Biometrical Journal

RESEARCH PAPER

Nowcasting fatal COVID-19 infections on a regional level in Germany

Marc Schneble[1] | **Giacomo De Nicola[1]** | **Göran Kauermann[1]** | **Ursula Berger[2]**

[1] Department of Statistics, Ludwig-Maximilians-University Munich, Munich, Germany

[2] Institute for Medical Information Processing, Biometry, and Epidemiology, Ludwig-Maximilians-University Munich, Munich, Germany

Correspondence

Marc Schneble, Department of Statistics, Ludwig-Maximilians-University Munich, Ludwigstr. 33, 80539 Munich, Germany.
Email: marc.schneble@stat.uni-muenchen.de

Abstract

We analyse the temporal and regional structure in mortality rates related to COVID-19 infections, making use of the openly available data on registered cases in Germany published by the Robert Koch Institute on a daily basis. Estimates for the number of present-day infections that will, at a later date, prove to be fatal are derived through a nowcasting model, which relates the day of death of each deceased patient to the corresponding day of registration of the infection. Our district-level modelling approach for fatal infections disentangles spatial variation into a global pattern for Germany, district-specific long-term effects and short-term dynamics, while also taking the age and gender structure of the regional population into account. This enables to highlight areas with unexpectedly high disease activity. The analysis of death counts contributes to a better understanding of the spread of the disease while being, to some extent, less dependent on testing strategy and capacity in comparison to infection counts. The proposed approach and the presented results thus provide reliable insight into the state and the dynamics of the pandemic during the early phases of the infection wave in spring 2020 in Germany, when little was known about the disease and limited data were available.

KEYWORDS

COVID-19, disease mapping, generalized regression model, nowcasting

1 | INTRODUCTION

In March 2020, COVID-19 became a global pandemic. From Wuhan, China, the virus spread across the whole world, and with its diffusion more and more data became available to scientists for analytical purposes. In daily reports, the WHO provides the number of registered infections as well as the daily death toll globally (https://www.who.int/). It is inevitable for the number of registered infections to depend on the testing strategy in each country (see, e.g., Cohen & Kupferschmidt, 2020). This has a direct influence on the number of undetected infections (see, e.g., Li et al., 2020), and first empirical analyses aim to quantify how detected and undetected infections are related (see, e.g., Niehus, De Salazar, Taylor, & Lipsitch, 2020). Though similar issues with respect to data quality hold for the reported number of fatalities

(see, e.g., Baud et al., 2020), the number of deaths can overall be considered a more reliable source of information than the number of registered infections. The results of the 'Heinsberg study' in Germany point in the same direction (Streeck et al., 2020). A thorough analysis of death counts can in turn generate insights on changes in infections as proposed in Flaxman et al. (2020) (see also Ferguson et al., 2020). In this paper, we pursue the idea of directly modelling registered death counts related to COVID-19 instead of registered infections. In other words, we restrict our analysis to fatal COVID-19 cases only, omitting recovered or symptom-free infections. We analyse data from Germany and break down the analyses to a regional level. Such regional view is apparently immensely important, considering the local nature of some of the outbreaks, for example in Italy (see, e.g. Grasselli, Pesenti, & Cecconi, 2020; Grasselli, Zangrillo, & Zanella, 2020), France (see, e.g., Massonnaud, Roux, & Crépey, 2020) or Spain and can assist local health authorities in monitoring the disease and planning infection control measures.

The analysis of fatalities has, however, an inevitable time delay and requires to take the course of the disease of COVID-19 patients into account. In particular, in this paper we consider the timespan between the registration of the infection through local health authorities and the report of its deadly outcome by the Robert Koch Institute (RKI). A first approach on modelling and analysing the time from illness and onset of symptoms to reporting and further to death is given in Jung et al. (2020) (see also Linton et al., 2020). Understanding the delay between onset and registration of an infection and, for severe cases, the time between registered infection and death, can be of vital importance. Knowledge on those timespans allows us to obtain estimates for the number of infections that are expected to be fatal based on the number of infections registered on the present day. The statistical technique to obtain such estimates is called nowcasting (see, e.g., Höhle & an der Heiden, 2014) and traces back to Zeger, See, and Diggle (1989) or Lawless (1994). Nowcasting in COVID-19 data analyses is not novel and is, for instance, used in Günther, Bender, Küchenhoff, Katz, and Höhle (2020) for nowcasting daily infection counts in Germany, that is to adjust daily reported new infections to include infections which occurred the same day but were not yet reported. Altmeijd, Rocklöv, and Wallin (2020) apply nowcasting techniques to Swedish data and Bird and Nielsen (2020) provide nowcasting fatalities in English hospitals. We extend this approach to model the duration between the registration date of an infection and its fatal outcome, accounting for additional covariates. To do so, we combine a nowcasting model with a spatio-temporal regression model.

We analyse the number of fatal cases of COVID-19 infections in Germany using district-level data. The data are provided by the RKI (www.rki.de), the German federal government agency and scientific institute responsible for health reporting, disease control and prevention in humans. They report the cumulative number of deaths in different gender and age groups for each of the 412 administrative districts in Germany, together with the date of registration of the infection. The data are available in dynamic form through daily downloads of the updated cumulated numbers of deaths. Comparing two consecutive daily downloads allows to construct a new dataset which contains both the date at which a COVID-19 disease is registered and the date at which a fatality is reported to the RKI, with the latter usually being reported at a later time point. We employ flexible statistical models with smooth components (see, e.g., Wood, 2017), assuming the district-specific number of fatalities to be negatively binomial distributed, which permits to also account for possible overdispersion in the data. The spatial structure in the death rate is incorporated in two ways: First, we assume a spatial correlation of the number of deaths by including a long-range smooth spatial death intensity. This allows to map a general pattern of the spread of the disease over Germany, which shows that regions of Germany are affected to different extents. On top of this long-range effect, we include two types of unstructured region-specific effects. An overall region-specific effect reflects the situation of a district as a whole, while a short-term effect mirrors region-specific variations of fatalities over time and captures local outbreaks as happened in, for example Heinsberg (North-Rhine-Westphalia) or Tirschenreuth (Bavaria). This effect can be seen as an unstructured time-space interaction. In addition to the spatial components, we include an overall temporal effect to capture dynamic changes in the number of fatal infections for Germany. The latter effect mirrors the overall flattening of the infectious situation in the considered time period, that is spring 2020. Besides the spatio-temporal character, our modelling approach further adjusts for the district-specific age and gender structure.

Modelling infectious diseases is a well-developed field in statistics, and we refer to Held, Meyer, and Bracher (2017) for a general overview of the different models. We also refer to the powerful R package `surveillance` (Meyer, Held, & Höhle, 2017). Since our focus is on analysing district-specific dynamics, both structured and unstructured, as well as dynamic behaviour of fatal infections, we prefer to make use of generalized additive regressions implemented in the `mgcv` package in R, which also allows to decompose the spatial component in more depth.

The paper is organized as follows. In Section 2 we describe the data. Section 3 introduces our model, while Section 4 discusses the necessity of incorporating a nowcasting model. Section 5 shows the results of our analysis which are then refined to subgroups of the data in Section 6. Section 7 concludes the paper by also discussing the limitations of our modelling exercise.

TABLE 1 Illustration of the data structure, showing downloads of the data from April 25 and April 26, 2020 as an example. To facilitate reproducibility, the original column names used in the RKI datasets are given in brackets below our English notation

		District (Landkreis)	Age Group (Altersgruppe)	Gender (Geschlecht)	Infections (Anzahl Fall)	Fatal Infections (Anzahl Todesfall)	Registration Date (Meldedatum)	Reporting Date (Datenstand)
Data downloaded on April 25, 2020	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		Munich City	60–79	F	3	1	April 22, 2020	April 25, 2020
		Munich City	60–79	M	5	1	April 22, 2020	April 25, 2020
		⋮	⋮	⋮	⋮	⋮	⋮	⋮
Data downloaded on April 26, 2020		District	Age Group	Gender	Infections	Fatal Infections	Registration Date	Reporting Date
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
		Munich City	60–79	F	6	2	April 22, 2020	April 26, 2020
		Munich City	60–79	M	5	1	April 22, 2020	April 26, 2020
		⋮	⋮	⋮	⋮	⋮	⋮	⋮

2 | DATA

We make use of the COVID-19 dataset (Esri Deutschland GmbH, 2020) provided by the RKI on a daily basis for the 412 districts in Germany (which also include the 12 districts of Berlin separately). The data are collected by the RKI, but originate from the district-based health authorities (*Gesundheitsämter*). Due to different population sizes in the districts, and certainly also because of different local situations, some health authorities transmit the daily numbers to the RKI with a delay. This happens in particular over the weekend, a fact that we need to take into account in our model. We have daily downloads of the data since March 27, 2020. We here choose to focus on a phase of the COVID-19 pandemic in which the death toll in Germany was high. The subsequent analysis was thus conducted with data up to May 14, 2020, and was performed considering only deadly infections with registration dates from March 26, 2020 until May 13, 2020 (the day before that of the analysis).

Table 1 illustrates an exemplary extract of the data that are available. For each of the 412 districts, the data contain the cumulated number of laboratory-confirmed COVID-19 infections as well as the cumulated number of deaths related to COVID-19 for each district of Germany, stratified by age group (15–34, 35–59, 60–79 or 80+), gender, and the date of registration of the infection by the local public health authorities. The time stamp for a fatal outcome always refers to the registration date of the infection and *not* to the individual's date of death. Therefore, the numbers in the column 'Fatal infections' cannot exceed the numbers shown in the column 'Infections'. Even though the time point of infection obviously precedes that of death, registration of an infection can also occur after death, for example when a post-mortem test is conducted, or when test results arrive after the patient has passed away. In the former case, the registration date are set to the day of death by the local health authority. Also note that it is not indicated in the dataset whether a fatal infection resulted from a post-mortem test, and that no information on whether the patient has died *with* or *because* of a COVID-19 infection is included.

The cumulative numbers are reported on a daily basis by the RKI, which is mirrored in the column 'Reporting date' in Table 1. The reporting date always corresponds to the query date and the download date of the data. In Table 1, we see that the number of reported infections with registration date April 22, 2020, which relate to females in the age group 60–79 living in the city of Munich, increases by three from April 25, 2020 to the following day. In the same period, the number of fatal infections increased by one. Thus, we can deduce that three registered infections in this sub-population were reported with a delay of 4 days. The single newly reported fatal infection belongs to an individual of this sub-population for which the time between registration by the local health authorities and reported death amounts to 4 days. In this paper, we are especially interested in the latter quantity, which we model as a duration time. It is of importance to note that we can derive such information only due to daily downloads of the dataset, which are not being provided retrospectively.

We refrain from providing general descriptive statistics on the spatio-temporal distribution of confirmed COVID-19 infections here, since these numbers are already visualized on the RKI dashboard (Robert Koch-Institut, 2020; see also StaBLab, LMU Munich, 2020). However, the number of fatal infections is less often taken into account. Thus, in Figure 1 we show the empirical duration between the day of registration as COVID-19 infected by the local health authorities and the day on which the death has been reported by the RKI (based on the data until May 14, 2020). Due to the aforementioned reporting delay, the minimum duration is 1 day. Note that these plots show stapled bar charts, highlighting the counts by gender. We see that considerably more fatal infections originate from the age group 80+. Regarding the age group 80− (aggregated age groups 15–39, 40–59 and 60–79), we see that males are much more affected than females, whereas in the age group 80+ the counts are more balanced. Finally, in both age groups there are a small number of deaths,
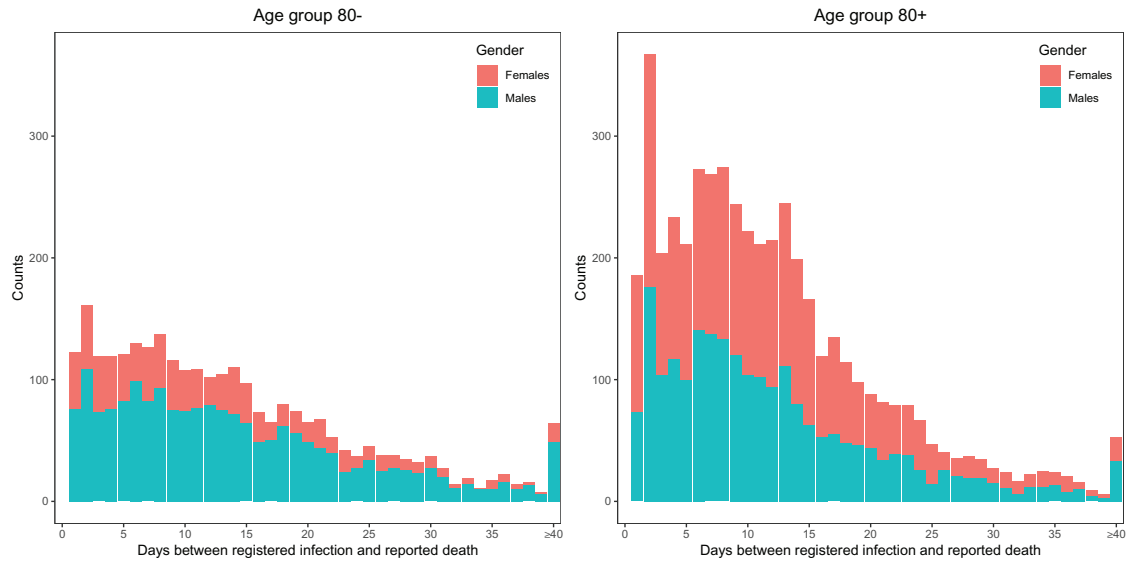
FIGURE 1 Stapled bar chart of the counts of fatal infections depending on days between registered infection and reported death. Only data reported until May 14, 2020 is considered here (left panel: age group 80− (less than 80 years), right panel: age group 80+ (80 years or older)
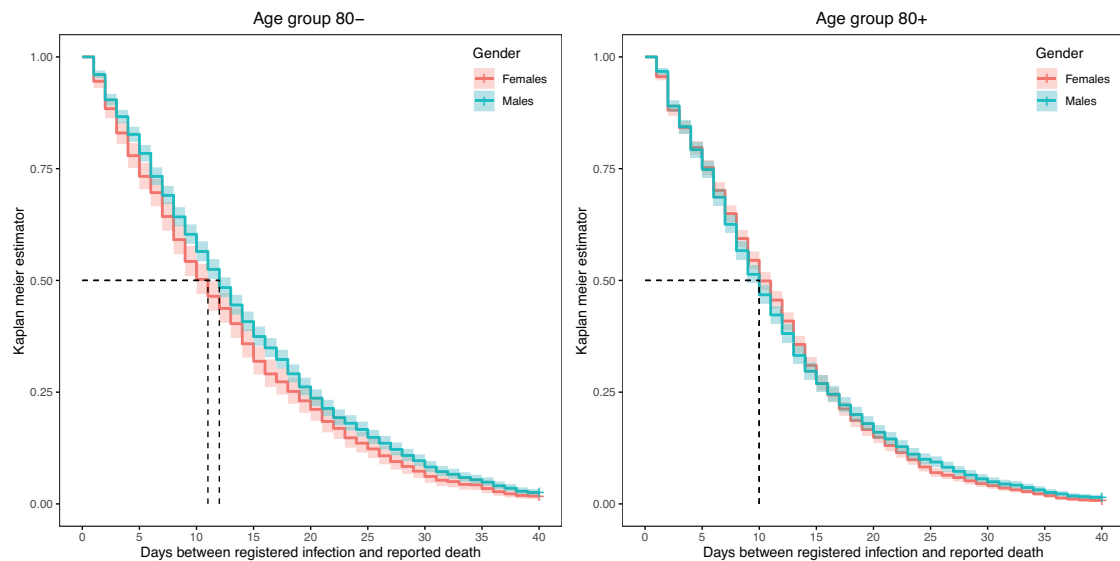

FIGURE 2 Kaplan–Meier estimators of the data shown in Figure 1 with 95% confidence intervals

which were reported 40 or more days after the registration of the COVID-19 infection. Kaplan–Meier estimators of the duration between registered infection and reported death are shown in Figure 2 for age groups 80− and 80+ by gender. Here we especially see that the median duration time of elderly patients is slightly shorter when compared to the younger age groups.

3 | MORTALITY MODEL

Let $Y_{t,r,c}$ denote the number of deaths due to COVID-19 with time point of registration $t = 0, \ldots, T$ in district/region r and cohort c, where the cohort c is characterized by age group and gender of the deceased. Time index $t = T$ corresponds to the day of analysis, which is May 14, 2020, and $t = 0$ corresponds to March 26, 2020. Not all fatalities with registered infection at time point t have been observed at time T, as some deaths will occur later. We therefore need a model for nowcasting, which is discussed in the next section.

For now, we assume all $Y_{t,r,c}$ to be known. A family of discrete distributions which is supported on the set of nonnegative integers and also allows to account for possible overdispersion in the data is the negative binomial distribution. Therefore, we model those numbers as according to

$$Y_{t,r,c} \sim \text{NB}(\lambda_{t,r,c}, \phi), \tag{1}$$

where $\mathbb{E}(Y_{t,r,c}) = \lambda_{t,r,c}$ and the constant dispersion parameter ϕ relates to the variance by $\text{Var}(Y_{t,r,c}) = \lambda_{t,r,c} + \phi\lambda_{t,r,c}^2$. We model the mean $\lambda_{t,r,c}$ of the response $Y_{t,r,c}$ through a regression model and specify

$$\begin{aligned}
\lambda_{t,r,c} = \exp\{&\beta_0 + \text{age}_c\beta_{\text{age}} + \text{gender}_c\beta_{\text{gender}} \\
&+ \text{age}_c\text{gender}_c\beta_{\text{age, gender}} + \text{weekday}_t\beta_{\text{weekday}} \\
&+ m_1(t) + m_2(s_r) + u_{r0} + \mathbb{1}_{\{t \geq T-14\}}u_{r1} + \log(\text{pop}_{r,c})\},
\end{aligned} \tag{2}$$

where the linear predictor is composed as follows:

- β_0 is the intercept.
- β_{age} and β_{gender} are the age- and gender-related regression coefficients, and $\beta_{\text{age, gender}}$ is the coefficient that models the interaction between age and gender.
- β_{weekday} are regression coefficients, which relate to the weekday of the registration date as COVID-19 infected.
- $m_1(t)$ is an overall smooth time trend, with no prior structure imposed on it.
- $m_2(s_r)$ is a smooth spatial effect, where s_r is the geographical centroid of district/region r.
- u_{r0} and u_{r1} are district-/region-specific random effects, which are independently and identically distributed (i.i.d.) and follow a normal prior probability model. While u_{r0} specifies an overall level of the death rate for district r over the entire observation time, u_{r1} is a spatio-temporal effect that reveals region-specific dynamics by allowing the regional effects to differ for the last 14 days.
- $\text{pop}_{r,c}$ is the gender and age group-specific population size in district/region r and serves as an offset in our model.

We here emphasize that we fit spatial effects of different types: We model a smooth spatial effect, that is $m_2(s_r)$, which takes the correlation between the fatal infections of neighbouring districts/regions into account and gives a global overview of the spatial distribution of fatal infections. In addition to that we also have unstructured district-/region-specific effects $\boldsymbol{u}_r = (u_{r0}, u_{r1})^\top$, which capture local behaviour related to single districts only. While u_{r0} captures the corresponding long-term effect, u_{r1} captures the short-term effect of the last 14 days; see (2). This means that we also model a dichotomous and unstructured interaction of space with time. The district-specific effects $\boldsymbol{u}_r$ are considered as random, with prior structure

$$\boldsymbol{u}_r \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_u) \text{ i.i.d} \tag{3}$$

for $r = 1, \ldots, 412$. The prior variance matrix $\boldsymbol{\Sigma}_u$ is estimated from the data. The predicted values $\hat{\boldsymbol{u}}_r$ (i.e. the posterior mode) exhibit districts that show unexpectedly high or low death tolls when adjusted for the global spatial structure and for age- and gender-specific population sizes.

While model (2) is complex and highly structured, note that no autoregressive components are included in the linear predictor in (2). We will demonstrate in Section 6.4 below that auto-correlation is of negligible size, and that time dependence is fully captured by $m_1(t)$ as well as the unstructured effects u_{r1}.

The mortality model defined through (1) and (2) belongs to the model class of generalized additive mixed model (see, e.g., Wood, 2017). The smooth functions are estimated by penalized splines without restrictions on the number of degrees of freedom, with a quadratic penalty that can be comprehended as a normal prior (see, e.g., Wand, 2003). The same type of prior structure holds for the region-specific random effects $\boldsymbol{u}_r$. In other words, smooth estimation and random effect estimation can be accommodated in one fitting routine, which is implemented in the R package mgcv. This package has been used to fit the model, so that no extra software implementation was necessary. This demonstrates the practicability of the proposed method. Our analysis is completely reproducible, with code and data openly available and downloadable from our GitHub repository.[1]

---

[1] https://github.com/MarcSchneble/Nowcasting-Fatal-COVID-19-Infections

4 | NOWCASTING MODEL

4.1 | Model description

The above model cannot be fitted directly to the available data, since we need to take the course of the disease on the individual level into account. This means that the final number of fatal outcomes for infections registered on date $t < T$ is not known at the time point of analysis $t = T$, since not all patients with a fatal outcome of the disease have died yet. This requires the implementation of nowcasting. Due to the sparsity of the data, we perform the nowcast on a national level, that is we cumulate the numbers over district/region r. For reasons of notation, we temporarily drop the gender and age-related subscript g, and we simply notate the cumulated number of deaths with registered infections at day t with Y_t.

Let $N_{t,d}$ denote the number of deaths reported on day $t + d$ for infections registered on day t. Assuming that the true date of death is at $t + d$, or at least close to it, we ignore any time delays between time of death and its notification to the health authorities. We call d the duration in days between the registration date as a COVID-19 patient and the reported day of death, where $d = 1, \ldots, d_{\max}$. Here, $d_{\max}$ is a fixed reasonable maximum duration, which we set to 40 days (see, e.g., Wilson, Kvalsvig, Barnard, & Baker, 2020). This is also motivated by the means of Figure 1. The minimum duration is one day, since the RKI daily reports the new numbers, which they have received from the public health departments the day before. In nowcasting, we are interested in the cumulated number of deaths for infections registered on day t, which we define as

$$Y_t = \sum_{d=1}^{d_{\max}} N_{t,d}.$$

Therefore, the total number of deaths with a registered infection at t becomes available only after $d_{\max}$ days. In other words, only after $d_{\max}$ days we know exactly how many deaths occurred due to an infection which was registered on day t. We define the partial cumulated sum of deaths as

$$C_{t,d} = \sum_{l=1}^{d} N_{t,l}$$

so that by definition $C_{t,d_{\max}} = Y_t$.

On day $t = T$, when the nowcasting is performed, we are faced with the following data constellation, where NA stands for not (yet) available:

			d		
t	1	2	$\cdots$	$d_{\max}$	Reported deaths
0	$N_{0,1}$	$N_{0,2}$	$\cdots$	$N_{0,d_{\max}}$	Y_0
1	$N_{1,1}$	$N_{1,2}$	$\cdots$	$N_{1,d_{\max}}$	Y_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$T - d_{\max}$	$N_{T-d_{\max},1}$	$N_{T-d_{\max},2}$	$\cdots$	$N_{T-d_{\max},d_{\max}}$	$Y_{T-d_{\max}}$
$T - d_{\max} + 1$	$N_{T-d_{\max}+1,1}$	$N_{T-d_{\max}+1,2}$	$\cdots$	NA	$C_{T-d_{\max}-1,d_{\max}-1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$T - 2$	$N_{T-2,1}$	$N_{T-2,2}$	NA	NA	$C_{T-2,2}$
$T - 1$	$N_{T-1,1}$	NA	NA	NA	$C_{T-1,1}$

We may consider the timespan between registered infection and (reported) death as a discrete duration time taking values $d = 1, \ldots, d_{\max}$. Let D be the random duration time, which by construction is a multinomial random variable. In principle, for each death we can consider the pairs (D_i, t_i) as i.i.d. and we aim to find a suitable regression model for D_i given t_i, including potential additional covariates $x_{t,d}$. We make use of the sequential multinomial model (see Agresti, 2010) and define

$$\pi(d; t, x_{t,d}) = P(D = d | D \leq d; t, x_{t,d}).$$

Let $F_t(d)$ denote the corresponding cumulated distribution function of D which relates to probabilities $\pi()$ through

$$
\begin{aligned}
F_t(d) = \mathrm{P}_t(D \leq d) &= \mathrm{P}(D \leq d | D \leq d + 1) \cdot \mathrm{P}(D \leq d + 1) \\
&= (1 - \pi(d + 1; \cdot)) \cdot (1 - \pi(d + 2; \cdot)) \cdot \ldots \cdot (1 - \pi(d_{\max}; \cdot)) \\
&= \prod_{k=d+1}^{d_{\max}} (1 - \pi(k; \cdot))
\end{aligned}
\tag{4}
$$

for $d = 1, \ldots, d_{\max} - 1$ and $F_t(d_{\max}) = 1$.

We generalize notation again by including the subscript g, which in the nowcasting model only distinguishes between the two age groups 80− and 80+. The available data on cumulated death counts now allow us to estimate the conditional probabilities $\pi(d;)$ for $d = 2, \ldots, d_{\max}$. In fact, the sequential multinomial model allows to look at binary data such that

$$
N_{t,d,c} \sim (\text{quasi-})\mathrm{Binomial}\big(C_{t,d,c}, \pi(d; t, c, x_{t,d})\big)
\tag{5}
$$

with

$$
\mathrm{logit}(\pi(d; t, c, x_{t,d})) = s_1(t) + s_2(d) + s_3(d) \cdot \mathbb{1}_{\mathrm{age}}\{80+\} + x_{t,d}\gamma,
\tag{6}
$$

where

- $s_1(t)$ is an overall smooth time trend over calendar days.
- $s_2(d)$ is a smooth duration effect, capturing the course of the disease.
- $s_3(d)$ is a varying smooth duration effect, capturing interaction between the dynamics of the disease and age, particularly for the age group 80+. Note that with effect $s_3(d)$ we take into account that for infections with a fatal outcome, the individual course of the disease for elderly patients might differ compared to younger patients.
- $x_{t,d}$ are covariates which may be time and duration specific.

By utilizing a quasi-likelihood model (Fahrmeir, Kneib, Lang, & Marx, 2007) as in (5), we account for possible overdispersion in the data, which results in adjusted standard errors of the parameter estimates, while, however, the estimates themselves are the same when compared to the fit of a binomial model.

Assuming that D, the duration between a registered fatal infection and its reported death, is independent of the number of fatal COVID-19 infections, we obtain the relationship

$$
\mathbb{E}(C_{t,d,c}) = F_{t,c}(d) \cdot \mathbb{E}(Y_{t,c}).
\tag{7}
$$

Note further that if we model $Y_{t,c}$ with a negative binomial model as presented in the previous section, we have no final observation $Y_{t,c}$ for time points $t > T - d_{\max}$. Instead, we have observed $C_{t,T-t,c}$, which relates to the mean of $Y_{t,c}$ through (7) by $C_{t,T-t,c} = F_{t,c}(T - t) \cdot \mathbb{E}(Y_{t,c})$. Including therefore $\log F_{t,c}(T - t)$ as additional offset in model (2) allows to fit the model as before, but with the nowcasted number of fatal infections included. That means, instead of $\lambda_{t,r,c}$ as in (2), the expected number of fatal infections are now parameterized by $\lambda_{t,r,c}^{\star} = \lambda_{t,r,c} \exp(\log F_{t,c}(T - t))$, where the latter multiplicative term is included as additional offset in the model.

4.2 | Results for nowcasting

We fit the nowcasting model (5) with parameterization (6). We include a weekday effect for the registration date of the infection with reference category 'Monday'. The estimates of the fixed linear effects are shown in Table 2. The fitted smooth effects are shown in Figure 3. The top panel shows the effect over calendar time, which is very weak and confirms that the individual course of the disease hardly varies over time. This is supported by the fact that the German healthcare

TABLE 2 Estimated fixed linear effects (standard errors in brackets) in the nowcasting model (6). Parameters and their standard errors are given on the log scale. The relative risk is given together with 95% confidence intervals. The reference for the weekdays is Monday

	Effect (*SE*)	exp(Effect) Relative risk	95% Confidence interval of relative risk
Intercept	−3.12 (0.045)	0.04	[0.04, 0.05]
Tuesday	0.06 (0.060)	1.06	[0.94, 1.19]
Wednesday	0.11 (0.059)	1.12	[0.99, 1.25]
Thursday	0.20 (0.058)	1.23	[1.09, 1.38]
Friday	0.26 (0.059)	1.30	[1.16, 1.45]
Saturday	0.27 (0.063)	1.31	[1.16, 1.48]
Sunday	0.20 (0.068)	1.22	[1.07, 1.40]

system remained stable over the considered period, and hence survival did not depend on the date on which the infection was notified.

The bottom panel of Figure 3 shows the course of the disease as a smooth effect over the time between registration of the infection and death. We see that the probabilities $\pi(d; \cdot)$ decrease in d, where this effect is the strongest in the first days after registration. Thus, most of the COVID-19 patients with fatal infections are expected to die not long after their registration date. We also see no overall significant difference in the duration effect between the age groups 80− and 80+, since the fitted curves $s_2(d)$ and $s_2(d) + s_3(d)$ hardly differ. To some extent, this was already visible from Figure 1. This shows that, given that a registered case ends with a fatal outcome, the individual's course of the disease does not depend on the age group. The effect of d becomes easier to interpret by visualizing the resulting distribution function $F_{t,c}(d)$, where here g refers to the age group 80+. This is shown in Figure 4 for two different values of t, that is April 13 and May 13. The plot also shows how the course of the disease hardly varies over calendar time: In fact, the small differences between the two distribution functions is dominated by the weekday effect, since the red curve is related to a Monday while the blue one is from a Wednesday.

4.3 | Nowcasted number of fatal infections

On the day of analysis, we do not observe the total counts of deaths for recently registered infections. This means that there are an unknown number of currently infected people which will die at a future point in time. We therefore nowcast those numbers, that is we predict the prospective deaths which can be attributed to all registration dates up to today. This is done on a national level, and the resulting nowcast of fatal infections for Germany is shown in Figure 5. For example, on May 14, 2020 there are 25 deaths reported where the infection was registered on May 5 (red bullets on May 5). We expect this number to increase to about 50 when all deaths due to COVID-19 for this registration date will have been reported (green triangles on May 5). Naturally, the closer a date is to the present, the larger the uncertainty in the nowcast will be. This is shown by the shaded bands. Details on how the statistical uncertainty has been quantified are provided below. In Section 5, we incorporate the nowcasting results into the mortality model as discussed before, but the nowcast results are interesting in their own right. The curve confirms that the number of fatal infections is decreasing since the beginning of April. Note that the curve also mirrors the 'weekend effect' in registration, as less infections are reported on Sundays.

Since we are now more than $d_{\max} = 40$ days after the day of analysis (May 14, 2020), we can assess the predictive accuracy of our nowcast. Therefore, we also show in Figure 5 the counts of fatal infections, which we observe 40 days after the respective registration date. We see that our nowcast performs in general very well. However, there are a handful of registration dates for which the nowcasted values were clearly outside of the prediction intervals. Most remarkably, the cumulative number of fatal infections for registered infections on April 8, 2020 has dropped after May 14, 2020. This happens in the rare case in which the database has been modified retrospectively by the local health authorities.

4.4 | Uncertainty quantification in nowcasting

In Figure 5, we have shown the nowcasting results along with uncertainty intervals shaded in grey. These were constructed using a bootstrap approach as follows. Given the fitted model, we simulate $n = 10,000$ times from the asymptotic joint
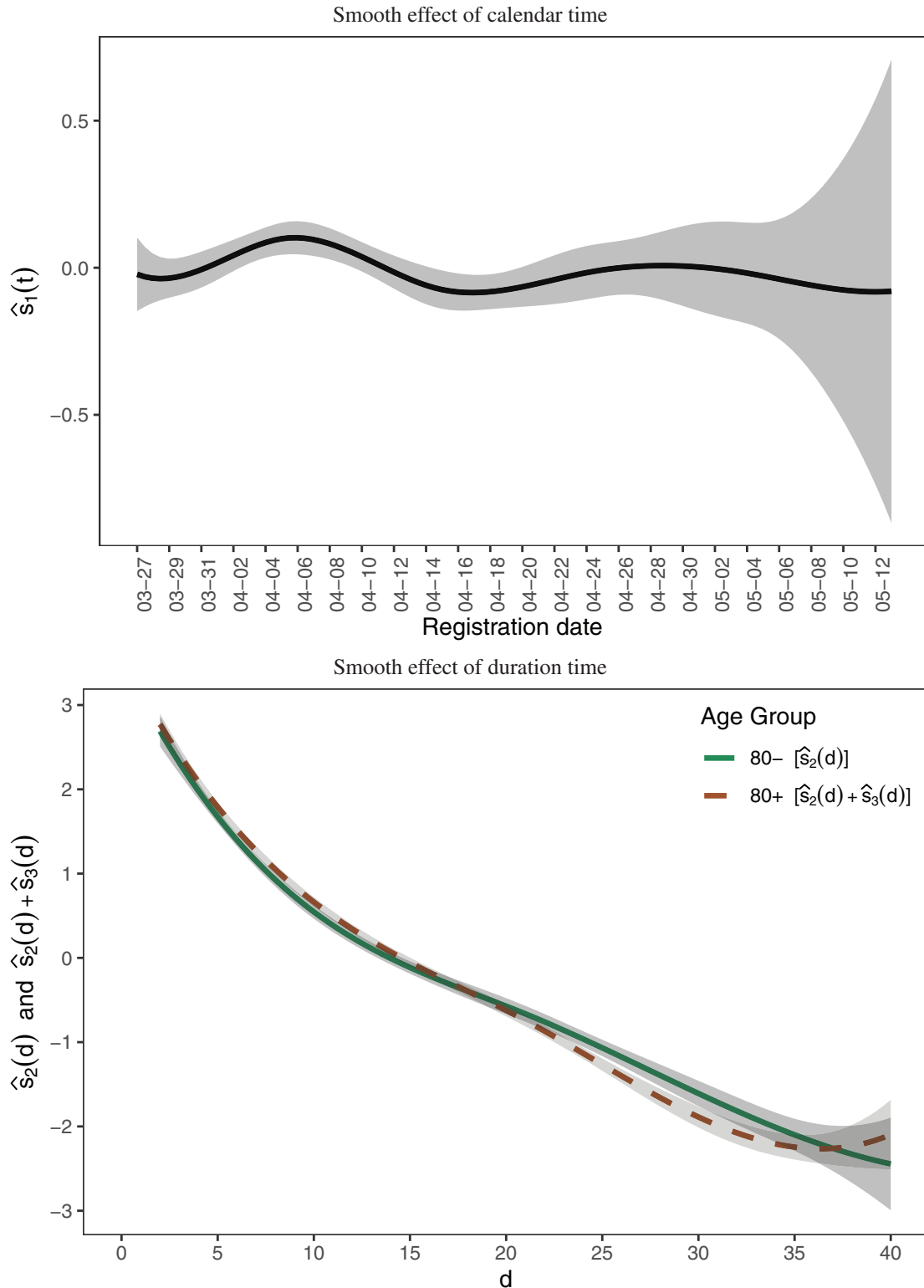
Smooth effect of calendar time

Smooth effect of duration time

FIGURE 3 Estimates of smooth effects in the nowcasting model

normal distribution of the estimated model parameters which results through (4). This leads to a set of bootstrapped distribution functions $\mathcal{F} = \{\widehat{F}_t^{(i)}(T-t), i = 1, \dots, n; t = T - d_{\max} + 1, \dots, T - 1\}$. This set is used to compute the simulated nowcasts $\widehat{Y}_t^{(i)} = C_{t,T-t}/\widehat{F}_t^{(i)}(T-t)$ applying (7), where $C_{t,T-t}$ is the observed partial cumulated sum of deaths at time point $T - t$ with registration date t. The point-wise lower and upper bounds of the 95% prediction intervals for the nowcast for Y_t are then given by the 2.5 and the 97.5 quantiles of the set $\{\widehat{Y}_t^{(i)}, i = 1, \dots, n\}$, respectively.
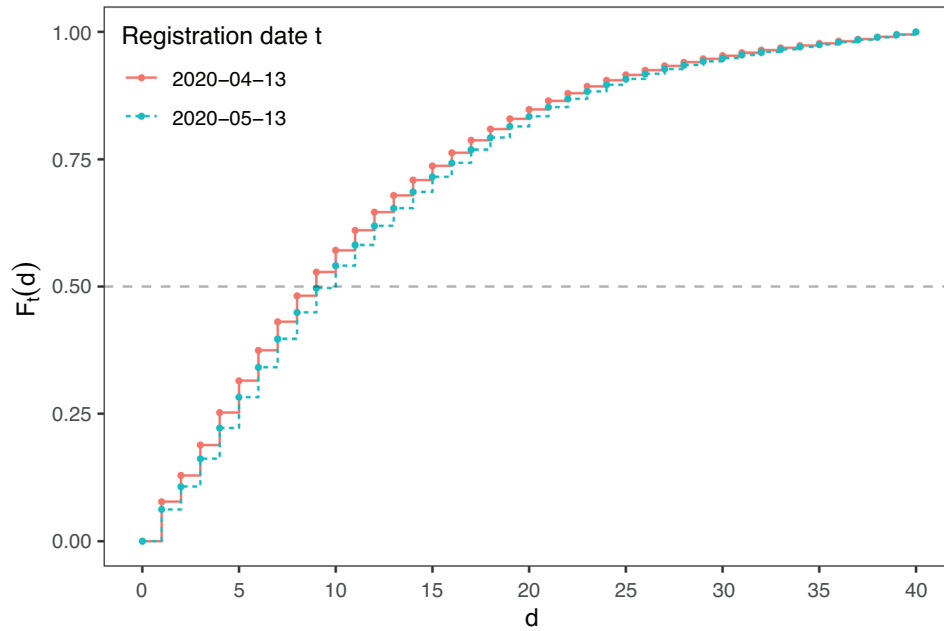
FIGURE 4 Fitted distribution function $F_t(d)$ for the age groups 80+ and 80−, where t corresponds to Wednesday, May 13, 2020
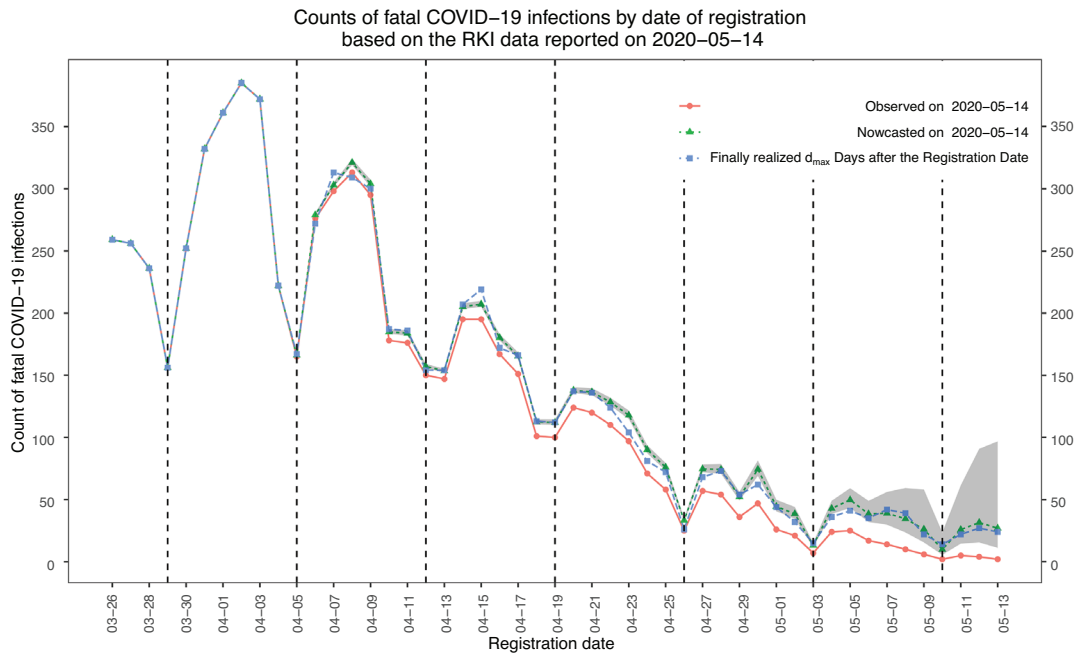

FIGURE 5 Observed (red line) and nowcasted (blue line) of daily death counts due to a COVID-19 infection on May 14, 2020 including 95% prediction intervals (shaded areas). Sundays are marked by a dashed vertical line. Finally realized death counts (d_{max} after the respective registration date) are shown as blue squares

5 | RESULTS OF THE MORTALITY MODEL

We first discuss the estimates of the fixed linear effects included in model (2), which are shown in Table 3. We see that both age and gender play a major role when estimating the numbers of fatal infections. Elderly people exhibit a much higher death rate from COVID-19, which is, for males (females) in the age group 80+, around 80 times ($148 \approx \exp(4.39 + 0.61)$ times) higher than in the reference age group 35–59. This already hints at a remarkable difference between genders, where the expected death rate for females in the reference age group is around 60% ($\approx 1 - \exp(-0.94)$) lower than the

TABLE 3 Estimated fixed linear effects (standard errors in brackets) in the mortality model (2). Parameters and their standard errors are given on the log scale. The relative risk is given together with 95% confidence intervals. The reference category for age is the age group 35–59. The reference for the weekdays is Monday

		Effect (S.E.)	exp(Effect) Relative risk	95% Confidence interval of relative risk
	Intercept	−15.90 (0.095)	$1.27 \cdot 10^{-7}$	$[1.05 \cdot 10^{-7}, 1.53 \cdot 10^{-7}]$
Patient related	Female	−0.94 (0.142)	0.39	[0.29, 0.51]
	Age 15–34	−2.53 (0.325)	0.08	[0.04, 0.15]
	Age 15–34 Female	−0.18 (0.674)	0.84	[0.22, 3.18]
	Age 60–79	2.61 (0.081)	13.58	[11.60, 15.90]
	Age 60–79 Female	0.07 (0.151)	1.07	[0.80, 1.45]
	Age 80+	4.41 (0.080)	81.9	[70.20, 95.90]
	Age 80+ Female	0.61 (0.147)	1.83	[1.38, 2.45]
Reporting related	Tuesday	0.20 (0.051)	1.22	[1.10, 1.35]
	Wednesday	0.23 (0.052)	1.26	[1.14, 1.39]
	Thursday	0.24 (0.050)	1.28	[1.16, 1.41]
	Friday	0.10 (0.051)	1.10	[1.00, 1.22]
	Saturday	−0.12 (0.054)	0.88	[0.79, 0.98]
	Sunday	−0.41 (0.058)	0.66	[0.59, 0.74]

corresponding death rate for males. When considering the total gender-related numbers of fatal infections in the age group 80+ (see Figure 1), the difference between the genders is seemingly very small. However, by respecting the district-, gender- and age-specific population sizes in our model we see that the death rate of females in the age group 80+ is still around 28% ($\approx 1 - \exp(-0.94 + 0.61)$) lower when compared to the male population in this age group. Furthermore, we see that significantly less deaths are attributed to infections registered on Sundays compared to weekdays, due to the existing reporting delay during weekends.

Our model includes a global smooth time trend representing changes in the death rate since March 26. This is visualized in Figure 6. The plotted death rate is scaled to give the expected number of deaths per 100,000 people in an average district for the reference group, that is males in the age group 35–59. Overall, we see a peak in the death rate on April 3 and a downwards slope until the end of April. However, our nowcast reveals that the rate remains constant since beginning of May. Note that such developments cannot be seen by simply displaying the raw death counts of these days. The nowcasting step inevitably carries statistical uncertainty, which is taken into account in Figure 6 by including best and worst case scenarios. The latter are based on bootstrapped confidence intervals, where details are provided in Section 6.3 later in the paper.

Our aim is to investigate spatial variation and regional dynamics. To do so, we combine a global geographic trend for Germany with unstructured region-specific effects, where the latter uncovers local behaviour. In Figure 7, we combine these different components and map the fitted nowcasted death counts related to COVID-19 for the different districts of Germany, cumulated over the last 14 days before the day of analysis, that is May 14, 2020. While in most districts of Germany, the death rate is relatively low, some hotspots can be identified. Among those, Traunstein and Rosenheim (in the south-east part of Bavaria) are the most evident, but Greiz and Sonneberg (east and south part of Thuringia) stand out as well, to mention a few. A deeper investigation of the spatial structure is provided in Section 6, where we show the global geographic trend and provide maps that allow to detect new hotspot areas, after correcting for the overall spatial distribution of the infection.

6 | MORE RESULTS AND MODEL EVALUATION

6.1 | Spatial effects

It is of general interest to disentangle the two spatial components that we introduced in Section 3. We visualize the fitted global geographic trend $m_2(\cdot)$ for Germany in Figure 8. The plot confirms that, up to mid May 2020, the northern parts
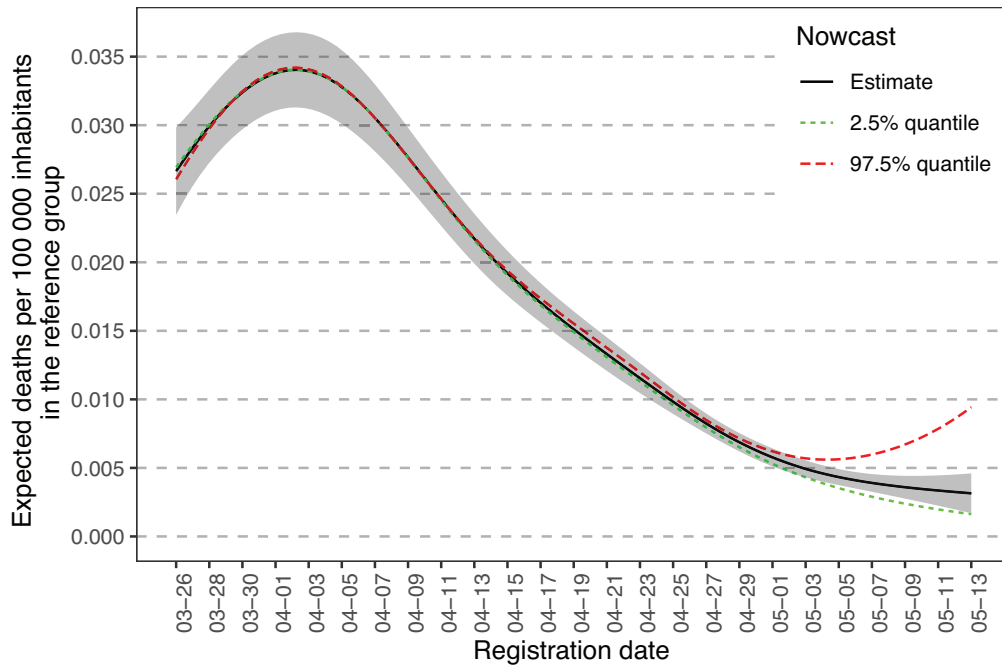
FIGURE 6 Fitted smoothed expected fatal COVID-19 infections per 100,000 inhabitants in the reference group (males aged between 35 and 59 in an average district) by registration date including 95% confidence bands as shaded area. Uncertainty resulting from the nowcast model is shown as dashed coloured lines

of the country are less affected by the disease in comparison to the southern states. The two plots in Figure 9 map the region-specific effects, that is the predicted long-term level of a district u_{r0} (left-hand side) and the predicted short-term dynamics u_{r1} (right-hand side). Both plots uncover quite some region-specific variability. In particular, the short-term dynamics u_{r1} (right plot) pinpoint districts with unexpectedly high nowcasted death rates in the last two weeks, after correcting for the global geographic trend and the long-term effect of the district. Some of the noticeable districts have already been highlighted in Section 3 above, but we can here detect further districts which are less evident in Figure 6: For instance, Steinfurt (in the north-west of North Rhine-Westphalia), Olpe (southern North Rhine-Westphalia) and Gotha (center of Thuringia) all show a relatively high rate of fatal infections.

6.2 | Age group-specific analyses

A large portion of the registered fatal infections related to COVID-19 stems from people in the age group 80+. Locally, high numbers are often caused by an outbreak in a retirement home. Such outbreaks apparently have a different effect on the spread of the disease, and the risk of an epidemic infection caused by outbreaks in this age group is limited. Thus, the death rate among elderly people could vary differently across districts when compared to regional peaks in the death rate of the rest of the population. In order to respect this, we decompose the district-specific effects $\boldsymbol{u}_r$ in (2) into $\boldsymbol{u}_r^{80-} = (u_{r0}^{80-}, u_{r1}^{80-})^\top$ for the age group 80− and $\boldsymbol{u}_r^{80+} = (u_{r0}^{80+}, u_{r1}^{80+})^\top$ for the age group 80+, where the age group 80− consists of the aggregated age groups 15–34, 35–59 and 60–79. We put the same prior assumption on the random effects as we did in (3), but now the variance matrix that needs to be estimated from the data has dimension 4×4.

The fitted age group-specific random effects are shown in Figure 10, where the $\boldsymbol{u}_r^{80-}$ are shown in the top panel and the $\boldsymbol{u}_r^{80+}$ in the bottom panel. Most evidently, the variation of the random effects is much higher in the age group 80+ when compared to the younger age groups, as more districts occur which are coloured dark blue or dark red, respectively. When comparing the district-specific short-term dynamics of the last 14 days (u_{r1}) in Figure 10 to those in Figure 9, we recognize that in most of the districts which recently experienced very high death intensities (with respect to the whole period of analysis), these stem from the age group 80+. As mentioned before, this can often be explained by outbreaks in retirement homes.

Nowcasted fatal infections per 100 000 inhabitants with registration dates from 2020-04-30 until 2020-05-13
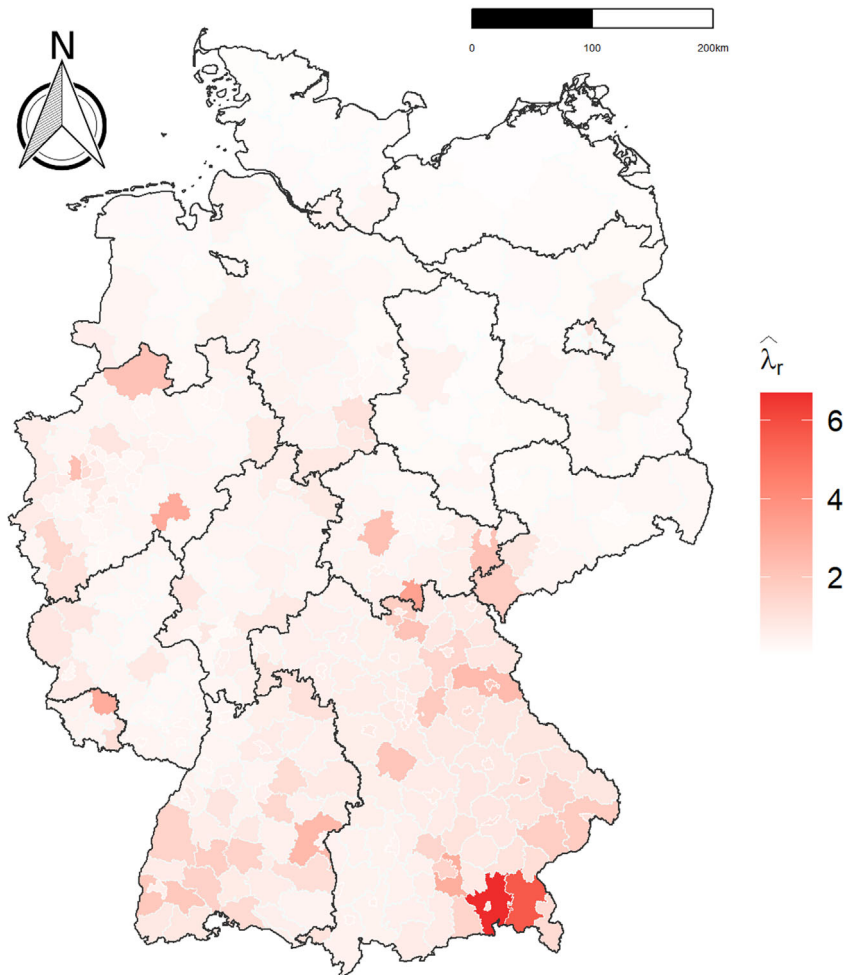
Based on data reported up to 2020-05-14.
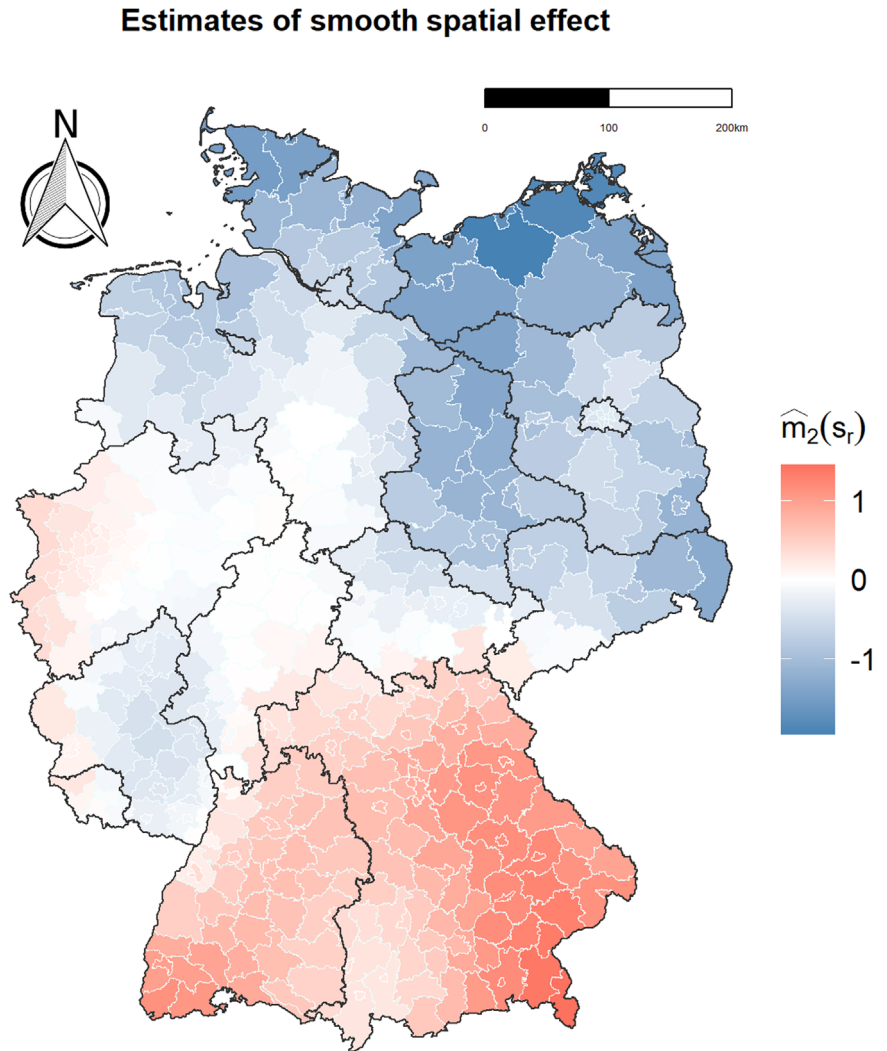Model includes registration dates from
2020-03-26 until 2020-05-13.

6.3 | Additional uncertainty in the mortality through the nowcast

When fitting the mortality model (1), we included the fitted nowcast model as offset parameter. This apparently neglects the estimation variability in the nowcasting model, which we explored via bootstrap as explained in Section 4.4 and visualized in Figure 5. In order to also incorporate this uncertainty in the fit of the mortality model, we refitted the model using (a) the upper end and (b) the lower end of the prediction intervals shown in Figure 5. It appears that there is little (and hardly any visible) effect on the spatial components, which is therefore not shown here. But the time trend shown in Figure 6 does change, which is visualized by including the two fitted functions corresponding to the 2.5% and 97.5% quantile of the offset function. We can see that the estimated uncertainty of the nowcast model mostly affects the last 10 days, with a strong potential increase in the death rate mirroring a possible worst case scenario.

6.4 | Auto-correlation of residuals in the mortality model

In the mortality model (2), we did not include an epidemic component accounting for possible temporal auto-correlation, as it is often done in endemic-epidemic models (see, e.g., Meyer et al., 2017). To check for possibly omitted auto-correlation

FIGURE 8 Smooth spatial effect of the death rate in Germany

Estimates of smooth spatial effect

$\widehat{m}_2(s_r)$

Based on data reported up to 2020-05-14.
Model includes registration dates from
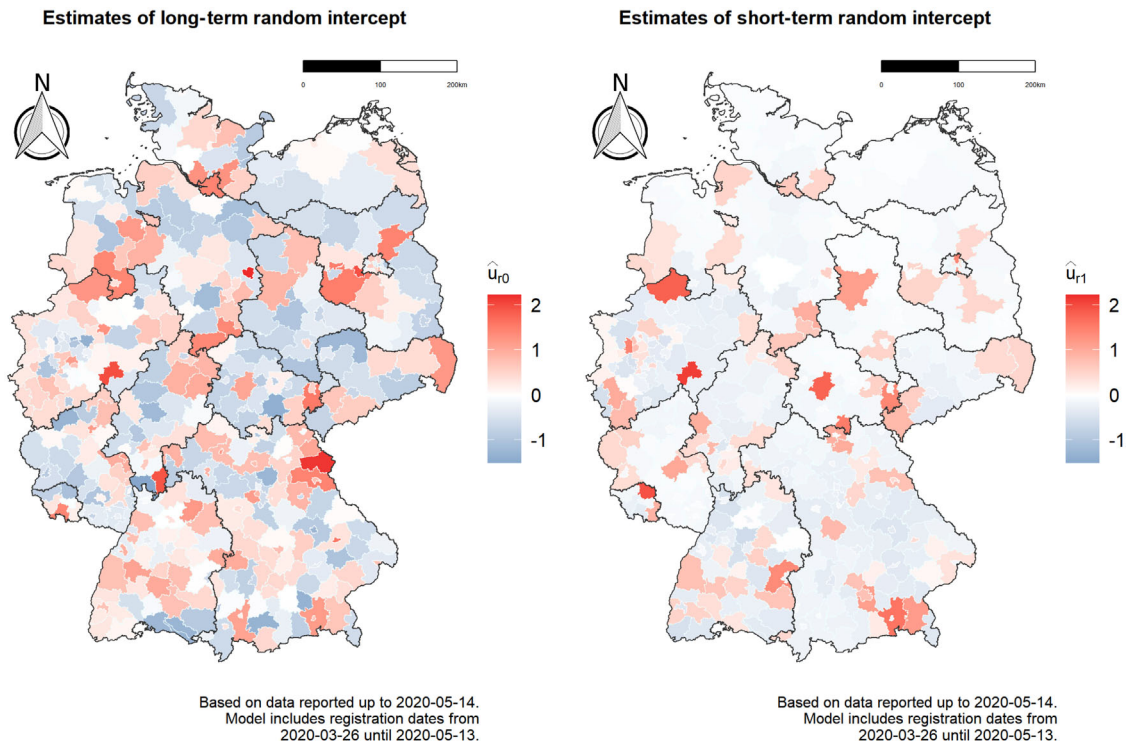2020-03-26 until 2020-05-13.

in our model, we explore the temporal correlation of the Pearson residuals in the mortality model (2). To do so, we compute the auto-correlation function (ACF) for all lags $k = 0, \ldots, T-1$. The corresponding ACF plot is shown in Figure 11. Apparently, the results do not show any pattern of auto-correlation and support the suitability of our model. We emphasize, however, that infection dynamics are included in the model through the time trend $m_1(t)$. Moreover, even if we ignore possibly existing auto-correlation, this time trend $m_1(t)$ is still estimated unbiased with penalized spline smoothing, which is robust against misspecification of the auto-correlation structure (Krivobokova & Kauermann, 2007).

We also think that the epidemic component is generally less impactful when modelling fatal infections in comparison to modelling the number of registered infections. The time between person-to-person transmission of the virus and a fatal outcome of a COVID-19 infection is much larger than the time until the registration of the infection, as shown in Figure 1, and hence any auto-correlation is rather indistinct for fatal cases.

7 | DISCUSSION

The paper presents a general approach for monitoring the dynamic behaviour of COVID-19 infections on a small-area level purely based on the analysis of the number of observed death counts. This in turn means that the results are less dependent on testing strategies, which may vary by region and over time.

Estimates of long-term random intercept

Estimates of short-term random intercept

Based on data reported up to 2020-05-14.
Model includes registration dates from
2020-03-26 until 2020-05-13.

Based on data reported up to 2020-05-14.
Model includes registration dates from
2020-03-26 until 2020-05-13.

FIGURE 9 Region-specific long-term level (left-hand side) and short-term dynamics of the 14 days prior to May 14, 2020 (right-hand side) of fatal COVID-19 infections

In addition, patients with fatal infections typically require intensive medical care and are therefore relevant in the planning of clinical capacities of the local health system. An analysis of fatal infections is especially interesting in situations in which reliable information on hospitalization is not available, as in the considered timespan of the COVID-19 pandemic in Germany.

The described nowcasting approach enables us to estimate the number of deaths following a registered infection even if the fatal outcome has not occurred yet, providing an up-to-date picture of the situation. The results of the nowcasting model confirm that the individual course of the disease for fatal infections did not change over calendar time nor did it differ by gender. More in particular, it uncovers that in Germany, during the considered timespan, elderly patients had, in the case of fatal infections, about the same course of the disease as younger patients.

Our analysis of the nowcasted number of fatal infections on a regional level allows to draw conclusions on the current dynamics of the disease on the spatial dimension. By separately estimating, for each district, a long-range effect which mirrors the overall situation as well as a short-term dynamic effect, we can timely identify districts with unexpectedly high nowcasted death rates. An additional interaction for elderly people allows us to distinguish between outbreaks which might be attributed to activity in retirement homes and those due to unexpected activities in the general population. Mapping the general pattern of the spread of the disease in Germany confirms that different regions are affected to different extents, with southern and western regions being generally more affected than northern states. In addition, a global smooth time trend captures the changes in death rate, showing the peak at the beginning of April and a constant decrease since then. Thanks to the implemented nowcasting, the time trend can be estimated up to the date of analysis. This spatially differentiated picture would not be achievable through a simple monitoring of district-specific observed deaths.

A natural next step would now be to consider the nowcasted fatal infections in relation to the number of newly registered infections, which is, in contrast, highly dependent on both testing strategy and capacity. We consider this as possible future research, but the proposed model allows to explore data in this direction. This might ultimately help us in shedding light on the relationship between registered and undetected infections as well as on the effectiveness of different testing strategies.

There are several limitations to this study, which we want to address as well. First and utmost, even though death counts are, with respect to cases counts, less dependent on testing strategies, they are not completely independent from them. This applies in particular to the handling of post-mortem tests. We therefore do not claim that our analysis of death counts is completely unaffected by testing strategies. Second, a fundamental assumption in the model is the independence between
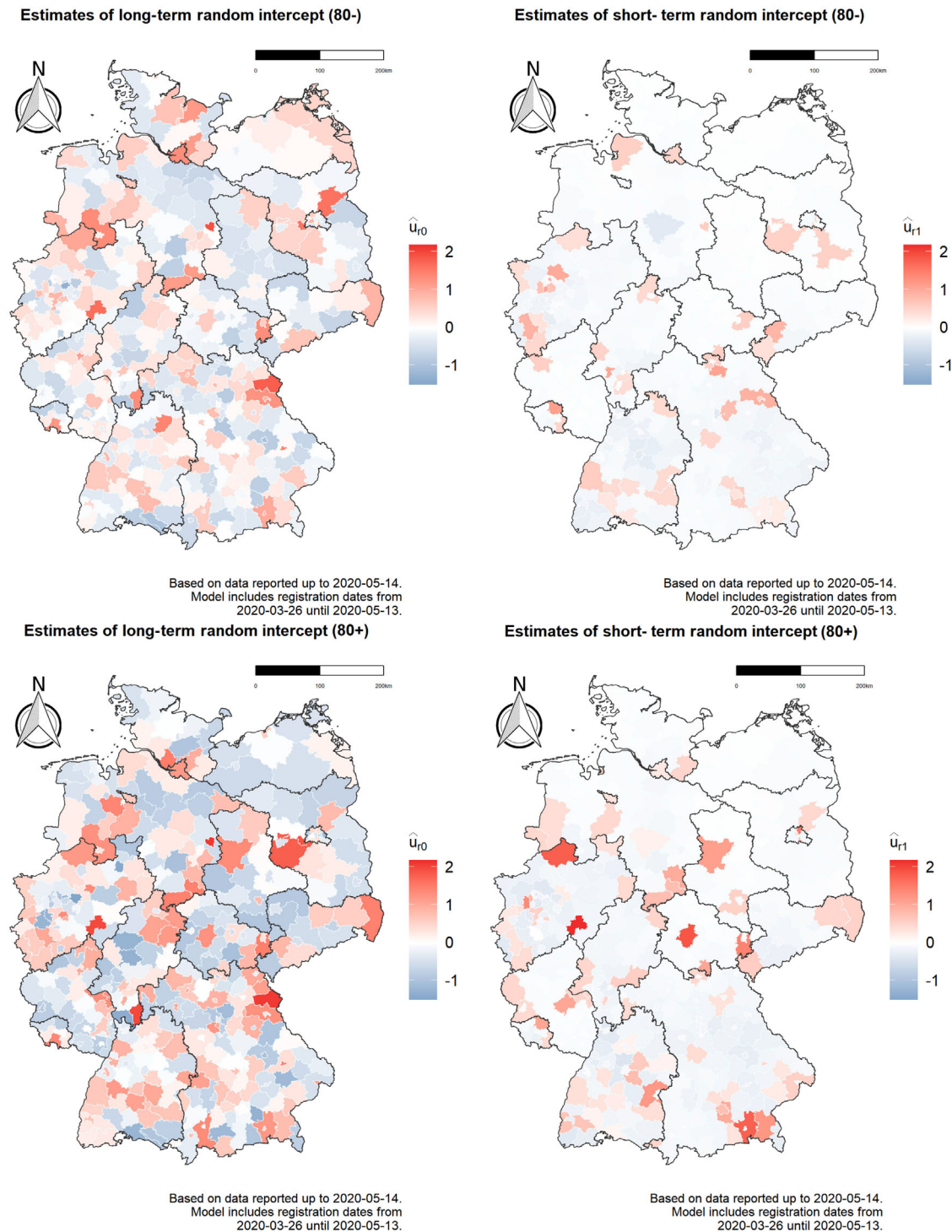
FIGURE 10 Region-specific long-term level (left-hand column) and short-term dynamics of the 14 days prior to May 14, 2020 (right-hand column) of fatal COVID-19 infections for the age groups under 80 (80−, upper row) and above 80 (80+, bottom row)

the course of the disease (on the population level) and the number of infections. Overall, if the local health systems have sufficient capacity and triage can be avoided, this assumption seems plausible, but it is difficult or even impossible to prove the assumption formally. However, the results of the nowcasting model empirically show a rather stable course of the disease, supporting our assumption. Furthermore, the registration of a COVID-19 case is related to the district of residence, while the infection does not necessarily occur in the district where the infected person resides. However, due to a lack of data we cannot explore this point further. Also, in the considered timespan, the mobility in the population has
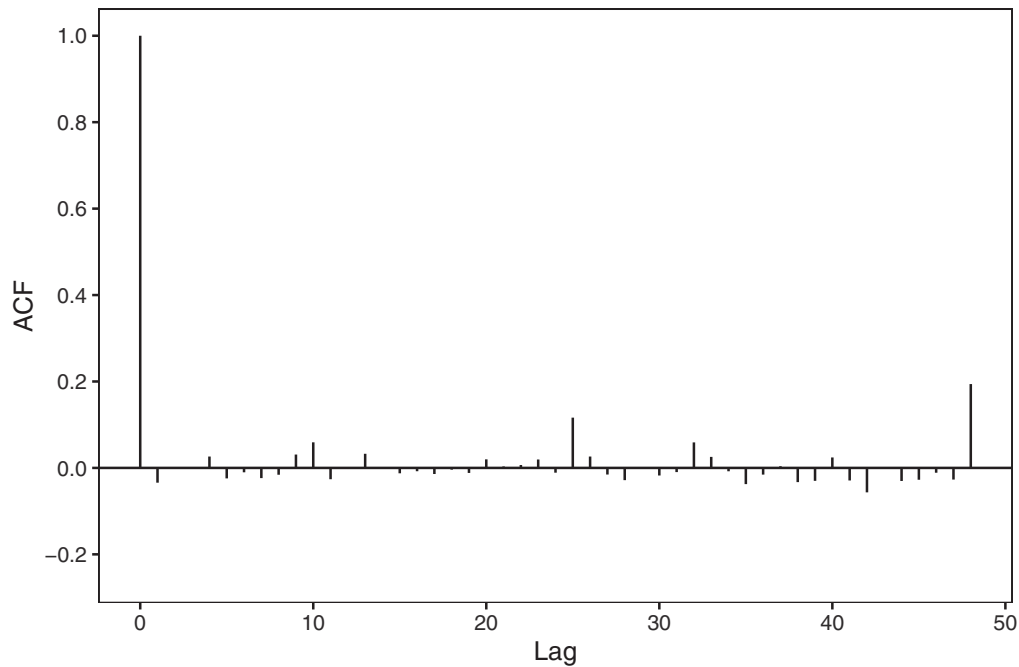
FIGURE 11   ACF plot of the Pearson residuals in the mortality model

been rather low due to governmental restrictions. Even though the model focuses on regional aspects of the pandemic, the nowcasting itself is carried out on a national level, due to sparse data. Given that our results show that the course of the disease from registration to death in Germany did not notably depend on age or gender, we do not expect it to depend on place of residence either.

A general limitation results through the availability of information. Our analyses are based on available data of all registered COVID-19 infections in Germany together with the information on fatalities, which is published daily by the RKI. While these data allow for an analysis of the occurrence of the disease in Germany, it lacks further detailed patient-specific information, for example on clinical aspects or on the differentiation between death with or because of COVID-19. This issue is shared with many other public disease registers. Note also that the methods we are proposing in this paper are not necessarily restricted to the use case of German COVID-19 data. For the purpose of applying our methodology to other countries, the data need to be in the same format as illustrated in Table 1, that is death counts need to be available in an aggregated form stratified by age (group), gender and district. For an appropriate interpretation of the results, it is critical that the reference date of every infection with a fatal outcome (here: registration date) corresponds to a time point at the early stages of the course of the disease. This could also be the date of infection with COVID-19, if known. The second date, which is needed for our nowcasting approach, is the reporting day of each fatal infection. While in Germany, this information can be deduced by considering the COVID-19 database daily over a longer period, the health authorities in other countries might supply historical reporting dates in a consecutively updated database.

Finally, the proposed approach demonstrates that valuable insight into the state and the dynamic of the disease can be obtained by disentangling spatial variation into a global pattern, district-specific long-term effects and current short-term dynamics in a spatio-temporal model. A particular virtue of the presented modelling approach over other proposals is that it also adjusts for the age and gender structure of the local population. This can provide relevant support for the monitoring of this new disease and can assist local health authorities in the planning of infection control measures as well as healthcare system capacities, in a further step towards the understanding and control of the COVID-19 pandemic.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

ORCID

Marc Schneble https://orcid.org/0000-0001-9523-4173

REFERENCES

Agresti, A. (2010). *Analysis of ordinal categorical data* (2nd ed.). Hoboken, NJ: Wiley.

Altmeijd, A., Rocklöv, J., & Wallin, J. (2020). Nowcasting COVID-19 statistics reported with delay: A case-study of Sweden. Preprint arXiv:2006.06840v1.

Baud, D., Qi, X., Nielsen-Saines, K., Musso, D., Pomar, L., & Favre, G. (2020). Real estimates of mortality following COVID-19 infection. *The Lancet. Infectious Diseases*, *20*(7), 773. https://doi.org/10.1016/S1473-3099(20)30195-X.

Bird, S., & Nielsen, B. (2020). Now-casting of COVID-19 deaths in English hospitals. Retrieved from http://users.ox.ac.uk/nuff0078/Covid/index.htm.

Cohen, J., & Kupferschmidt, K. (2020). Countries test tactics in "war" against COVID-19. *Science*, *367*(6484), 1287–1288.

Esri Deutschland GmbH (2020). Daily COVID-19 case numbers provided by the Robert-Koch-Institute. Retrieved from https://www.arcgis.com/home/item.html?id=f10774f1c63e40168479a1feb6c7ca74. Accessed: 30/09/2020.

Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2007). *Regression*. Berlin, Germany:Springer.

Ferguson, N., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., … Ghani, A. C. (2020). Report 9 - Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. London: Imperial College London. https://doi.org/10.25561/77482.

Flaxman, S., Mishra, S., Gandy, A., Unwin, H., Coupland, H., Mellan, T., … Bhatt, S. (2020). Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. Retrieved from https://spiral.imperial.ac.uk:8443/handle/10044/1/77731.

Grasselli, G., Pesenti, A., & Cecconi, M. (2020). Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: Early experience and forecast during an emergency response. *JAMA*, *323*(16), 1545–1546.

Grasselli, G., Zangrillo, A., & Zanella, A. (2020). Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2. *JAMA*, *323*(16), 1574–1581.

Günther, F., Bender, A., Katz, K., Küchenhoff, H., & Höhle, M. (2020). Nowcasting the COVID-19 pandemic in Bavaria. *Biometrical Journal*, 1–13. https://doi.org/10.1002/bimj.202000112.

Held, L., Meyer, S., & Bracher, J. (2017). Probabilistic forecasting in infectious disease epidemiology: the 13th Armitage lecture. *Statistics in Medicine*, *36*(22), 3443–3460.

Höhle, M., & an der Heiden, M. (2014). Bayesian nowcasting during the STEC O104:H4 outbreak in Germany, 2011. *Biometrics*, *70*, 993–1002.

Jung, S.-M., Akhmetzhanov, A., Hayashi, K., Linton, N., Yang, Y., Yuan, B., .. Nishiura, H. (2020). Real-time estimation of the risk of death from novel coronavirus (COVID-19) infection: Inference using exported cases. *Journal of Clinical Medicine*, *9*, 523.

Krivobokova, T., & Kauermann, G. (2007). A note on penalized spline smoothing with correlated errors. *Journal of the American Statistical Association*, *102*(480), 1328–1337.

Lawless, J. (1994). Adjustment for reporting delays and the prediction of occurred but not reported events. *Canadian Journal of Statistics*, *22*(1), 15–31.

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., & Shaman, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, *368*(6490), 489–493.

Linton, N. M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A. R., Jung, S.-m., … Nishiura, H. (2020). Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. *Journal of Clinical Medicine*, *9*(2), 538.

Massonnaud, C., Roux, J., & Crépey, P. (2020). COVID-19: Forecasting short term hospital needs in France. *medRxiv 2020.03.16.20036939*. https://doi.org/10.1101/2020.03.16.20036939.

Meyer, S., Held, L., & Höhle, M. (2017). Spatio-temporal analysis of epidemic phenomena using the R package surveillance. *Journal of Statistical Software, Articles*, *77*(11), 1–55.

Niehus, R., De Salazar, P. M., Taylor, A., & Lipsitch, M. (2020). Quantifying bias of COVID-19 prevalence and severity estimates in Wuhan, China that depend on reported cases in international travelers. *medRxiv* 2020.02.13.20022707. https://doi.org/10.1101/2020.02.13.20022707.

Robert Koch-Institut (2020). COVID-19-dashboard. Retrieved from https://experience.arcgis.com/experience/478220a4c454480e823b17327b2bf1d4.

StaBLab, LMU Munich (2020). CoronaMaps. Retrieved from https://corona.stat.uni-muenchen.de/maps/.

Streeck, H., Schulte, B., Kümmerer, B. M., Richter, E., Höller, T., Fuhrmann, C., . . . Hartmann, G. (2020). Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. *medRxiv* 2020.05.04.20090076. https://doi.org/10.1101/2020.05.04.20090076.

Wand, M. (2003). Smoothing and mixed models. *Computational Statistics*, *18*(2), 223–249.

Wilson, N., Kvalsvig, A., Barnard, L. T., & Baker, M. G. (2020). Case-fatality risk estimates for COVID-19 calculated by using a lag time for fatality. *Emerging Infectious Diseases*, *20*(6), 1339–1441.

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. Boca Raton, FL: CRC Press.

Zeger, S. L., See, L. C., & Diggle, P. J. (1989). Statistical methods for monitoring the AIDS epidemic. *Statistics in Medicine*, *8*, 3–21.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.