or decrease the mortality. Therefore the excess mortality does not directly reflect the mortality caused by COVID-19. For example, individuals can die in excess numbers during the epidemic due to *other causes* unrelated to SARS-CoV-2, such as anxiety and related health complications (e.g. stroke), suicides or inability of the healthcare system to accommodate planned hospitalizations and surgeries at that time. At the same time epidemics may also have a decreasing effect on the overall mortality, since people may travel less and the number of travel-related accidents and deaths can decrease<sup>55</sup>. Therefore, the overall mortality in our study should be interpreted as the *overall COVID-19 burden* on the society as a whole rather than the COVID-19 induced death.

Moreover, the exact (clinical) causes of each death are not reported for Belarus. As a consequence, the mortality summaries by clinical cause cannot be directly compared or attributed to COVID-19.

The fourth challenge comes from the study of the Google Trends search data<sup>56</sup>. We had to refer to them to complement the limited knowledge and to study the alternative sources of information about mortality during the pandemic. As a result, such analyses indeed provided us with evidence about the increased funeral-related search trends during the pandemic. Such data, however, are prone to severe reporting biases since the number of Google users changes constantly. In particular, people may constantly change frequency of use or use multiple search engines at the same time. The other limitation comes from variability of search options i.e. the list of potential key phrases can be endless. We have only focused on the three most obvious key phrases and even those had quite different time series. As a result, the Google Trends search data are *not complete* by any means and the results are *expected* to be biased in one way or the other. Still, from all studied trends the excess searches have been observed during the pandemic which were more or less pronounced.

## Code availability

The analysis source code in R has been made publicly available on GitHub and can be downloaded<sup>57</sup>.

Received: 12 November 2021; Accepted: 17 March 2022

Published online: 31 March 2022

## References

1. Zhu, N. *et al.* A novel coronavirus from patients with pneumonia in china, 2019. *N. Engl. J. Med.* **2**, 1 (2020).
2. Lu, H., Stratton, C. W. & Tang, Y.-W. Outbreak of pneumonia of unknown etiology in Wuhan China: The mystery and the miracle. *J. Med. Virol.* **2**, 1 (2020).
3. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, china. *Lancet* **395**, 497–506 (2020).
4. Chen, N. *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *Lancet* **395**, 507–513 (2020).
5. Ciotti, M. *et al.* The covid-19 pandemic. *Crit. Rev. Clin. Lab. Sci.* **57**, 365–388 (2020).
6. WHO. <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-COVID-19---11-march-2020> (2020). Accessed: 2021-08-30.
7. Åslund, A. Responses to the covid-19 crisis in russia, ukraine, and belarus. *Eur. Geogr. Econ.* **2**, 1–14 (2020).
8. Banik, A., Nag, T., Chowdhury, S. R. & Chatterjee, R. Why do covid-19 fatality rates differ across countries? an explorative cross-country study based on select indicators. *Glob. Bus. Rev.* **21**, 607–625 (2020).
9. Chen, W.-H., Strych, U., Hotez, P. J. & Bottazzi, M. E. The SARS-CoV-2 vaccine pipeline: an overview. *Curr. Trop. Med. Rep.* **2**, 1–4 (2020).
10. Burki, T. K. The russian vaccine for covid-19. *Lancet Respir. Med.* **8**, e85–e86 (2020).
11. Sputnik V. <https://sputnikvaccine.com/about-us/> (2020). Accessed: 2021-08-30.
12. Mahase, E. Covid-19: What do we know about the late stage vaccine candidates? (2020).
13. Mullard, A. How COVID vaccines are being divvied up around the world. *Nature* <https://doi.org/10.1038/d41586-020-03370-6> (2020).
14. Mahase, E. Covid-19: UK approves pfizer and BioNTech vaccine with rollout due to start next week. *BMJ* <https://doi.org/10.1136/bmj.m4714> (2020).
15. Ledford, H. US authorization of first COVID vaccine marks new phase in safety monitoring. *Nature* <https://doi.org/10.1038/d41586-020-03542-4> (2020).
16. Health Canada. <https://healthycanadians.gc.ca/recall-alert-rappel-avis/hc-sc/2020/74543a-eng.php> (2020). Accessed: 2021-08-30.
17. Logunov, D. Y. *et al.* Safety and efficacy of an rad26 and rad5 vector-based heterologous prime-boost covid-19 vaccine: an interim analysis of a randomised controlled phase 3 trial in russia. *Lancet* **397**, 671–681 (2021).
18. United Nations Human Development Reports. <http://hdr.undp.org/en> (2021). Accessed: 2021-08-30.
19. WHO COVID-19 Technical mission of experts to the Republic of Belarus: 8–11 April 2020. Executive summary. <https://www.euro.who.int/en/countries/belarus/publications/covid-19-technical-mission-of-experts-to-the-republic-of-belarus-8-11-april-2020.-executive-summary> (2020). Accessed: 2021-08-30.
20. Huynh, T. L. D. Does culture matter social distancing under the covid-19 pandemic?. *Saf. Sci.* **130**, 104872 (2020).
21. Betsch, C. *et al.* Social and behavioral consequences of mask policies during the covid-19 pandemic. *Proc. Natl. Acad. Sci.* **117**, 21851–21853 (2020).
22. minzdrav.gov.by. <https://minzdrav.gov.by/ru/sobytiya/v-belarusi-nachata-vaktsinatsiya-protiv-koronavirusa/> (2020). Accessed: 2022-02-25.
23. Mathieu, E. *et al.* A global database of covid-19 vaccinations. *Nat. Hum. Behav.* **5**, 947–953 (2021).
24. voxex.org. <https://voxex.org/article/understanding-setup-and-speed-global-covid-19-vaccination-campaigns> (2022). Accessed: 2022-02-25.
25. Karlinky, A. & Kobak, D. Tracking excess mortality across countries during the covid-19 pandemic with the world mortality dataset. *Elife* **10**, e69336 (2021).
26. Our World in Data - ourworldindata.org A pandemic primer on excess mortality statistics and their comparability across countries. <https://ourworldindata.org/covid-excess-mortality> (2021). Accessed: 2021-08-30.
27. EuroMOMO is a European mortality monitoring activity (www.euromomo.eu). <https://www.euromomo.eu/> (2021). Accessed: 2021-08-30.
28. Google Trends - Belarus. <https://trends.google.com/trends/explore?geo=BY> (2021). Accessed: 2021-08-30.
29. Nemira, A. *et al.* Sars-cov-2 transmission dynamics in belarus in 2020 revealed by genomic and incidence data analysis. *Communications Medicine* **1**, 31 (2021).
30. National Statistical Committee of the Republic of Belarus. <https://www.belstat.gov.by/en/> (2021). Accessed: 2021-08-30.
31. National statistical committee of the republic of belarus - population data 1991–2020 (in russian). [https://www.belstat.gov.by/ofitsialnaya-statistika/ssrd-mvf\\_2/natsionalnaya-stranitsa-svodnyh-dannyh/naselenie\\_6/dinamika-chislennosti-naseleniya/](https://www.belstat.gov.by/ofitsialnaya-statistika/ssrd-mvf_2/natsionalnaya-stranitsa-svodnyh-dannyh/naselenie_6/dinamika-chislennosti-naseleniya/) (2021). Accessed: 2021-08-30.
32. National statistical committee of the republic of belarus - demographic yearbook of the republic of belarus 2019. [https://www.belstat.gov.by/ofitsialnaya-statistika/publications/izdania/public\\_compilation/index\\_14356/](https://www.belstat.gov.by/ofitsialnaya-statistika/publications/izdania/public_compilation/index_14356/) (2021). Accessed: 2021-08-30.
33. United nations data (undata) - deaths by month of death. <http://data.un.org/Data.aspx?d=POP&f=tableCode:65#POP> (2021). Accessed: 2021-08-30.
34. Belarus in numbers 2021. [https://www.belstat.gov.by/ofitsialnaya-statistika/publications/izdania/public\\_compilation/index\\_28455/](https://www.belstat.gov.by/ofitsialnaya-statistika/publications/izdania/public_compilation/index_28455/) (2021). Accessed: 2021-05-05.

35. Belarus in numbers 2020. [https://www.belstat.gov.by/ofitsialnaya-statistika/publications/izdania/public\\_compilation/index\\_16711/](https://www.belstat.gov.by/ofitsialnaya-statistika/publications/izdania/public_compilation/index_16711/) (2021). Accessed: 2021-05-05.
36. Ministry of Public Health of the Republic of Belarus (Website). <http://minzdrav.gov.by/en/> (2021). Accessed: 2021-08-30.
37. Ministry of Public Health of the Republic of Belarus - COVID-19 - F.A.Q. (in Russian). <http://minzdrav.gov.by/ru/dlya-beloruski-kh-grazhdan/COVID-19/> (2021). Accessed: 2021-08-30.
38. Ministry of Public Health of the Republic of Belarus - Official Telegram Channel (in Russian). <https://t.me/s/minzdravbelarus> (2021). Accessed: 2021-08-30.
39. onliner.by COVID-19 (in Russian). <https://sp.onliner.by/covid-19/> (2021). Accessed: 2021-08-30.
40. Centers for Disease Control and Prevention - Excess Death. [https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess\\_deaths.htm](https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm) (2021). Accessed: 2021-08-30.
41. EuroMOMO is a European mortality monitoring activity Z-scores. <https://www.euromomo.eu/how-it-works/what-is-a-z-score/> (2021). Accessed: 2021-08-30.
42. Davis, R. E., Knappenberger, P. C., Michaels, P. J. & Novicoff, W. M. Seasonality of climate-human mortality relationships in us cities and impacts of climate change. *Clim. Res.* **26**, 61–76 (2004).
43. Rau, R. *Seasonality in human mortality: a demographic approach* (Springer, Berlin, 2006).
44. Perrotta, F. *et al.* Covid-19 and the elderly: insights into pathogenesis and clinical decision-making. *Aging clinical and experimental research* 1–10 (2020).
45. Taylor, S. J. & Letham, B. Forecasting at scale. *Am. Stat.* **72**, 37–45 (2018).
46. Hyndman, R. J. *et al.* Automatic time series forecasting: the forecast package for r. *J. Stat. Softw.* **27**, 1–22 (2008).
47. cran.r-project.org - prophet package. <https://cran.r-project.org/web/packages/prophet/index.html> (2018). Accessed: 2021-08-30.
48. Shumway, R. H. & Stoffer, D. S. Arima models. In *Time series analysis and its applications*, 75–163 (Springer, 2017).
49. Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **2**, 424–438 (1969).
50. ourworldindata.org BLR, RUS, UKR, LVA, LTU, POL Comparison. <https://ourworldindata.org/grapher/excess-mortality-p-scores?time=earliest..2021-04-30&country=BLR~RUS~UKR~LVA~LTU~POL> (2020). Accessed: 2021-05-05.
51. Ihme - estimation of total mortality due to covid-19. <http://www.healthdata.org/node/8660> (2020). Accessed: 2021-05-05.
52. Aron, J. & Muellbauer, J. Measuring excess mortality: the case of england during the covid-19 pandemic. *Institute for New Economic Thinking, Oxford Working Paper* (2020).
53. ourworldindata.org BLR, GBR, ESP, BEL, ITA, NLD, FRA Comparison. <https://ourworldindata.org/grapher/excess-mortality-p-scores?time=earliest..2021-04-30&country=BLR~GBR~ESP~BEL~ITA~NLD~FRA> (2020). Accessed: 2021-05-05.
54. Balmford, B., Annan, J. D., Hargreaves, J. C., Altoè, M. & Bateman, I. J. Cross-country comparisons of covid-19: Policy, politics and the price of life. *Environ. Resource Econ.* **76**, 525–551 (2020).
55. Max Roser, E. O.-O., Hannah Ritchie & Hasell, J. Coronavirus pandemic (covid-19). *Our World in Data* (2020). <https://ourworldindata.org/coronavirus>.
56. Nuti, S. V. *et al.* The use of google trends in health care research: A systematic review. *PLoS ONE* **9**, e109583 (2014).
57. GitHub—Belarus Mortality Analysis Code. <https://github.com/akirpich-ap/belarus-mortality> (2021). Accessed: 2021-08-30.

## Author contributions

A.K.: conceptualization, methodology, investigation, data curation and analysis, statistical modeling, code writing, writing—original draft, writing—review and editing, supervision, project administration. A.S.: Writing—Original Draft. T.A.W.: Writing—Original Draft. A.P.T.: Conceptualization, Writing—Original Draft. P.S. Conceptualization, Methodology, Writing—Review and Editing. Y.G.: Conceptualization, Writing—Original Draft, Writing—Review and Editing. All authors reviewed the manuscript.

## Funding

PS was supported by the National Institutes of Health grant 1R01EB025022 and by the National Science Foundation grant 2047828.

## Competing Interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-09345-z>.

**Correspondence** and requests for materials should be addressed to A.K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022