Monitoring the spatiotemporal epidemiology of Covid-19 incidence and mortality: a small-area analysis in Germany

Sven Rohleder
Universitat Bielefeld Fakultat fur Gesundheitswissenschaften
Kayvan Bozorgmehr (kayvan.bozorgmehr@uni-bielefeld.de)
https://orcid.org/0000-0002-1411-1209

Research article

Keywords: Covid-19, infectious disease monitoring, Bayesian spatial analysis, infectious disease modelling

Posted Date: June 11th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-34572/v1

Version of Record: A version of this preprint was published at Spatial and Spatio-temporal Epidemiology on May 1st, 2021. See the published version at https://doi.org/10.1016/j.sste.2021.100433.

Abstract

Background: As response to the pandemic of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), countries worldwide have implemented mitigation and control measures at national and subnational level. Timely monitoring of risks of SARS-CoV-2 incidence and associated deaths at small-area level is essential to inform local response strategies. However, the potentials of spatial epidemiology to contribute to this aim are yet untapped in most countries. Using the example of Germany, we analysed the spatiotemporal epidemiology of SARS-CoV-2 incidence and associated deaths at district level to develop a tool for monitoring incidence and mortality rates and to estimate district-specific risks of disease incidence. Methods: We conducted a longitudinal small-area analysis for 401 districts to assess the district-specific risks of SARS-CoV-2 incidence by using nationally representative data from the national surveillance system in Germany on a daily basis (January 28 th to May 4 th 2020). We used a Bayesian spatiotemporal model to estimate the district-specific risk ratios (RR) of SARS-CoV-2 incidence and the posterior exceedance probability for RR thresholds greater than 1, 2 or 3, respectively. We further calculated standardised incidence (SIR) and mortality ratios (SMR) stratified by sex and age groups to assess the spatial distribution of SARS-CoV-2 incidence and deaths. Results: A total of 85 districts (21 % of all districts) showed a RR greater than 3, and 63 districts (16 % of all districts) exceed the RR threshold with a probability of greater than 80 %. Median RR was 1.19 (range 0-523.08), and the median SIR and SMR were 0.34 (range 0-423.94) and 0 (range 0-343.39), respectively. Elevated RR, and correspondingly high SIR and SMR, were observed in at-risk districts (identified by the spatiotemporal model) in southern and western districts of Germany. Daily updates of district-specific risk, SIR and SMR are implemented in a web-based platform. Conclusions: Our approach provides an informative and timely tool to monitor the district-specific risks of SARS-CoV-2 incidence and associated deaths. This approach can be used to inform local authorities for decision-making and strategy planning on containing the SARS-CoV-2 pandemic.

Background

The pandemic of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused yet more than 3,500,000 notified cases and about 240,000 deaths worldwide, which are still progressively increasing [1]. Since the first notified case in Germany (January 28th, 2020), a total of 167,300 cases and 7,266 associated deaths have been registered until 7th May 2020 [2].

Measures proposed by the WHO to contain the outbreak and protect the population are adopted on governmental and societal levels within the countries. These measures include a wide range of non-pharmaceutical interventions (NPI) on national and subnational level considering restrictions for public events, institutions, spaces and transport as well as traveling, strengthening healthcare systems, protecting vulnerable populations, and testing policies [3]. Ongoing monitoring and epidemiological risk assessments of SARS-CoV-2 incidence and mortality on a local level are essential to identify high-risk areas and to inform policy decision-making and planning for implementation, withdrawal, or re-implementation of control and mitigation strategies [3]. The use of spatial epidemiological methods as a

tool to assess the pandemic course and spatial distribution of SARS-CoV-2 is an essential approach for evaluating the outbreak in terms of different regional characteristics and for informing local decision-makers on containing the epidemiological spread of SARS-CoV-2. However, only few studies focus on analysis and monitoring SARS-CoV-2 incidence or associated deaths on a small area level using real-time data [4–7]. The majority of geographical information system dashboards or tools predominantly reflect crude cumulative counts and rates of cases and deaths [8], although risk measures on a regional level are needed, which are necessary for planning locally required tailored strategies.

After an initial phase of concerted policy measures and NPI aimed at flattening the epidemiological curve in Germany, questions of regional differentiation and risk-adapted mitigation strategies have emerged and are subject to ongoing intensive policy debates [9]. Such an approach, however, requires timely, continuous and reliable monitoring of regional risks of SARS-CoV-2 incidence at small-area level, and measures that go beyond crude cases and cumulative incidence rates per population. Using the example of Germany, the aim of our study was to (i) analyse the spatiotemporal epidemiology of SARS-CoV-2 incidence and associated deaths at district level from the beginning of the pandemic in Germany on January 28th 2020 until May 4th 2020, and to (ii) implement the approach into a tool for ongoing monitoring of standardised incidence and mortality rates as well as district-specific risks.

Methods

Study design

We conducted a longitudinal small-area analysis in Germany at level of 401 districts (Nomenclature of Territorial Units for Statistics, NUTS 3) to assess the distribution of SARS-CoV-2 incidence and associated deaths using standardised incidence (SIR) and mortality ratios (SMR). We further estimate the district-specific risks and probabilities of exceeding risk thresholds for disease incidences using a Bayesian spatiotemporal model. The analysis based on daily-notified cases of SARS-CoV-2 from 28th January 2020 to 4th May 2020.

Data sources

We linked three different nationally representative data sources, including district data on SARS-CoV-2 incidence and deaths, population statistics, and cartographical data. Daily data on SARS-CoV-2 incidence and associated deaths (28/01/2020 to 04/05/2020) was obtained from the Robert-Koch Institute, which provides the national surveillance system of infectious diseases in Germany, stratified by sex and age groups (0–4, 5–14, 15–34, 35–59, 60–79, 80+) [10]. We used most recent district-level population statistics (2019) from a database of the system of social reporting in official statistics to calculate expected incidences and deaths [11]. Cartographical data were taken from the Federal Agency of Cartography and Geodesy [12].

Statistical analysis

We first assessed the distribution of SARS-CoV-2 incidence and associated deaths by calculating weekly standardised incidence (SIR) and mortality ratios (SMR) at district level stratified for sex and age groups (0–4, 5–14, 15–34, 35–59, 60–79, 80+) and on federal states level. For the calculation of SIR and SMR, and corresponding Poisson 95% confidence intervals (95%-CI), the ratio of observed versus expected counts was calculated [13, 14].

We then used a Bayesian spatiotemporal model, fitted by the integrated nested Laplace approximation (INLA) approach [15], to assess the spatial risk of SARS-CoV-2 incidence. For this analysis, we used daily notifications from 28/01/2020 to 04/05/2020. In order to select the likelihood distribution for the model that fit the count data best (e. g. Poisson, zero-inflated Poisson, negative binomial or zero-inflated negative binomial), four intercept-only models were specified at first. Using the Watanabe-Akaike information criterion (WAIC), the likelihood of the intercept-only model with lowest WAIC was selected for fitting two spatiotemporal models, which included a parametric and a non-parametric time trend, respectively. Based on lowest WAIC, we then selected the model that fit the data best for further calculation (see Additional file 1: Table S1 and Table S2) [14].

A zero-inflated negative binomial spatiotemporal model fit the data best. The model included a Besag-York-Mollié (BYM) model and a non-parametric dynamic time trend. For districts $i = \{1, \ldots, 401\}$ and notification days $t = \{1, \ldots, 84\}$ the rate η_{it} of observed SARS-CoV-2 incidence and expected number of cases integrated as offset $\log(E_{it})$ was modelled on a logarithmic scale:

$$\eta_{it} = \alpha + u_i + v_i + \gamma_t + \varphi_t$$

.

Where α is the intercept, $u_i + v_i$ is the BYM-model with u_i as spatially structured random effect modelled using intrinsic conditional auto-regression (neighbouring structure of districts) and v_i as spatially unstructured effect modelled exchangeable among the districts (independent and identically distributed, iid). The dynamic temporal trend incorporates a temporal structured random effect γ_t modelled dynamically using a random-walk model of order two, and an unstructured temporal effect φ_t modelled using the iid [14, 16].

The district-specific risk ratios (RR) and corresponding 95% credibility intervals (95%-CrI) are calculated by combining the spatial structured and unstructured random effects, e. g. $RR_i = e^{u_i + v_i}$ [16]. For assessing the uncertainty associated with the RR_i, we compute the posterior probability of exceeding an RR_i threshold greater than 1, 2 and 3, respectively. Following the Richardson criterion, an exceedance probability of greater than or equal to 80% determines a high certainty of exceeding one of the mentioned RR_i thresholds in the respective district [17]. We further assess the posterior temporal mean trend at national level to provide an estimate for the time-dependent change in the risks of SARS-CoV-2 incidences in Germany. Therefore, we combined the temporal effects trough a linear combination and calculated time-specific risk ratios (TRR_t) and 95%-CrI [14].

The analysis was conducted using the R-programming language for statistical computing (V. 3.6.3). The R-INLA package [18] was used to fit the Bayesian models, the tmap [19] and the leaflet [20] packages were used for generating the maps.

Calculation of expected values

The expected SARS-CoV-2 incidence for district $i = \{1, \ldots, 401\}$ and notification day $t = \{1, \ldots, 84\}$ used for the Bayesian spatiotemporal models were calculated on the basis of the cumulative incidence rate of SARS-CoV-2 (with observed incidence O_{it}) until the current notification day weighted by the respective district population size p_{it} (formula 1) [14, 16]:

$$E_{it} = p_{it} \times \left(\frac{\sum_{i=1,t=1}^{i=401,t=n} O_{it}}{\sum_{i=1,t=1}^{i=401,t=n} p_{it}} \right)$$

$$(formula1)$$

.

The calculation of the weekly expected incidence and deaths used for estimating SIR and SMR is based on formula 1. The cumulative incidence rate for the most recent notification week t_n with $n = \{1, \ldots, 15\}$ was calculated on the basis of the respective preceding notification weeks $t_{1,\ldots,n}$ (formula 2):

$$E_{it_n} = p_{it_n} \times \left(\frac{\sum_{i=1,t=1}^{i=401,t=n} O_{it_{1,\ldots,n}}}{\sum_{i=1,t=1}^{i=401,t=n} p_{it_{1,\ldots,n}}} \right)$$

$$(formula2)$$

.

The same calculation procedure was applied for estimating stratified SIR and SMR for age groups (0–4, 5–14, 15–34, 35–59, 60–79, 80+) and sex, and for the 16 federal states in Germany.

Approach for ongoing monitoring

In order to ensure an ongoing daily monitoring, we implement the statistical approach explained above in an automatic process and generate interactive maps for SIR, SMR and district specific risk estimates for SARS-CoV-2 incidence and mortality on a daily basis. The data management and data analysis is carried out with the R programming language, and is automatically started when the database of the Robert-Koch Institute is updated for the previous notification day at 00:00. The output of the analysis contains maps grouped by calendar weeks for daily and weekly SIR and SMR estimates (including age group and sex stratified estimates and counts for SARS-CoV-2 morbidity and mortality), as well as a daily updated map for district specific risk ratios based on the model configuration and selection process explained

above. The daily updated interactive maps are then automatically integrated in an interactive web surface, the “Covid-19 Small Area Monitor” (CoV-19 SAM) [21], which is programmed using JavaScript and the JavaScript library jQuery.

Results

We analysed a total of 163,719 incident SARS-CoV-2 cases (52% female) and 6822 associated deaths (44% female) between 28/01/2020 and 04/05/2020. See additional file for descriptive statistics (see Additional file 2: Table S3 and Table S4).

Standardised incidence and mortality ratios

Legend: A SIR greater than 1 means that the observed number of SARS-CoV-2 cases (O) is higher than expected (E).

Legend: A SMR greater than 1 means that the observed number of SARS-CoV-2 cases (O) is higher than expected (E).

On federal states level (Fig. 3), the epidemiological curve of SARS-CoV-2 cases and deaths declined over time in every state with highest peaks between the weeks 13 and 16. The states Bavaria, Baden-Württemberg, Northrhine-Westphalia, Hamburg and also Saarland show highest SIR and SMR trends over time. On week 14, the highest SIR was observed in Bavaria with 5.0 (4.9–5.1) and highest SMR in Saarland with 8.3 (6.2–10.7). SIR and SMR estimates on federal states level are provided in the additional file (see Additional file 4: Table S5 and Table S6).

Legend: A SIR or SMR greater than 1 means that the observed number of SARS-CoV-2 cases (O) is higher than expected (E). Estimates illustrated on a log-scale.

District-specific risks and exceedance probability

The district-specific RR and the posterior probability of exceeding an RR threshold at level of 401 districts are illustrated in Fig. 4. RR and corresponding 95%-CrI grouped by RR thresholds are shown in Fig. 5. Median RR over all districts was 1.19 (95%-CrI: 0.68–1.91), ranging from 0 to 523.08. Elevated RR for SARS-CoV-2 incidence are predominantly clustered in southern and western districts of Germany with highest RR in the districts Heinsberg (RR = 523.08, 95%-CrI: 311.54-824.49) and Straubing (25.93, 14.98–41.84). Also, Hamburg (2.44, 1.56–3.64) and Berlin (1.13, 0.73–1.68) and its periphery districts in Eastern Germany are at higher risk for SARS-CoV-2 incidence. In total, 224 of 401 districts have a RR greater than 1, and the threshold was exceeded by 186 regions with a probability of 80%. A total of 133 and 85 districts have a RR greater than 2 and 3, respectively, of which 97 and 63 districts exceeded the threshold with a probability of 80%, respectively. District-specific risks estimates and exceedance probabilities for given threshold are provided in the additional file (see Additional file 5: Table S6).

Legend: A RR above 1 means that the posterior mean of district-specific risks is higher than the average incidence risk across Germany. An exceedance probability above 80% determines a high certainty of exceeding one of the RR thresholds.

Legend: A RR above 1 means that the posterior mean of district-specific risks is higher than the average incidence risk across Germany. An exceedance probability above 80% determines a high certainty of exceeding one of the RR thresholds.

Legend: Temporal risk ratio (TRR) and corresponding 95% credibility intervals (95%-CrI) of SARS-CoV-2 incidence by notification day on a log-scale.

Discussion

Our analysis showed that the spatiotemporal risk for SARS-CoV-2 incidence was highest in 85 districts, predominantly across southern and western parts of Germany, of which 63 exceeded the risks threshold with high certainty. The temporal trend for the risks of SARS-CoV-2 incidence increased across Germany until the end of March. Since then, the time trend decreased to a lower, but still elevated, level. Also the SIR and SMR estimates showed visible dynamics of observed to expected incidence and associated deaths, especially in high-risk areas identified through the modelling approach.

As situations may change dynamically, the ongoing and continuous re-calculation and provision of small-area risk estimates by means of the presented methods will form a relevant and solid basis for local and regional decision-making. Beside the district-specific risk estimates, the calculation of SIR and SMR based on population-weighted observed vs. expected numbers of confirmed SARS-CoV-2 cases and associated deaths detects "at-risk" districts that deviate from the mean national incidence or death rate without introducing a fix threshold.

The analysis generates a robust basis for local decision-making and makes best use of best-available data by considering daily updated risks relative to a national reference, and by considering temporal dynamics and spatial dependencies. Implications of the estimates could be that mitigation measures are partially ceased, or withdrawn in districts with estimated lower risk of infections and low probability of exceeding an RR threshold with high certainty.

As the presented analysis is re-run on a daily basis in our web-platform “Covid-19 Small Area Monitor” (CoV19-SAM) [21], a re-introduction or intensification of mitigation strategies is also possible if districts change their risk profile from low to higher risks with high probabilities of exceeding RR thresholds. In comparison to other suggested measures (such as a fix incidence threshold of 50 per 100,000) [10], the focus on relative risks may be both more sensitive and flexible by considering local and age-sex-specific disease incidence relative to the overall course of the epidemic in the country.

While we provide a tool that deliver timely information on district-specific risk to guide political decisions, limitations of the underlying notification data cannot be resolved by our analysis. For example, as notified

cases are also a function of testing strategies, local variations and adaptations of the national testing guideline may affect incidence variations. Local planners and public health officers should be aware of such potential sources of variation and consider these carefully when interpreting the estimates. As the analysed outcomes can be considered as rare, daily analysis of SIR and SMR leads to strong variations. These are, however, levelled by analysis of weekly periods. Overall RR in a few districts, which were particularly affected by the pandemic, show extremely high values (RR > 100) at the beginning of the pandemic as the national references are comparably low. As the median RR across all districts in Germany was 1.19 (range 0-523.08), these extreme values (outliers) are exceptional and need to be interpreted very carefully. The approach allows in the course of the pandemic to considering health system resources, mobility measures or time-projections in estimation of risk, as well as socio-economic and other potential determinants of SARS-CoV-2 incidence at small-area level. The approach may also serve as blueprint for monitoring approaches elsewhere.

Conclusions

A spatial epidemiology approach for monitoring the risk of SARS-CoV-2 incidence and associated deaths provides an informative and timely fundament for local policy planning and decision-making. Areas at high risk can be sensitively detected considering both the spatial and temporal dynamics of disease incidence, complementing the monitoring of crude cases and population-weighted cumulative incidence rates. Based on the presented approach, the small-area risk of disease incidence can be routinely monitored and strategies for containing the epidemiological outbreak of SARS-CoV-2 can be locally adopted.

Abbreviations

95%-CI: Poisson 95% confidence intervals

95%-CrI: 95% credibility intervals

BYM: Besag-York-Mollié

CoV-19 SAM: Covid-19 Small Area Monitor

RR: district-specific risk ratio

iid: independent and identically distributed

NPI: non-pharmaceutical intervention

NUTS 3: Nomenclature of Territorial Units for Statistics (level 3)

SARS-CoV-2: severe acute respiratory syndrome coronavirus 2

SIR: standardised incidence ratio

SMR: standardise mortality ratio

TRR: time-specific risk ratio

WAIC: Watanabe-Akaike information criterion

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

No specific funding was raised for this study.

Authors' contributions

KB conceived the study. SR and KB designed the analysis methodology. SR collected the data, conducted the data analysis and modelling, generated figures and maps, and wrote the first and final version of the manuscript. KB revised and edited the manuscript for important intellectual content. All authors approved the final version for publication.

Acknowledgements

We acknowledge the contributions of Markus Qreini (Heidelberg University hospital) for technical support in data management and joint development of the monitoring tool.

References

1. World Health Organization (WHO). Coronavirus disease (COVID-19) Pandemic: WHO. 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019. Accessed 6 May 2020.
2. Robert-Koch Institute (RKI). Coronavirus Disease 2019 (COVID-19) Daily Situation Report of the Robert Koch Institute. 08/05/2020 - Updated status for Germany: RKI. 2020. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-05-08-en.pdf?__blob=publicationFile. Accessed 8 May 2020.
3. World Health Organization (WHO). Covid-19 Strategy Update Geneva: WHO. 2020. https://www.who.int/docs/default-source/coronaviruse/covid-strategy-update-14april2020.pdf?sfvrsn=29da3ba0_19. Accessed 18 May 2020.
4. Guliyev H. Determining the spatial effects of COVID-19 using the spatial panel data model. Spat Stat. 2020:100443.
5. Huang R, Liu M, Ding Y. Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis. J Infect Dev Ctries. 2020;14(3):246-53.
6. Lakhani A. Introducing the Percent, Number, Availability, and Capacity [PNAC] Spatial Approach to Identify Priority Rural Areas Requiring Targeted Health Support in Light of COVID-19: A Commentary and Application. J Rural Health. 2020.
7. Mollalo A, Vahedi B, Rivera KM. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Sci Total Environ. 2020;728:138884.
8. Kamel Boulos MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. Int J Health Geogr. 2020;19(1):8.
9. Abele-Brehm A, Dreier H, Fuest C, Grimm V, Kräusslich H, Krause G, et al. Die Bekämpfung der Coronavirus-Pandemie tragfähig gestalten. Empfehlungen für eine flexible, risikoadaptierte Strategie. 2020.
10. Robert-Koch Institute (RKI). COVID-19 Dashboard: RKI. 2020. https://corona.rki.de. Accessed 5 May 2020.
11. Statistische Ämter des Bundes und der Länder. Regionaldatenbank Deutschland Düsseldorf: Information und Technik Nordrhein-Westfalen (IT.NRW). 2019. https://www.regionalstatistik.de/genesis/online/logon. Accessed 1 May 2020.
12. Bundesinstitut für Kartographie und Geodäsie. Verwaltungsgebiete 1:1.000 000 Frankfurt am Main: BKG. 2018. http://www.geodatenzentrum.de/geodaten/gdz_rahmen.gdz_div?gdz_spr=deu&gdz_akt_zeile=5&gdz_anz_zeile=1&gdz_unt_zeile=16&gdz_user_id=0. Accessed 30 Apr 2018.
13. Bivand R, Pebesma E, Gómez-Rubio V. Applied Spatial Data Analysis with R. New York: Springer Science+Business Media; 2008.
14. Blangiardo M, Cameletti M. Spatial and Spatio-temporal bayesian Models with R-INLA. New Jersy: John Wiley & Sons, Ltd; 2015.

15. Martins T, Simpson D, Lindgren F, Rue H. Bayesian computing with INLA: New features. Computational statistics & Data Analysis. 2013;67:68-83.
16. Knorr-Held L. Bayesian modelling of inseparable space-time variation in disease risk. Statist Med. 2000;19:2555-67.
17. Richardson S, Thomson A, Best N, Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies. Environ Health Perspect 2004;112(9):1016-25.
18. Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximation. JR Statist Soc B. 2009;71:319-92.
19. Tennekes M. tmap: Thematic Maps in R. Journal of Statistical Software, . 2018;84(6):1–39.
20. Graul C. leafletR: Interactive Web-Maps Based on the Leaflet JavaScript Library. R package version 0.4-0. 2016.
21. Rohleder S, Qreini M, Bozorgmehr K. Covid-19 Small Area Monitor (Cov-19 SAM): University of Bielefeld. 2020. https://covidmonitor.de/. Accessed 6 May 2020.

Figures

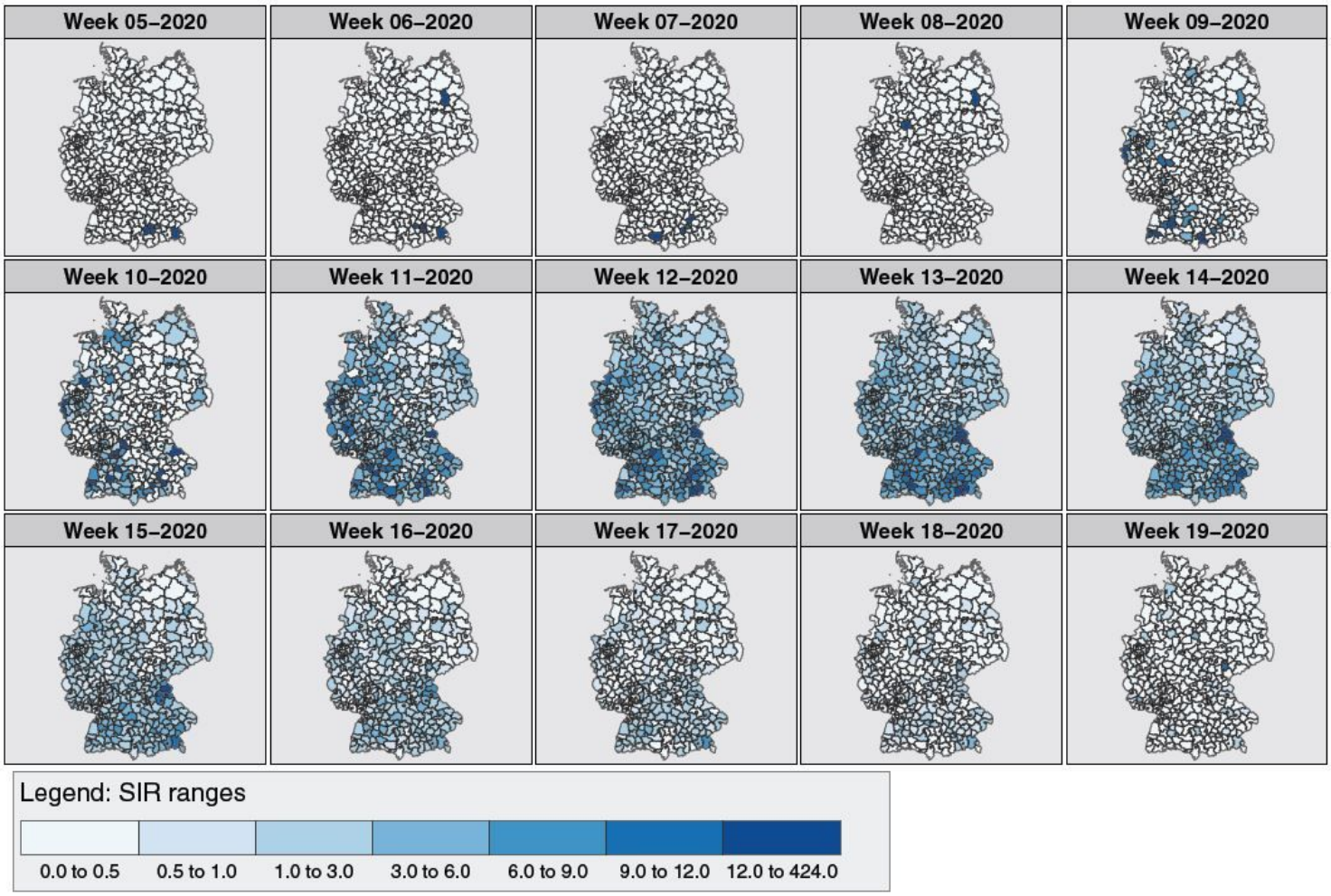

Figure 1

Standardised incidence ratios (SIR) on district level in Germany by notification week. A SIR greater than 1 means that the observed number of SARS-CoV-2 cases (O) is higher than expected (E).

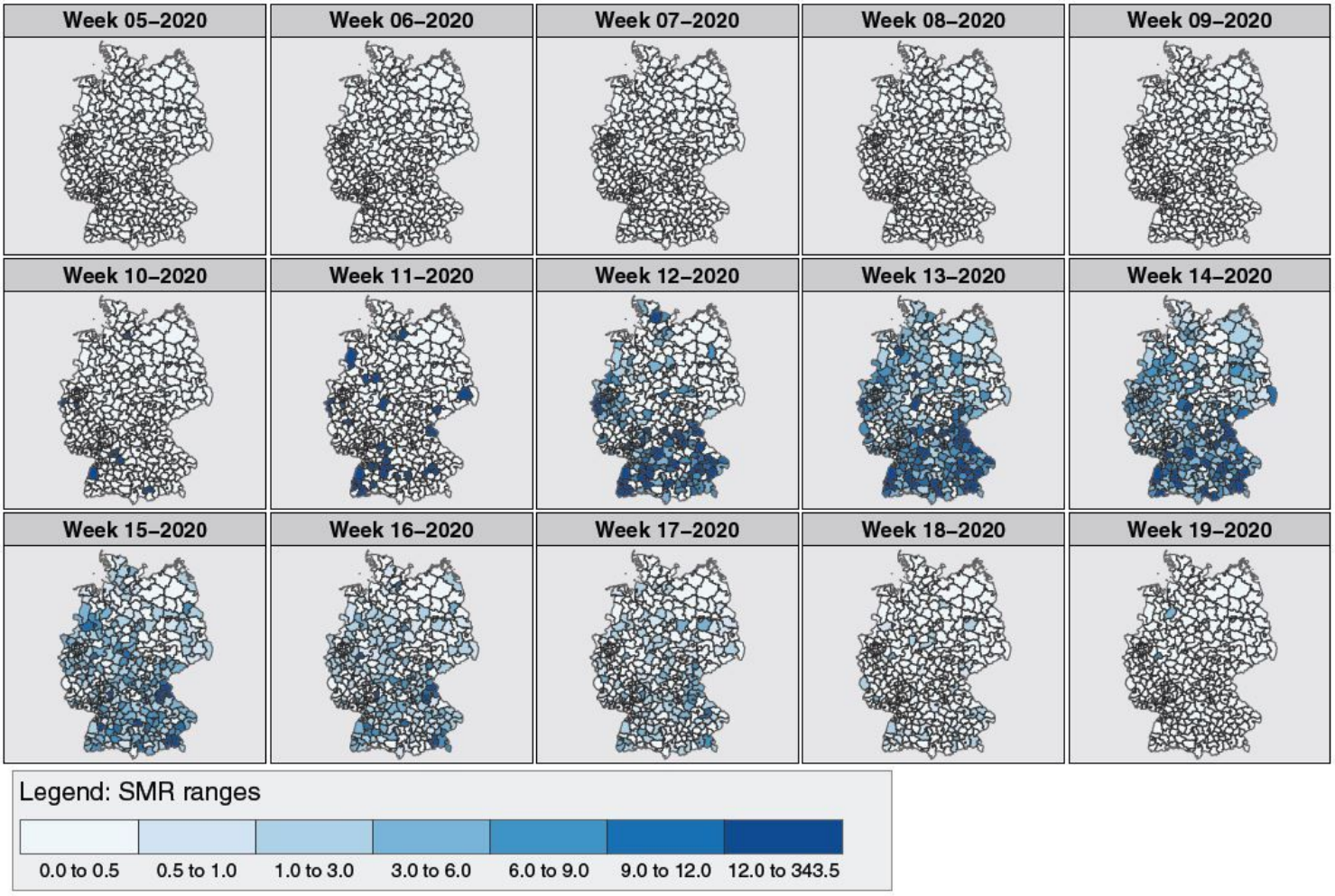

Figure 2

Standardised mortality ratios (SMR) on district level in Germany by notification week. A SMR greater than 1 means that the observed number of SARS-CoV-2 cases (O) is higher than expected (E).

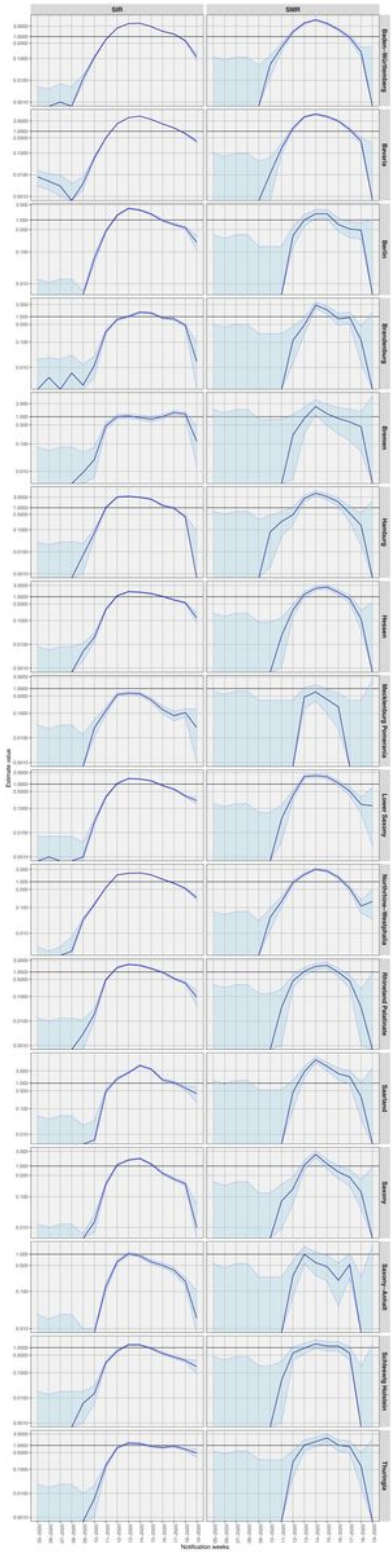

Figure 3

Standardised incidence (SIR) and mortality ratios (SMR) of federal states in Germany by notification week. A SIR or SMR greater than 1 means that the observed number of SARS-CoV-2 cases (O) is higher than expected (E). Estimates illustrated on a log-scale.

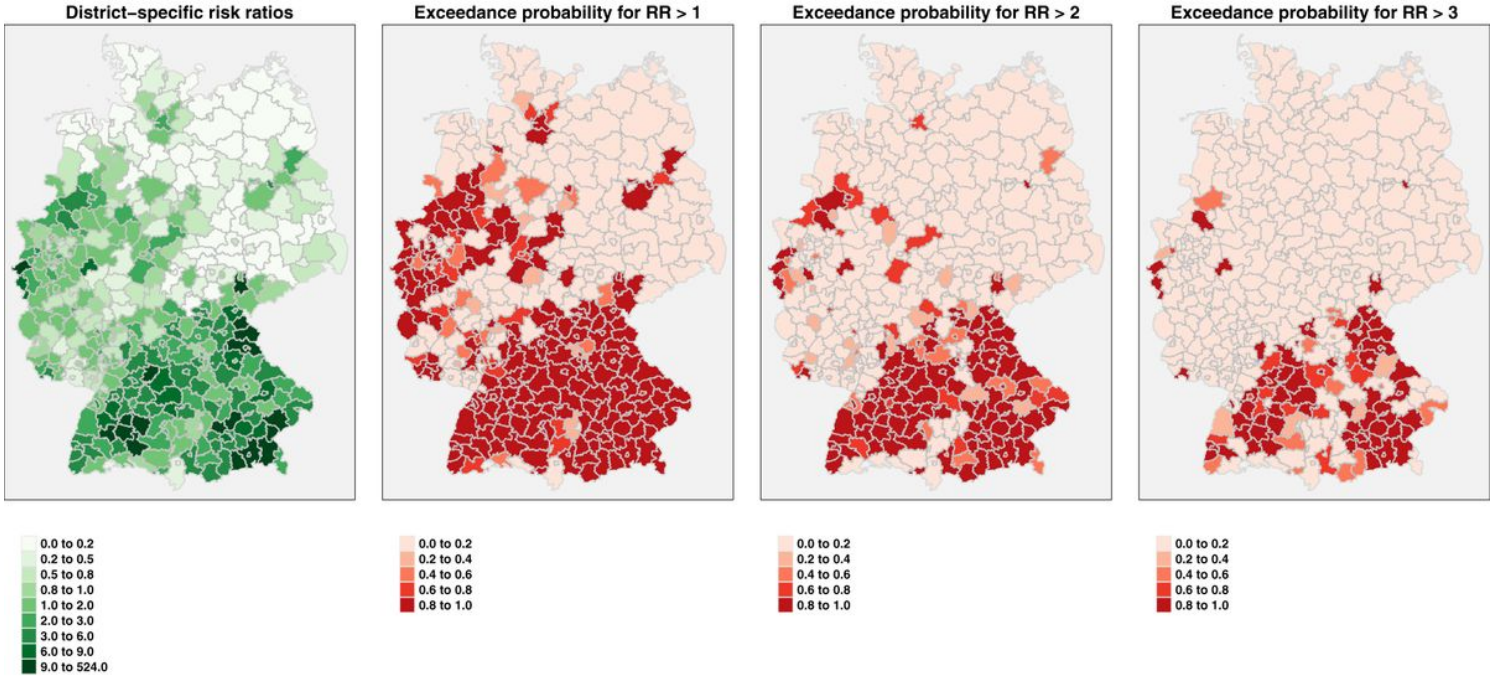

Figure 4

District-specific risk ratios (RR) and exceedance thresholds. A RR above 1 means that the posterior mean of district-specific risks is higher than the average incidence risk across Germany. An exceedance probability above 80% determines a high certainty of exceeding one of the RR thresholds.

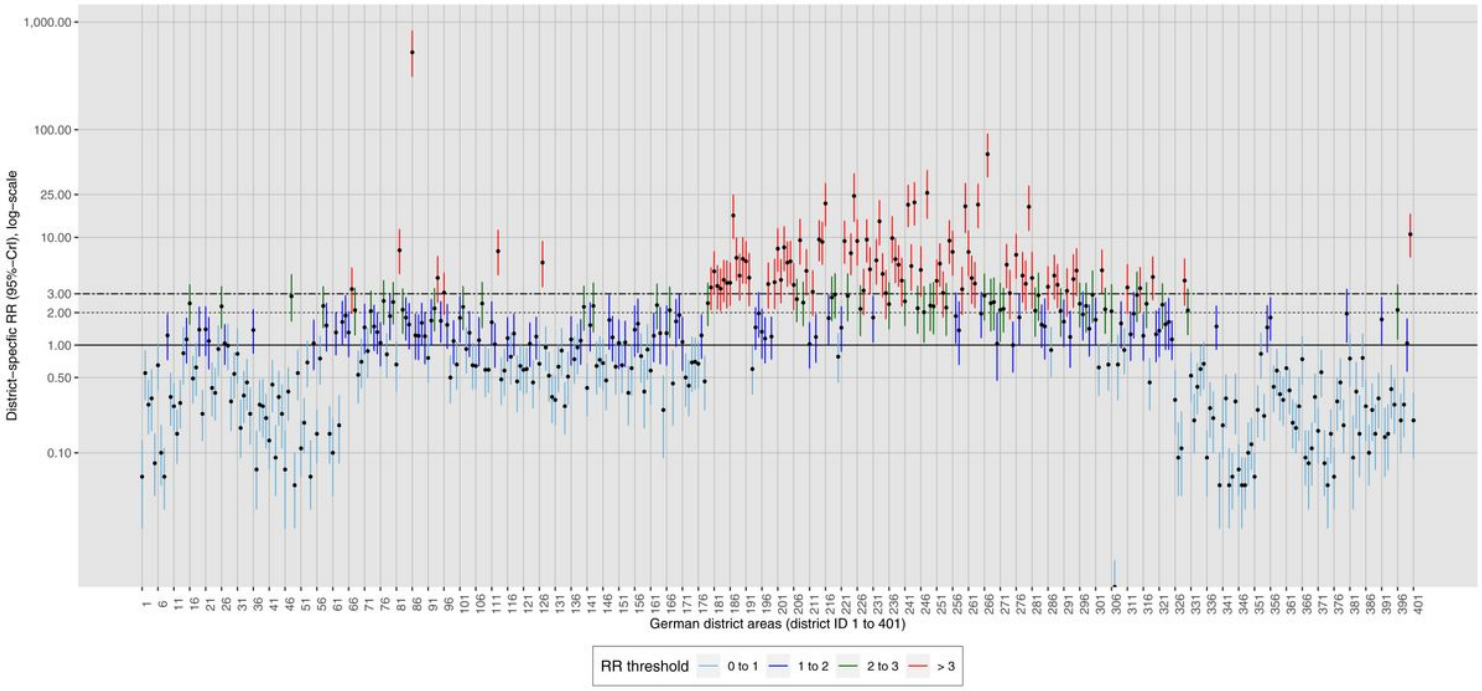

Figure 5

District-specific risk ratios (RR) and corresponding 95% credibility intervals (95%-CrI). A RR above 1 means that the posterior mean of district-specific risks is higher than the average incidence risk across

Germany. An exceedance probability above 80% determines a high certainty of exceeding one of the RR thresholds.

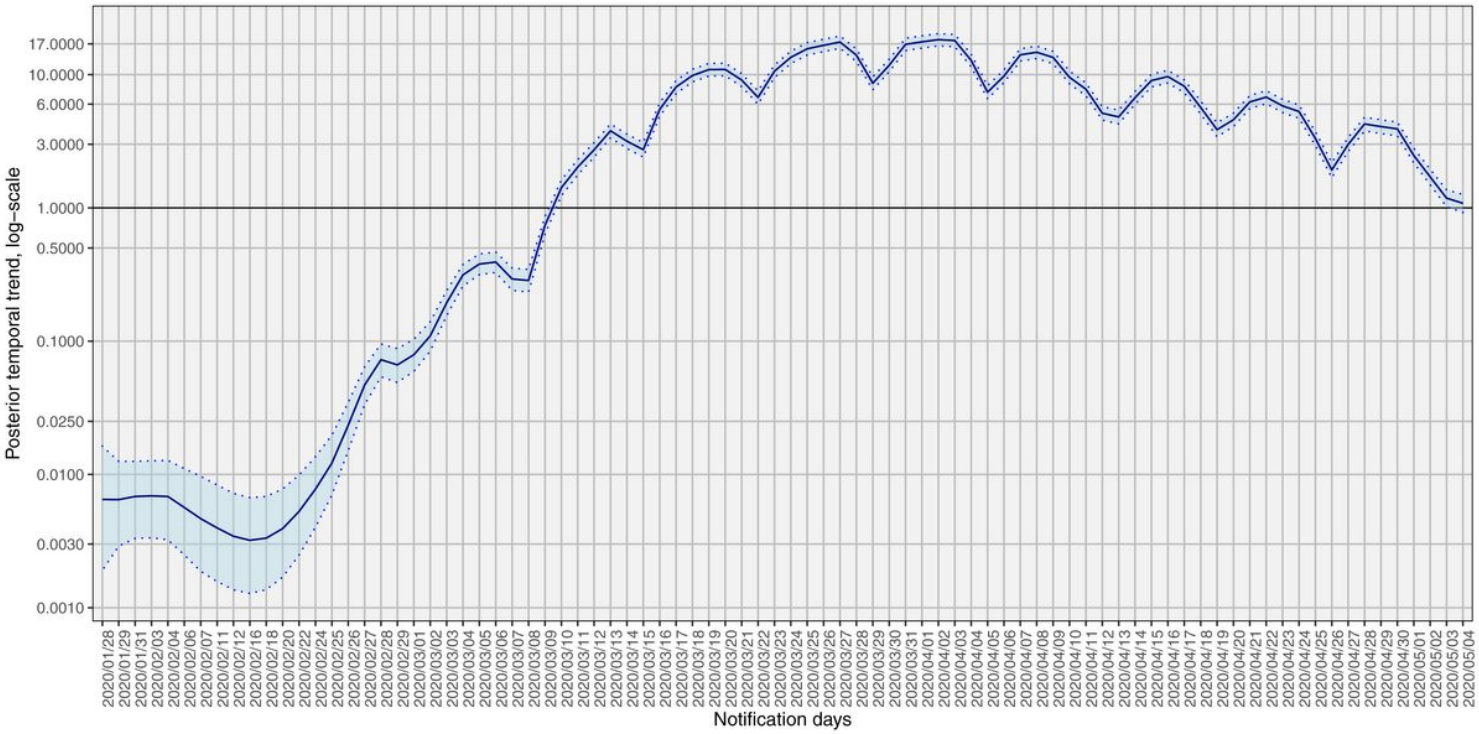

Figure 6

Posterior temporal mean trend for SARS-CoV-2 incidences in Germany. Temporal risk ratio (TRR) and corresponding 95% credibility intervals (95%-CrI) of SARS-CoV-2 incidence by notification day on a log-scale.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement10.pdf
- supplement11.pdf
- supplement12.pdf
- supplement13.pdf
- supplement14.pdf