medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20222265; this version posted October 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Analysis of the early Covid-19 epidemic curve in Germany by regression models with change points

Helmut Küchenhoff[1], Felix Günther[1,2],
Michael Höhle[3], Andreas Bender[1]
[1]Statistical Consulting Unit StaBLab, LMU Munich, Germany
[2]Department of Genetic Epidemiology,
University of Regensburg, Germany
[3]Department of Mathematics, Stockholm University, Sweden
E–Mail: kuechenhoff@stat.uni-muenchen.de
URL: corona.stat.uni-muenchen.de

October 29, 2020

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20222265; this version posted October 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

We analyze the Covid-19 epidemic curve from March to end of April 2020 in Germany. We use statistical models to estimate the number of cases with disease onset on a given day and use back-projection techniques to obtain the number of new infections per day. The respective time series are analyzed by a Poisson trend regression model with change points. The change points are estimated directly from the data without further assumptions. We carry out the analysis for the whole of Germany and the federal state of Bavaria, where we have more detailed data. Both analyses show a major change between March 9th and 13th for the time series of infections: from a strong increase to a stagnation or a slight decrease. Another change was found between March 24th and March 31st, where the decline intensified. These two major changes can be related to different governmental measures. On March, 11th, Chancellor Merkel appealed for social distancing in a press conference with the Robert Koch Institute (RKI) and a ban on major events with more than 1000 visitors (March 10th) was issued. The other change point at the end of March could be related to the shutdown in Germany.

Our results differ from those by other authors as we take into account the reporting delay, which turned out to be time dependent and therefore changes the structure of the epidemic curve compared to the curve of newly reported cases.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20222265; this version posted October 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

1 Introduction

The first phase of the Covid-19 pandemic in Germany was managed relatively successful in comparison to other countries in Europe. Therefore, it is worth taking a closer look at the course of the pandemic in Germany, which has already led to controversial discussions. This particularly concerns the important question about the effectiveness of various control measures. There are several publications using data from different countries on the effects of control measures, see, e.g., [1] , [2] and [3]. As [4] point out, many of such studies are undermined by unreliable data on incidence. Many papers use data provided from the Johns Hopkins University (JHU) [5]. These data are based on cumulative registered cases in different countries, which induces several problems, particularly the fact that not all cases are reported and that there is delay between the day of infection and the reporting day. Furthermore, the systems of reporting vary between countries, which makes comparisons between countries difficult. Therefore we focus on the analysis of the epidemic curve in Bavaria and Germany. The availability of case based data for Bavaria and detailed data for disease onset for Germany is essential for our analysis. In a recent paper on Germany by [6], the authors use a complex Bayesian modeling approach based on the daily registrations in the JHU data. An important claim by [6] is that the lock-down-like measures on March 23rd were necessary to stop exponential growth, however, this result contradicts for example results by the German RKI [7]. Furthermore, these approaches were critically questioned by [8] and [9], where the latter emphasized the importance of taking into account the delay by reporting and incubation time, when analyzing the possible effect of non pharmaceutical interventions. We follow this line of argument and use a statistical model to estimate daily numbers of infected and of persons with disease onset on a certain day. For Bavaria, we use detailed case-based data, while a modeling approach is utilized for German data. We analyze the respective epidemic curves using a segmented regression model with change points. The paper is organized as follows. In Section 2, we present the data and the the strategy of estimating the relevant daily counts. Then the segmented regression model, which is the basis for further analyses, is presented. In Section 3, we present the results followed by a discussion in Section 4.

2 Data and Methods

2.1 Estimation of diseases onset

For the analysis of the Bavarian data, we use the Covid-19 reporting data of the Bavarian State Office for Health and Food Safety (LGL), that is collected within the framework of the German Infection Control Act (IfSG). At case level, this data includes the reporting date (the date at which the case was reported to the LGL) as well as the time of disease onset (here: symptom onset). However,

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20222265; this version posted October 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the latter is not always known: partly because it could not be determined and partly because the case did not (yet) have any symptoms at the time of entry into the data base. A procedure for imputation of missing values regarding the disease onset has been developed by [10], using a flexible generalized additive model for location, scale and shape (GAMLSS; [11]), assuming a Weibull-distribution for time $t_d > 0$ between disease onset and reporting date. We estimate the delay time distribution from data with disease onset and impute missing disease onsets based on this model.

For the German data, no individual case data were available, so instead we used estimated disease onset data provided by the Robert Koch Institute (RKI), see [12] and [7]. The method used by the RKI is similar to our approach applied to Bavarian data [10].

2.2 Back-projection

To interpret the course of the epidemic and possible effects of interventions, case based data on time of infection is essential. However, as such data is generally not available, one simple approach is to shift the curve to the past by the average incubation period. The average incubation period for COVID-19 is about five days [13]. A more sophisticated approach is to use the incubation period distribution as part of an inverse convolution, also known as back-propagation, in order to estimate the number of infections per day from the time series of disease onsets [14, 15]. We assume a log-normal distribution for the incubation time with a median of 5.1. days and a 95% percentile at 11.5 days [13]. These are the same values as used by [6]. For our calculation, we use the back-projection procedure implemented in the R package `surveillance` [16].

2.3 The segmented regression model

To analyze the temporal course of the infection we use the following Poisson regression model with over-dispersion and change points (see [17], [18]):

$$E(Y_t) = \exp\left(\beta_0 + \beta_1 t + \sum_{k=1}^{K} \gamma_k (t - CP_k)_+\right), \tag{1}$$

where $E(Y_t)$ is the expected number of new reported cases at time t, K is the number of change points, and $x_+ = \max(x, 0)$ is the positive part of x. The change points are used to partition the epidemic curve Y_t into $K+1$ phases. These are characterized by different growth parameters. In the phase before the first change point CP_1 the growth is characterized by the parameter β_1, in the 2nd phase between CP_1 and CP_2 by $\beta_2 = \beta_1 + \gamma_1$. The next change is then at time CP_2. In the 3rd phase between CP_2 and CP_3 the growth parameter is given by $\beta_3 = \beta_1 + \gamma_1 + \gamma_2$. This applies accordingly until the last phase after CP_K. The quantities $\exp(\beta_j)$, $j = 1, \ldots, K+1$ can be interpreted as daily growth factors. Since model (1) is a generalized linear model given

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20222265; this version posted October 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the change points, the parameters of the model (including the change points) can be estimated by minimizing the quasi likelihood function for the Poisson model. Due to the estimation of the change points the numerical optimization problem is not straight forward. For the estimation of the model we use the R-package `segmented`, see [19]. The starting values are estimated by discrete optimization using all possible integer suitable combinations of change points. The number of change points K is varied step-wise up to a maximum of $K = 4$. It is examined whether the increase of the number of break points leads to a relevant improvement of the model fitting (over-dispersion parameter). Models with more than 4 change points since they are hardly interpretable and the danger of overfitting is high.

We apply the segmented regression model to time series of the estimated daily numbers of infections for Bavaria and Germany. Since the back propagation algorithm yields an estimate for the expected values of the number of daily infections and does so by inducing a smoothing effect, as a sensitivity analysis for the location of the breakpoints, we also apply the model to the time series of the daily number of disease onsets. When comparing the results, it should be taken into account that the onset of the disease is on average 5 days after infection.

3 Results

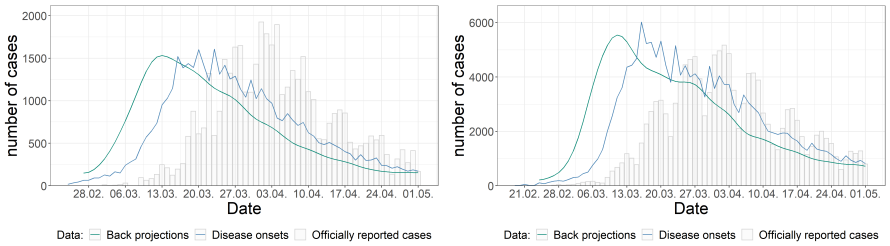

Figure 1: Comparison of daily time series of reported, disease onsets and estimated infection dates for Bavaria (left panel) and Germany (right panel).

In Figure 1, the three different time series of daily cases (reported, disease onset and estimated infection date) are presented. The delay between the three time series for Bavaria and Germany is evident. Furthermore, the curves do not just differ by a constant delay, but there is some change in structure of the curves. The curve relating to the date of infection is clearly smoothed due to the back projection procedure (cf. Section 2.2) and has a clear maximum both for Bavaria and Germany.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20222265; this version posted October 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

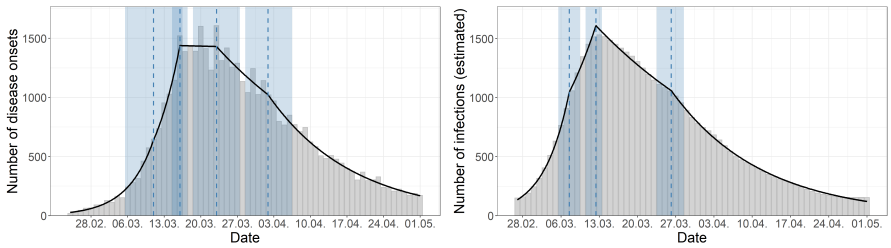

Figure 2: Segmented Poisson models for Bavaria with four estimated change points for disease onset (left) and estimated infection date (right). The solid lines are the fitted curves according to the segmented regression models. The bars are the values of the estimated daily numbers of new cases according to the disease onset model (left panel) and the estimated (cf. Section 2.2) newly infected cases (right panel). Dashed lines and surrounding shaded ribbons indicate estimated change points and respective, approximate 95% confidence intervals.

3.1 Bavarian data

For the Bavarian data on disease onset, the model with $K = 4$ change points gives the best result with an estimate of the over-dispersion parameter of 3.8, i.e., the variance of Y_t is 3.8 times higher than the value of $\mathrm{Var}(Y_t) = E(Y_t)$ otherwise expected under the assumption of the Poisson regression model. The over-dispersion for a model with $K = 3$ change points is substantially higher (4.5), which suggests that the model with four change points should be preferred. Figure 2 (left panel) depicts a graphical representation of the change point analysis. The estimated model coefficients and change points are summarized in Table 1, left panel. The model delivers five phases starting with a steep increase, which is slowed down in the second phase. In the third phase starting at 15th-17th of April, the increase is stopped and there is decrease in number of disease onsets in the fourth phase, which is accelerated in the fifth phase.

The Poisson model for the infection date gives a substantially better fit than the disease onset model. The model with four change points has an underdispersion by the the factor 0.39 compared to 0.79 for the model with three change points. Therefore, we use the model with three change points to avoid overfitting. The result can be seen in Figure 2 (right panel) and in Table 1 (right panel). The phases are similar to those for the disease onset, but the change from increasing curve (phase 1 and 2) to a decreasing phase (phase 3) is direct without a plateau in between.

Taking the mean incubation period of five days into account, we combine the results for the two models for Bavaria, which implies four phases of the epidemic:

1st phase There is a substantial increase in new infections in both models.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20222265; this version posted October 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Estimated change points for Bavaria

Change point	disease onset	infection date
1	03/11 (03/06 - 03/16)	03/08 (03/06-03/09)
2	03/16 (03/15 - 03/17)	03/12 (03/11-03/13)
3	03/23 (03/19 - 03/27)	03/26 (03/24-03/28)
4	04/02 (03/29 - 04/06)	-

Estimated multiplication factor

Phase (j)	disease onset	infection date
1	1.22 (1.19 - 1.24)	1.24 (1.23-1.25)
2	1.17 (1.14 - 1.2)	1.09 (1.08-1.11)
3	1.00 (0.97 - 1.03)	0.97 (0.97-0.98)
4	0.97 (0.95 - 0.98)	0.94 (0.94-0.95)
5	0.94 (0.94 - 0.95)	-

Table 1: Summary table of the segmented Poisson model for Bavaria with four estimated change points for disease onset. The dates of the estimated change points and the corresponding 95% confidence intervals are given. For the date of the lower/upper limit of the confidence intervals, the values were rounded up or down to the more extreme value. In the second part of the table the estimated multiplication factors of the number of cases per day with the confidence intervals for the 5 phases are given.

The first phase ends at March 7th-8th in the infection model, while the change point has a broad confidence interval in the disease onset model.

2nd phase The exponential increase slows down to a value of 1.17 (disease onset model) and 1.09 (date of infection).

3rd phase The second change point can be seen at March 10th-12th (15th-17th in the disease onset model) and at March 12th-13th in the infection model. There is a clearly visible change in the course of the epidemic in both models. It marks the turning point of the number of new infections (change form 1.09 to 0.97). For the disease onset curve, there is change from 1.17 to 1.

4th phase From March 27th onward, there is a decrease in new cases. The multiplication factor is now 0.94 (confidence interval 0.94-0.95).

The two models differ in the behavior of the curve after date of infection on March 12th. While the curve is nearly constant in the disease onset model followed by a change point on March 23rd (infection date March 18th), there is no further change point between March, 12th and March 27th in the infection date model.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20222265; this version posted October 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

3.2 German data

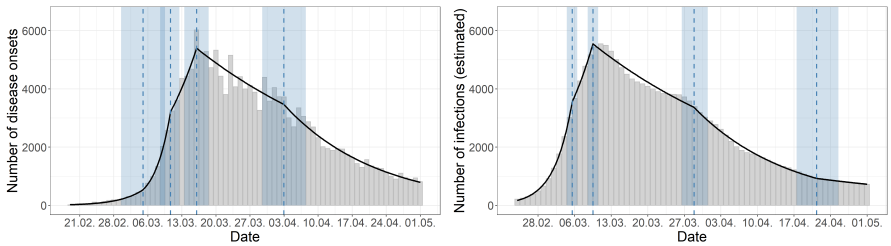

Figure 3: Segmented Poisson models for Germany with four estimated change points for disease onset (left) and estimated infection date (right). The solid lines are the fitted curves according to the segmented regression models. The bars are the values of the estimated daily numbers of new cases according to the disease onset model (left panel) and the back-projected (cf. Section 2.2) newly infected cases (right panel). Dashed lines and surrounding shaded ribbons indicate estimated change points and respective, approximate 95% confidence intervals.

The results for the German data are presented in Figure 3 and in Table 2. For the disease onset model, the overdispersion is substantially lower for the model with four change points than with three change points. However, the overdipersion of the model with four change points is rather high (20.1) and the confidence intervals for the change points are rather big, especially for the first part of the time series where two change points are estimated. While the distinction in the first three phases is unclear, there is a clear turning point between March 14th and March 18th. The fifth phase with a further slowdown starts at the end of April. As can be seen from Figure 3 (right panel) and Table 2 (right panel) the model for the estimated infection date (cf. Section 2.2) gives a much better fit with an overdispersion of 3.37 for 4 change points (5.12 for 3 change points). The first change point is on the 6th of March, where we see a first reduction of the multiplication factor from 1.3 to 1.1. The turning point of the curve is on March 9th - 10th, where the multiplication factor changes to 0.97 inducing a decrease of the curve. At the end of March 27th-30th, there is a change to a multiplication factor of 0.94. The last estimated change point at the end of April has a large confidence interval and marks the beginning of a phase of almost constant infection numbers.
Overall, the models for Germany indicate three consistent phases of the epidemic:

1st phase There is a strong increase in new infections in the beginning. With high multiplication factors in both models. In the infection model we see a first change point at March 5th-6th. However, there is no clear change point in the disease model. The first clear change point, which can be seen in both models is on March 9th-10th (change point 2 in the infection

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20222265; this version posted October 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Estimated change points for Germany		
change point	disease onset	infection date
1	03/05 (03/01 - 03/09)	03/06 (03/05 - 03/06)
2	03/11 (03/09 - 03/12)	03/09 (03/09 - 03/10)
3	03/16 (03/15 - 03/17)	03/29 (03/27 - 03/31)
4	04/03 (03/30 - 04/07)	04/21 (04/18 - 04/25)

Estimated multiplication factor		
Phase (j)	disease onset	infection date
1	1.23 (1.17 - 1.29)	1.34 (1.32 - 1.36)
2	1.37 (1.27 - 1.48)	1.12 (1.09 - 1.15)
3	1.10 (1.07 - 1.14)	0.97 (0.97 - 0.98)
4	0.98 (0.97 - 0.98)	0.95 (0.94 - 0.95)
5	0.95 (0.94 - 0.95)	0.98 (0.96 - 0.99)

Table 2: Summary table of the segmented Poisson model for Germany with four estimated change points for disease onset. The dates (first time point corresponds to February, 24 2020) of the estimated change points and the corresponding 95% confidence intervals are given. For the date of the lower/upper limit of the confidence intervals, the values were rounded up or down to the more extreme value. In the second part of the table the estimated multiplication factors of the number of cases per day with the confidence intervals for the 5 phases are given.

model) and March 10th-12th (change point 3 in the disease onset model), respectively.

2nd phase After these change points, there is a clearly visible change in the course of the epidemic. There is change from an increasing to a decreasing curve.

3rd phase In both models there is a further change point from March 27th onward, where there is a decrease in new cases. The multiplication factor is now 0.94 (confidence interval 0.939-0.95).

There are further estimated change points in both models, which have wide confidence intervals and do not fit well together in the models.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20222265; this version posted October 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

4 Discussion

The present analysis is a retrospective, exploratory analysis of the German and Bavarian COVID-19 reporting data during Mar-Apr 2020.

4.1 Limitations

The analysis does not include cases that have not been recorded. If the proportion of undetected cases changes over time, this can distort the curve and thus the determination of the change points. Therefore, additional data on daily deaths and hospital admissions and the number of tests performed should be considered. Furthermore, it is possible to estimate the proportion of undetected cases with the help of representative studies such as the one currently conducted in Munich, see [20].

Our analysis is based to a considerable extent on imputed data, see [10], which is a results of missing data w.r.t. the disease onset. Since changes in behavior do not occur abruptly, the assumption of change points is also problematic in itself. Therefore, the interpretation of change points should always be done in conjunction with a direct observation of the epidemic curve.

4.2 Interpretation of Results

Our analysis is based on the onset of the disease (more precisely: the onset of symptoms) and a back projection to the date of infections, and therefore, despite its limitations, is better suited to describe the course of the epidemic than the more common analysis of daily or cumulative reported case numbers.

In the analysis of the Bavarian and the German data in different settings, the main result is the change point, where the exponential growth was stopped and is clearly visible between March 9th and 13th. The timing of this change point coincides to the implementation of the first control measures: the partial ban of mass events with more than 1000 people. Furthermore, in a press conference on March 11th chancellor Merkel and the president of the RKI appealed to self-enforced social distancing (`https://www.bundesgesundheitsministerium.de/en/coronavirus/chronologie-coronavirus.html`.

Furthermore, the extended media coverage from Bergamo, Italy, as well as the voluntary transition to home-office work could be related to this essential change in the course of the pandemic.

In Bavaria and in Germany, the change point at the end of March of infection date is apparent. This change point is associated with different measures taken in March (closing of schools and stores on March 16th and the shutdown including contact ban on March 21st). Since there were many measures administered simultaneously, it is not possible to attribute individual measures to the development of the epidemic curve.

The claim by [6], that the shutdown on March 21st was necessary to stop the growth of the epidemic is not supported by our analysis. There is a change

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20222265; this version posted October 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

point in the epidemic curve after that date, but the major change from an exponential growth to a decrease was before the shutdown. The difference in results can be explained by the different data bases used for the respective analyses. While [6] use data based on daily registered cases, in our analysis, data on disease onset are included. As can be seen from Figure 1 and from the results of our data analysis, the delay distribution of the time between disease onset and reporting day changed over time. Using this information is a crucial difference between our analysis and that of [6]. In a recent technical addendum [21] the authors re-fit their model on more appropriate data. These analysis – in our opinion – clearly show that the effective reproduction number decreased earlier than in their initial analysis, however, they attribute the decrease to a SIR model peculiarity, where a linear decrease in the contact rate can lead to the incidence curve dropping despite $R(t) > 1$.

The above discussions illustrate how complex the interpretation of even simple SIR models is and the question is, if such SIR modeling is not too simple to really allow for questions to be answered model based (no age structure, no time varying reporting delay, no incubation delay). In contrast, our approach is more data driven with a minimum of modeling assumptions and without the need to include strong prior information about the change points. Directly using a segmented curve with exponential growth (decline) is in line with common models of infectious diseases in its early stages, where the limitation of the spread by immune persons plays no role. The problem of using complex models with many parameters for the evaluation of governmental measures has also been highlighted by [22].

Our approach is similar to that of [9]. However, using change point analysis for variables derived from daily new infections appears problematic, since assumptions modeling the reproduction number $R(t)$ or the cumulative numbers are questionable. More specifically, the use of the time-varying reproduction number $R(t)$, a standard measure to describe the course of an epidemic is challenging, as different definitions have been proposed in the literature that also imply different interpretations (see [23, 24]). However, the analysis of R(t) as a relative measure can be useful, when one wants to analyze data from different countries with non comparable reporting systems, see [3].

We prefer the direct use of a Poisson regression model with a more plausible assumptions about the error terms instead of using OLS for logarithmic case numbers. Furthermore, we apply a direct maximum likelihood estimation of the change points of the segmented regression model. For the interpretation of the model based on disease onsets, we also use a simple difference of five days to take the incubation time into account. Our results are similar to that of [9]. However, the claim that there is no evidence for the effect of governmental measures is not supported by our analysis.

Our result is in line with that of [25], where a stop of exponential growth in Great Britain has been before the shutdown. Furthermore, the effect of governmental measures as whole is clearly documented in the literature, see, e.g., [1] and [2].Our result on a possible effect of the ban of mass events is also in line with the results of [3].

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20222265; this version posted October 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The temporal connection between the change points in our analysis and various control measures should be interpreted as an association, rather than a direct causal relationship. In the end many other explanations exists and from a simple time series analysis it is not possible to say to what extent the population already had changed their behavior voluntarily, as for example observed in mobility data [26], and in what way the measures contributed to this. More speculative alternative explanations would include the possibility of a seasonal effect on coronavirus activity (e.g. related to temperature) or changes in test capacity or the case detection ratio. However, given the re-emergence of the epidemic in the fall of 2020 at high test capacity and at relatively high temperatures shows that contact behavior is the major explanatory factor for virus activity. Nevertheless, any analysis of observational time series data including only a limited amount of explanatory factors has to be interpreted with care and with respect to the many uncertainties which remain regarding COVID-19 [27].

Despite the limitations of the approach, we argue that it is advantageous and important to directly interpret the epidemic curve and the absolute number of cases, rather than indirect measures like the $R(t)$. Furthermore, the reproduction rate does not contain information about how many people are currently affected, or whether the infected persons belong to risk groups. The course of the time-varying reproduction number calculated by us for Bavaria fits well with the change point analysis [10]. A value of $R(t) > 1$ corresponds to a rate of increase >1, noting that the time delays in the interpretation of $R(t)$ must be kept in mind.

It should be noted, that the presented analysis is retrospective. Control measures have to be decided based on a completely different level of information than what the retrospectively established epidemic curve suggests. The simple observation of the course of the reported case numbers by reporting date is also problematic because this course is strongly influenced by the reporting behavior and the methods and capacities of the test laboratories. Typically, substantially fewer cases are reported at weekends than during the week. Therefore, the estimation [10] is an important step to estimate the better interpretable curve of new cases, but is limited by assumptions and limitations itself, that need to be considered when interpreting the results.

Since the impact of the measures also depends on how they are implemented by the population (compliance), the results cannot be directly transferred to the future. Nevertheless, it remains a remarkable result that the clear turning point of the early COVID-19 infection data in Germany is associated with non drastic measures (no shutdown) and strong appeals by politicians.

Data availability

All data used for the analyses and all code to reproduce the models, figures and tables in the manuscript are openly and freely available from `https://github.com/adibender/covid19-changepoint-analysis-germany-bavaria`. All

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20222265; this version posted October 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

analyses were performed using the R programming language [28]. Figures were created using R package `ggplot2` [29].

Acknowledgements

We would like to thank Katharina Katz and Manfred Wildner from the Bavarian State Office for Health and Food Safety (LGL) for providing the data and for useful discussions. We also thank Nadja Sauter for help with visualizations.

References

[1] Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. NATURE. 2020 8;584(7820):257+.

[2] Islam N, Sharp SJ, Chowell G, Shabnam S, Kawachi I, Lacey B, et al. Physical distancing interventions and incidence of coronavirus disease 2019: natural experiment in 149 countries. BMJ-BRITISH MEDICAL JOURNAL. 2020 7;370.

[3] Li Y, Campbell H, Kulkarni D, Harpur A, Nundy M, Wang X, et al. The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (R) of SARS-CoV-2: a modelling study across 131 countries. The Lancet Infectious Diseases. 2020. Available from: `http://www.sciencedirect.com/science/article/pii/S1473309920307854`.

[4] Floyd TM, Endowed JR. Lockdown-type measures look effective against covid-19 But evidence is undermined by unreliable data on incidence. BMJ-BRITISH MEDICAL JOURNAL. 2020 7;370.

[5] Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases. 2020;20(5):533 – 534.

[6] Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science. 2020.

[7] Robert-Koch Institut. Nowcasting; 2020. Accessed: 2020-09-23. Available from: `"https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Projekte_RKI/Nowcasting.html"`.

[8] Bryant P, Elofsson A. The limits of estimating COVID-19 intervention effects using Bayesian models. medRxiv. 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20222265; this version posted October 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[9] Wieland T. A phenomenological approach to assessing the effectiveness of COVID-19 related nonpharmaceutical interventions in Germany. SAFETY SCIENCE. 2020 11;131.

[10] Günther F, Bender A, Katz K, Küchenhoff H, Höhle M. Nowcasting the COVID-19 pandemic in Bavaria. medRxiv. 2020.

[11] Stasinopoulos MD, Rigby RA, Heller GZ, Voudouris V, De Bastiani F. Flexible regression and smoothing: using GAMLSS in R. The R series. Boca Raton London New York: CRC Press, Taylor & Francis Group, a Chapman & Hall book; 2017. OCLC: 987291390.

[12] an der Heiden M, Hamouda O. Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland – Nowcasting. Epid Bull. 2020;17:10–15.

[13] Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Annals of Internal Medicine. 2020 03.

[14] Becker NG, Watson LF, Carlin JB. A method of non-parametric back-projection and its application to AIDS data. Statistics in Medicine. 1991;10:1527–1542.

[15] Werber D, King L, Müller L, Follin P, Bernard H, Rosner B, et al. Associations of Age and Sex on Clinical Outcome and Incubation Period of Shiga toxin-producing Escherichia coli O104:H4 Infections, 2011. American Journal of Epidemiology. 2013;178(6):984–992.

[16] Salmon M, Schumacher D, Höhle M. Monitoring Count Time Series in R: Aberration Detection in Public Health Surveillance. Journal of Statistical Software. 2016;70(10). Also available as vignette of the R package surveillance.

[17] Muggeo VMR. Estimating regression models with unknown breakpoints. Statistics in Medicine. 2003 10;22(19):3055–3071.

[18] Muggeo V, Sottile G, Porcu M. Modelling COVID-19 outbreak: segmented regression to assess lockdown effectiveness; 2020.

[19] Muggeo VMR. segmented: an R Package to Fit Regression Models with Broken-Line Relationships. R News. 2008;8(1):20–25.

[20] Radon K, Saathoff E, Pritsch M, Noller JMG, Kroidl I, Olbrich L, et al. Protocol of a population-based prospective COVID-19 cohort study Munich, Germany (KoCo19) (vol 20, 1036, 2020). BMC PUBLIC HEALTH. 2020 9;20(1).

medRxiv preprint doi: https://doi.org/10.1101/2020.10.29.20222265; this version posted October 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

[21] Dehning J, Spitzner FP, Linden M, Sebastian BM, Neto JP, Zierenberg J, et al.. Model-based and model-free characterization of epidemic outbreaks - Technical notes on Dehning et al., Science, 2020; 2020. Available from: https://raw.githubusercontent.com/Priesemann-Group/covid19_inference_forecast/master/technical_notes_dehning_etal_2020.pdf.

[22] Bryant P, Elofsson A. The limits of estimating COVID-19 intervention effects using Bayesian models. medRxiv. 2020.

[23] Cori A, Ferguson NM, Fraser C, Cauchemez S. A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics. American Journal of Epidemiology. 2013 Nov;178(9):1505–1512. Publisher: Oxford Academic.

[24] Lipsitch M, Joshi K, Cobey S. Comment on Pan A, Liu L, Wang C, et al. Association of Public Health Interventions With the Epidemiology of the COVID-19 Outbreak in Wuhan, China. JAMA.; 2020. Accessed 22.4.2020. https://github.com/keyajoshi/Pan_response.

[25] Wood SN. Did COVID-19 infections decline before UK lockdown ?; 2020.

[26] Bryant P, Elofsson A. Estimating the impact of mobility patterns on COVID-19 infection rates in 11 European countries. PeerJ. 2020 Sep;8:e9879.

[27] Davey Smith G, Blastland M, Munafò M. Covid-19's known unknowns. BMJ. 2020;371.

[28] R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2020. Available from: https://www.R-project.org/.

[29] Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016.